FIELD OF VISION

915  How does the ‘environment’ come to the person? The ‘ecology of the person’ and addiction
Tretter F, Loeffler-Stastka H

EVIDENCE REVIEW

937  Therapeutic termination of pregnancy and women’s mental health: Determinants and consequences
di Giacomo E, Pessina R, Santorelli M, Rucco D, Placenti V, Aliberti F, Colmegna F, Clerici M

REVIEW

954  G-protein coupled receptors and synaptic plasticity in sleep deprivation
Parmar S, Tadavarty R, Sastry BR

981  Agmatine as a novel candidate for rapid-onset antidepressant response
Valverde AP, Camargo A, Rodrigues ALS

997  Healthy diet, depression and quality of life: A narrative review of biological mechanisms and primary prevention opportunities
Pano O, Martínez-Lapiscina EH, Sayón-Orea C, Martínez-Gonzalez MA, Martínez JA, Sanchez-Villegas A

MINIREVIEWS

1017  Neurodevelopmental disorders: An innovative perspective via the response to intervention model
Rodríguez C, Areces D, García T, Cueli M, Gonzalez-Castro P

1027  How non-rapid eye movement sleep and Alzheimer pathology are linked
Falter A, Van Den Bossche MJA

1039  Empirically supported psychological treatments: The challenges of comorbid psychiatric and behavioral disorders in people with intellectual disability
Gómez LE, Navas P, Verdugo MÁ, Tassé MJ

1053  Phantom bite syndrome: Revelation from clinically focused review
Tu TTH, Watanabe M, Nayanaar GK, Umezaki Y, Motomura H, Sato Y, Toyofuku A

1065  Breast cancer in schizophrenia could be interleukin-33-mediated
Borovcanin MM, Vesic K
ORIGINAL ARTICLE

Basic Study

1075 CPEB1, a novel risk gene in recent-onset schizophrenia, contributes to mitochondrial complex I defect caused by a defective provirus ERVWE1

Xia YR, Wei XC, Li WS, Yan QJ, Wu XL, Yao W, Li XH, Zhu F

Retrospective Study

1095 Psychiatric hospitalization during the two SARS-CoV-2 pandemic waves: New warnings for acute psychotic episodes and suicidal behaviors

Panariello F, Longobardi S, Cellini L, De Ronchi D, Atti AR

Observational Study

1106 Self-compassion and resilience mediate the relationship between childhood exposure to domestic violence and posttraumatic growth/stress disorder during COVID-19 pandemic


1116 Subgrouping time-dependent prescribing patterns of first-onset major depressive episodes by psychotropics dissection

Chen HC, Hsu HH, Lu ML, Huang MC, Chen CH, Wu TH, Mao WC, Hsiao CK, Kuo PH

SYSTEMATIC REVIEWS

1129 New-onset depression after hip fracture surgery among older patients: Effects on associated clinical outcomes and what can we do?

Qin HC, Luo ZW, Chou HY, Zhu YL

META-ANALYSIS

1147 Effectiveness of cognitive behavioral therapy-based interventions on health outcomes in patients with coronary heart disease: A meta-analysis

Li YN, Buys N, Ferguson S, Li ZJ, Sun J
ABOUT COVER
Peer Reviewer of World Journal of Psychiatry, Li Zhang, MD, PhD, Professor, Department of Geriatric Neurology, Affiliated Brain Hospital of Nanjing Medical University, Nanjing 210029, Jiangsu Province, China.
neuro_zhangli@163.com

AIMS AND SCOPE
The primary aim of World Journal of Psychiatry (WJP, World J Psychiatr) is to provide scholars and readers from various fields of psychiatry with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.
WJP mainly publishes articles reporting research results and findings obtained in the field of psychiatry and covering a wide range of topics including adolescent psychiatry, biological psychiatry, child psychiatry, community psychiatry, ethnopsychology, psychoanalysis, psychosomatic medicine, etc.

INDEXING/ABSTRACTING
The WJP is now abstracted and indexed in Science Citation Index Expanded (SCIE, also known as SciSearch®), Current Contents/Clinical Medicine, Journal Citation Reports/Science Edition, PubMed, and PubMed Central. The 2021 edition of Journal Citation Reports® cites the 2020 impact factor (IF) for WJP as 4.571; IF without journal self cites: 4.429; 5-year IF: 7.697; Journal Citation Indicator: 0.73; Ranking: 46 among 156 journals in psychiatry; and Quartile category: Q2.

RESPONSIBLE EDITORS FOR THIS ISSUE
Production Editor: Xu Guo; Production Department Director: Yu-Jie Ma; Editorial Office Director: Jia-Ping Yan.
How does the ‘environment’ come to the person? The ‘ecology of the person’ and addiction

Felix Tretter, Henriette Loeffler-Stastka

ORCID number: Felix Tretter 0000-0002-1076-1220; Henriette Loeffler-Stastka 0000-0001-8785-0435.

Author contributions: Tretter F and Loeffler-Stastka H conceived the topic in iterative discussions; Loeffler-Stastka H contributed the question how comes “the social into the mind” and added the psychoanalytic concepts; Tretter F provided expertise for systems theory, concepts on addiction, and the ecology of the person.

Conflict-of-interest statement: The authors declare no conflict of interest regarding our manuscript.

Country/Territory of origin: Austria

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in

Felix Tretter, Bertalanffy Center for the Study of Systems Science, Vienna 1040, Austria
Henriette Loeffler-Stastka, Department of Psychoanalysis and Psychotherapy, Medical University Vienna, Vienna 1090, Austria

Corresponding author: Henriette Loeffler-Stastka, MD, Dean, Professor, Department of Psychoanalysis and Psychotherapy, Medical University Vienna, Währinger Gürtel 18-20, Vienna 1090, Austria. henriette.loeffler-stastka@meduniwien.ac.at

Abstract

Currently, psychiatry lacks a field that can be called “theoretical psychiatry”, which uses theoretical concepts and explanatory models: The main stream of research is to collect data of all kinds in the hope that the computational Big Data approach will shed a bright light on the black box of mental disorders. Accordingly, the biology-based Research Domain Criteria of the National Institute of Mental Health have been established. However, as philosophical analyses of concepts and methods have shown, several epistemological gaps stand in the way of a consistent multilevel understanding of mental disorders. Also, the implicit ontological problems in the biological reduction of the psychosocial level and in the integration of so-called hard and soft disciplines are mostly left out. As a consequence, a non-reductive psychological theory of mental disorders is sought that also integrates correlating biological and sociological issues. In this context, one example of promising nonreductive psychiatric research is the option of systems/network psychopathology. The possibilities for integrating different psychological perspectives are highlighted for the field of addiction research and treatment, where pragmatic behaviorist approaches dominate over the theory-based practice of psychoanalysis. In comparing the theoretical constructs of these two approaches, the relevance of the concept of “(social) environment” as the wealth of influential sociocultural factors is discussed at levels superior to the interpersonal micro-level, namely the organizational meso- and societal macro level, which is not sufficiently considered in current biopsychiatry. On this basis of argumentation, the usefulness of grounding and framing psychiatry through the field of ecological sciences, especially human ecology, is demonstrated. Finally, to this end, an outline of an ecological model of mental health and illness is presented.

Key Words: Systems/network psychopathology; Theoretical psychiatry; Ecology of the person; Psychoanalytic object relations theory; Addiction; Mind; Resilience
©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Similar to theoretical medicine, theoretical psychiatry and its engagement with systems/network psychopathology has a research gap, and so we shed light on the question of how the “social”, respectively the environment, manifests itself in the person. Using addiction as an example, we explore systems theoretical and psychoanalytic concepts to provide a framework for understanding the sociocultural, interpersonal, and human ecological factors that impact mental health and illness. This theoretical framework provides a way to understand conceptually and structure computer-collected Big Data. The fact that humans are “situated subjects” implies a broader view of a systemic ecology of the person to guide processes of change in mental health, whether for prevention or treatment.

INTRODUCTION

Nothing is as practical as a good theory[1].

Theory in psychiatry?
The call for “theoretical psychiatry” is not often made, but it seems a useful, but challenging, task[2,3]. But what is a “theory”? Even in the philosophy of science there is not much consensus about essential features of theories. Empiricists see theories as final stages of induction, while constructivists see even observations as a kind of theories. Here the hybrid Kantian position might be useful, which distinguishes between empirical and theoretical knowledge but claims that they are related. In this context, and with respect to applications in medicine/psychiatry, there are several descriptions of theories to consider: Theories can be characterized essentially as “because” answers to “why” questions that appeal to causal principles, such as in classical physics: The apple falls from the tree because of gravity, namely, reciprocal forces that are inherent in all objects and are supposed to be exchanged between them[4]. Applied to medicine/psychiatry this could mean: The disease arose because the person was exposed to stressors. Scientific theories are logically consistent statements about reality. Their central concepts must be clearly defined (“disease”) and should be supported by empirical data[5]. In this regard, theories are based on a set of empirically verified hypotheses, where hypotheses are mainly statements with an “if x, then y” or “the more x, the more y” structure. Hypotheses are inductions from empirical data and, once strongly confirmed, are called empirical laws and can be developed into a more complex construct (gravity) that requires and allows for further explanation. In medicine (and psychiatry), a hypothesis is very often already classified as a theory. However, hypotheses are not sufficient to understand complex phenomena such as health and disease, because confirmed empirical laws are only a basis of theories, which expand into highly formalized concepts, at least in physics and chemistry. Of course, as derived statements about reality, theories should be falsifiable and therefore must be “grounded” by observations and data, otherwise they are according to Karl Popper only immunized belief systems. Theories intend causal explanations and should thereby also the range of their validity: Isaac Newton’s theory of gravity is useful for everyday mechanics, but Albert Einstein’s relativistic theory of gravity also covers processes in the subatomic world.

In summary, and consistent with such global descriptions, theories can be understood as explanations, in particular, of new observations and data by “nomological deduction”, which also enable predictions, even if these are only probabilistic and not deterministic in character[6]. In short, theories enable the ordering of observations.
Although this metatheoretical conception of a theory has been questioned in philosophy of science, at least with respect to the relevance of principles (or laws) in biology, psychology and sociology (nomological explanations), mechanistic models can be a sufficient formulation of a theory[7]. In this context, the systems theory view, as we partially apply it here, has a paradigmatic epistemological position that refers to principles such as “dynamic equilibrium” and focuses on building testable models of networks and circuits[8-11].

**THEORY IN PSYCHIATRY – ONLY A BIOLOGICAL REDUCTION WITHOUT INTEGRATION?**

Again, the call for a “theoretical psychiatry” is a challenging task[2,3] due to its complexity. Therefore, further elaboration is needed.

**Diversity of psychiatric “theories” – Different perspectives**

Within the description of theories and a brief review of the history of the scientific roots of psychiatry at the transition from the 19th to the 20th century, different approaches can be identified[12]: Kraepelin[13] emphasized experimental psychology in order to obtain more accurate data on mentally ill patients[13], while Jaspers[14] focused on the world of experiences with his influential phenomenological psycho-pathology. Another psychiatric giant of the times, Bleuler[15] had sympathies for the broader framework of psychoanalysis launched by Freud[16] and based on qualitative data and extensive theories[16].

After this period of more or less psychological approaches to mental disorders, the discovery of successful psychopharmaceutical approaches in the 1950s ushered in biological psychiatry, which continues to dominate psychiatric thinking today[17,18]. Also in the 1950s, social psychiatry began the scientific study of the social dimension of mental disorders leading, for example, to the discovery of class-specific inequality in the incidence of mental disorders[19]. However, the study and organization of health care was also increasingly a topic of social psychiatry[20]. In particular, international comparisons of epidemiology and culture-specific conceptions of mental health and health care demonstrated the need for cultural psychiatry[21].

In agreement with these main disciplinary directions of psychiatric research, it must be acknowledged that mental disorders as objects of knowledge in psychiatry are multidimensional phenomena, which - already for their differentiated description - require biological, psychological and sociological concepts and research methods, and which are in principle explained by the integrative bio-psycho-social model proposed by Engel[22] in the late 1970s. This model was intended to allow each individual mental disorder, such as addictions, to be represented within an integrative conceptual framework[23].

However, in the 1980s, and contrary to Engel’s suggestion, scientific psychiatry became increasingly one-dimensional through the extensive use of, of course valuable, (neuro) biological methods, data, and perspectives. As a result, sophisticated imaging techniques can be used in clinical research together with molecular biological tools that generate a huge amount of data, and new psychiatric drugs have also been developed. In line with this development, the book “Brave New Brain” in 2001 summarized the value of the biological approach in psychiatry as a strong programmatic perspective that is still growing and even dominating psychiatric thinking[24]. In terms of practical psychiatry, psychological techniques [e.g., cognitive behavioral therapy (CBT)][25] have been explored and added to pharmaceutical treatment. Planning and intervention in family and community relationships also became an important topic in multidimensional practical psychiatry.

Hypotheses were formulated at the theoretical level, derived from biological data and obtained using increasingly sophisticated technical methods. These hypotheses have been empirically tested, but almost no complex theoretical concepts have been developed that are “explanatory” in the strict sense. Mainly statistically based constructs like the “(individual’s) susceptibility to stress” were supposed to explain the variance in the data world[26]. However, as mentioned at the onset, such constructs are not theory in the sense of theory in the context of physics.

**Criteria for the research domain – an integrative framework?**

As a consequence of data, and technology-driven biological psychiatry based on the initial brief description of theory, we are currently confronted with an almost theory-
free psychiatry that wants to refer to data, hopefully Big Data[27]. This-critically speaking-empirical and biocentric complex research program was supposed to be integrated a few years ago by the conceptual framework of Research Domain Criteria (RDoC)[28]. RDoC essentially aims to integrate behavioral science and neuroscience into a matrix of five functional domains, namely social functions, cognitive functions, positive valence functions, negative valence functions, and basic regulatory functions described by different types of data such as genetics, molecules, cells, circuits, physiology, behavior, and even self-reports. This comprehensive approach promises more order in understanding the wealth of knowledge in psychiatry. It ultimately aims at a bioscience-based new taxonomy of diagnoses and thus stands in some contrast to the Diagnostic Statistical Manual[29] with its symptom-centered taxonomy of classical clinical psychiatry as used by clinicians.

One of the most impressive positive examples that fits well with RDoC is the neurobiology of addiction, which was developed before RDoC. Most notably, Koob and Le Moal[30], Koob[31], and Koob and Volkow[32] have stimulated theory development in addiction research through neurobiology[30-32]. For example, they consider the opposing functional organization of the reward system (nucleus accumbens) and the stress/punishment system (amygdala) and outline a stage-based concept that shifts from a brain center-oriented paradigm to a (“systemic”) neuro-network view. It also draws on the behavioral biology of addiction but is likely to be of limited use in clinical work. We will add to this approach in the second part of this paper.

In general, RDoC seems to favor “behavioristic psychology” over phenomenological approaches to psychological issues, especially psychopathology. Apparently, there is also a relevant epistemological gap between objective and detached psychiatric research in the laboratory (third person perspective), based on physiology and behavioral science, and subjective and interactive clinical observation and practice at the bedside, albeit standardized, based on self-reports (first and second person perspective). Some of these basic epistemological contradictions have already been discussed (e.g., in relation to the interdisciplinary study of working memory[33] and also in relation to the foundations of RDoC)[34].

**Epistemological gaps – challenges for bridging concepts**

These philosophical (or metatheoretical) issues in psychiatry are rarely discussed and only by a few researchers[35-38]. Regarding the theoretical integration of knowledge, due to the multiplicity of methods that are used to study mind and brain[39-40], there are at least three main gaps that cannot be reduced by physicalism and cannot be captured by a non-reductive physicalism within an “integrative neuroscience”, since these multiple dimensions seem to be emergent properties of humans[41,42]: (1) Intradisciplinary integration in biological substrate research: The main problem of contemporary biology is related to the part-whole problem and implies that molecular biological knowledge cannot be easily linked to questions of whole-organism physiology: This problem also implies that data from physics and chemistry are difficult to combine. For example, there is no consistent theory of information processing of the synapse that relates electrical digital phenomena, such as spiking frequency of presynaptic neurons, to proportional transmitter release and analogous changes in membrane potential of the postsynaptic neuron by integrating electrophysiology with molecular biology/pharmacology that can “explain” and justify robust formal theories of macro-physiological processes of the brain such as epileptic seizures[43-45]; (2) Interdisciplinary integration between biology and “psychological sciences” (psychology, psychopathology, psychotherapy, clinical psychiatry, etc.): In addition to the gaps in intradisciplinary integration, no “theory” (as described at the beginning) but only mechanistic neurobiological models are developed to explain mental phenomena and their disorders. A simple reduction of the mind to the brain has not yet been realized and is perhaps impossible in principle[46]. As noted neuropsychiatrist Nancy Andreasen has also pointed out[47], “The application of technology without the guidance of astute clinicians with specific expertise in psychopathology will be a lonely, sterile and perhaps fruitless endeavor.” Consistent with this insight, the foundation of clinical psychiatry remains descriptive psychopathology as the “interdisciplinary, theoretical, and an empirical study of subjectivity”[48]. The use of technology in clinical psychiatry has been a major challenge. Accordingly, it is obvious that not only biological research is fruitful, but also a structural-functionalist psychology without explicit reference to neurobiology can produce new insights: A “network psychopathology” that emphasizes the interactions between symptoms, as realized by Denny Borsboom’s group at the University of Amsterdam using graph theoretical tools, offers new possibilities to reconstruct psychopathology without...
neurobiology[49]. The results of this approach can be aligned with neurobiology, but there is no must[50]. This view fits well with our systems model of mind proposed later; (3) Interdisciplinary integration between psychiatry and social sciences (e.g., integration of mental/behavioral disorders and sociocultural issues). In other words: How does the sociocultural environment enter the mind/brain? Basically, it must be conceded with Popper[51] that there is an ontological uniqueness of the sociocultural world (“World 3”[51]). His analysis also justifies a more or less substrate-free social and cultural research in psychiatry, which cannot be replaced by the very limited “social neuroscience”. The latter seems to be limited to the micro-level of interpersonal interaction. Moreover, the social world is only one dimension of the real environment, because there are natural objects, technical objects, social objects, cultural objects etc., as will be explained in the third section of this paper. In particular, the closures due to coronavirus disease 2019, by deprivation show the influence of the whole environment on mental health. For this reason, we focus on the last point, namely the notion of environment.

To conclude this first chapter, our next step is to highlight some of these theoretical and metatheoretical issues, namely the difference between external and internal conditions of mental health—through challenges to a theory of addiction. We focus on the clinical limitations of the stimulus response model and the need for subjectivity and a differentiated model of the mind that also includes essential elements of the diversity of the factor “environment” in an ecological perspective. One goal is to find a productive theoretical bridge between behavior therapy and psychoanalysis, with an emphasis on the latter, even while being aware of basic criticisms from philosophy of science[52,53].

THEORIES OF ADDICTION – FROM PSYCHOANALYSIS TO THE STIMULUS-RESPONSE MODEL TO PSYCHOANALYSIS AND THE ECOLOGICAL PERSPECTIVE

As mentioned earlier in reference to RDoC, addiction is currently discussed primarily from the perspective of neurobiology. Addiction psychiatry is now a laboratory- and animal-centered perspective and has fewer and fewer connections to the clinical perspective. Addiction research is even conceptually reshaping clinical thinking. Therefore, the question now arises as to how the critical arguments made earlier fit with a “practical theory” of addiction: Are clinical observations more “real” than laboratory data? Since addiction is primarily a psychological problem, it should be kept in mind that the history of concepts in psychology (and psychiatry) shows that the subjective perspective has been increasingly avoided, and “objectification” has become the guiding feature of scientific psychology/psychiatry. Although psychoanalytic explanations of alcohol addiction were among the first theories of addiction, due to various shortcomings in empirical research, concepts, methodology, theory, and the practicality of treating mental disorders through psychoanalysis, the behaviorist perspective prevailed among professionals. However, since addiction is a chronic disease, in rehabilitation and long-term treatment some of the aspects proposed by psychoanalysis, especially in relation to the holistic domain (the person in the world), still seem relevant. If the concept of environment is constructed in a differentiated but integrated way, encompassing both significant people and drugs as components of the person’s environment, a human ecological perspective might also be conceptually connectable to psychoanalysis. This will be discussed in the final section. In this section, we will focus on the components of theoretical compatibility of psychoanalysis and cognitive behavioral therapy theory as conceptual building blocks for an integrative model of the mental.

The early psychoanalytic view of addiction – the topic and the structural model
Briefly characterizing some of the roots of psychoanalysis, it is a complex theory of human behavior and a treatment method that emphasizes unconscious repressed affective-motivational drives of behavior. It focuses on the hypothetical functional structure and dynamics of the mind and assumes a powerful unconscious layer of the mind. These are the key points that Sigmund Freud elaborated with the first topic model (conscious/preconscious/unconscious) and the second topic model, the structural model (id, ego, superego). In relation to addiction, Freud[54] around 1898 interpreted addiction as a kind of masturbation as a substitute behavior for (sexual) dissatisfaction[54,55]; it is a purely self-centered behavior necessary to experience
pleasure. Later, in the 1930s, Fenichel[56] interpreted excessive drinking as a merging of the id with the ego, in order to fight together against a rigid but mainly implicit superego, which can additionally cause guilt and shame, even in a circular causal process[56]. With regard to systems theory, this model identifies the central driving force of addiction as an intrapsychic difference between a set of desired values and real values that cannot be changed by the person, which is why he or she experiences a tension that can be subjectively reduced only by consuming psychoactive substances.

Remarkably, the therapeutic effectiveness of psychoanalytic treatment of addicts did not reach widespread use, and soon the behavioral approach in psychotherapy began its triumph. The problem of addicted clients at first contact – even after detoxification - did not show stable sobriety for analytic treatment, the lack of experimental and quantitative empirical evidence, and also the apparent arbitrariness of theoretical concepts devalued psychoanalysis of addiction and hindered the development of psychoanalysis in this clinical field. As a basic result, a fundamental reorientation toward learning theory occurred in clinical psychology. In the following, some of its basic theoretical features are briefly described and compared with psychoanalytic concepts.

**Behaviorist view of addiction – the stimulus-organism-response model**

Since the beginning of the 20th century, the behaviorist conceptualization of mental states and processes has been increasingly favored in psychiatry and also in addictive disorders with reference to the stimulus-response (SR) model, with the organism as a black box (or grey) box [stimulus-organism-response (SOR) model] without presupposing mental states and processes. For this basic epistemological position of analytic behaviorism/positivism[57], animal experiments in particular have provided the empirical basis for theoretical considerations. Here, we also start from this clear but superficial perspective; however, we will also criticize it in order to outline finally an ecological perspective through an onion skin-like systemic multilevel model of the situated person that takes into account a macro/micro-layering of environment and person.

Fundamentally, addiction can be characterized within a SR model as a learned extensive affinity for an object, which usually is a drug as a chemical stimulus[58]. A more theoretical definition characterizes addiction as “a syndrome at the center of impaired control over a reward seeking behavior”[59]. This implicitly assumes antagonistic forces (drivers and brakes) that determine overt behavior.

As an extension of the SOR model, the SORKC model has been successfully proposed, emphasizing the power of sustained contingencies (K) of consequences (C) [60]. In addition, classical conditioning considering contextual stimuli that become conditional stimuli and social learning through a behavioral model has been shown to be useful in understanding specifics of addiction development[61]. In treatment, all these perspectives are used as theoretical reference points for the development of therapeutic strategies. Recently, even the simple SR concept of habitual addictive behavior was successfully applied to a computer-based training to avoid alcoholic beverages presented as an image on the screen that had to be pushed away with a joystick. This training with virtual reality tools helps to balance addictive sensorimotor routines[62]. Here, we go into some conceptual details of the behaviorist approach.

**Reinforcement as reward – does it correspond to pleasure?**

Within the SORKC model, the occurrence of addiction is conceptually explained by reinforcement mechanisms, which represent the animal model of operant conditioning. These mechanisms are often referred to as reward mechanisms, especially in the context of neurobiology. For clinical tasks, a subject-oriented view is helpful: The (increasing) affinity to a drug can be experienced by the addicted person himself, after drinking alcohol most people experience an improvement of mood, as already impressively described by the alcohol dependent writer London[63] in 1913. In other words, the reinforcer (or reward) is experienced, in anthropomorphic terms, as a pleasant state and/or a reduction of an unpleasant emotional state. This improvement in emotional state acts as a positive feedback loop on the behavior, increasing the likelihood of its occurrence and thus “reinforcing” it[64]. Some pleasurable experiences are associated with the reduction of tension or stress. Consistent with this view, the early Freud’s version of the pleasure principle[65] appears within new guise[65].

However, another question arises: Why do some drug-exposed individuals develop addiction and others do not? Is there an intrinsic risk for addiction? This question can be explained by looking more closely at the external conditions but much more so at internal conditions. For example, psychiatric comorbidity (anxiety, depression,
borderline personality disorder, etc.) may “drive” alcohol use, although some of these psychopathological conditions may also be “caused” by alcohol[66]. This organism-based explanation is theoretically consistent with the concept of affect regulation or stress management through drug use, and in the context of psychoanalysis of addiction, it is consistent with the complex self-medication hypothesis[67].

The cognitive turn – modeling thought cascades without a self-model?

In the experimental context, the classical SOR model was soon improved by admitting that “internal” cognitive processes and representations matter for understanding behavior, a view effectively proposed by Tolman[68] and Neisser[69]. The cognitive approach also implies that the organism must have an internal model of the external world and also of the organism itself, which will be discussed later.

In the clinical context, Beck et al[25] introduced the cognitive turn for almost all mental disorders, focusing on the mood-organizing power of cognitive processes[25, 70]. The cognitive model of addiction is based primarily on conscious self-descriptions: Situational perceptions, expectations (e.g., of drug effects), thought processes, decision making, explicit memory, behavior planning, and other states and processing have been considered relevant to addiction. For example, in a stressful social situation, an addicted user thinks that he or she is disliked by others, with these thoughts reinforced by underlying dysfunctional cognitive schemata (“I am a loser”). These cognitive schemas and cascades of dysfunctional thinking are the targets of therapeutic intervention in CBT for mental disorders: Change in addictive behavior is targeted through changes in situational processing, cognitive coping with stress, consideration of positive and negative consequences of drinking and abstinence, etc.

Although a foundation of CBT is cognitive schemas, which are also associated with a concept of a self-model of the person through the notion of “dysfunctional assumptions”, an explicit definition of the self/self-model is not usually used, as it is conceptualized only as a recursive (rigid) self-description of the person (“I am a loser”). In terms of cognitive schemata as a theoretical building block in CBT, there is a conceptual correspondence to the psychoanalytic concept of (dysfunctional) self-representation. In relation to this theoretical concept of self-representation, Kohut[71, 72], a pioneer in analytic self-psychology, has influentially explained drug use as a pathological narcissistic act[71,72]: The developmental non-existence of the emotional other, caused, for example, by distant parents, induces a deficient self-representation (and object representation), with the drug literally filling up these empty spaces of images (defect of the self). The drug functions as a “self-object” and contributes to psychic homeostasis. This theoretical position also makes it possible to understand the difficult therapeutic relationship in which the addicted user uses the therapist in a narcissistic mode, especially when the client is abstinent, he uses the therapist in the transference situation in a parallel mode to the drug.

Behaviorist models of self-regulation – who is the regulator?

The old notion that the drug initially serves as a regulator of the addict’s affective state has also been successfully developed within the framework of learning theory and cognitive theory and can be subsumed under “self-regulation” models. As early as 1970s, some researchers such as Kanfer and Goldstein[73] and Kanfer[74] in the context of learning theory criticized the concept of stimulus determination of behavior and passivity of the person as seen in the framework of SOR model[73,74]. He emphasized the possibility of “self-management” for patients (e.g., in tobacco addiction). Self-observation, self-evaluation, and self-appraisal are circularly linked process components of smoking behavior control that can be influenced by behavior therapy. This influential model corresponds conceptually with models that focus on “self-control” and “self-regulation”[75,76]. Interestingly, why and how regulation is regulated is often not clarified. From a metatheoretical perspective, these models are process-oriented and neglect to model structures of the mind. In other words, these models do not explicitly refer to the concept of a “self” or “self-model” or an “ego” operating in the loop. In contrast, within the psychoanalytic structural model, the ego is the regulator that can use drugs to relieve tensions between external reality, the id, and the superego. Moreover, in relation to the topic model, drug use can be understood as a defensive strategy of the ego to suppress or overcome implicit emotions such as anxiety, anger, and depression that occupy and deform explicit consciousness (i.e. “self-medication hypothesis”).

The integrative power of the psychoanalytic model – translation of concepts

So far, it has been shown that some phenomena of addiction are well covered by data
and models of academic psychology, and others can be explained by psychoanalysis. Theoretical overlaps between these two approaches - for example regarding the function of the “self” - differ in the definitions and theoretical elaboration of the respective models. To achieve better integration, a combination of behaviorist and psychoanalytic perspectives might be useful, especially with regard to the heuristic significance of the assumption of drug-induced mental homeostasis based on an unbalanced mental system. An interesting special theoretical aspect of psychoanalytic models is that they are implicitly “ecosystemic” models of the mind, in that they account for multiple internal interactions of environmental feedback loops. They also appear as a kind of “ecological” model, since it is assumed that the concrete social environment of the person, namely the family into which that person was born, represents a set of factors that determine the mind. However, this (implicit) ecosystemic view also makes it difficult to identify, for example, a “causal locus of control”. Is the ego too weak, is the superego or id too strong, is the self model or object model too deficient, or are the resulting conflicts too strong? Here we can only refer to some theoretical questions. (1) Regarding the conceptual connection between the reward and pleasure principle, *i.e.* mood elevation through drug use, the question arises: What are the conditions for unpleasure and/or pleasure? Pleasure is difficult to define because of the qualia problem (see first paragraph), but it can be experienced in the release of tension. Tension can be explained in systems analytical terms as the result of the clash of opposing forces such as urgent drives and behavioral inhibitions (go/no-go condition). In the context of the structural model of theoretical psycho-analysis, this description reflects the contradictory relations within and between the affective-motivational id and the repressive superego experienced by the ego. But drug exposure raises the question: Who is stronger, the driving id or the suppressing superego? From the psychodynamic perspective, drug use can also be interpreted as an additional defense strategy of and for the ego, aimed at reducing negative emotions and the experience of helplessness or powerlessness[77]; (2) A pathogenetic predisposition to addiction through the dominance of the superego was favored by Wurmsen[78]. Wurmsen[78] emphasized, in agreement with Fenichel, that drug-induced gratification of the infantile narcissism by removing the control of the superego is a major cause of persistent drug use. However, this model needs to be diversified in terms of the architecture of the conflicting forces: For example, the urge to work until a task is completed is usually implemented and enforced by the parents, in most cases the father. Failure to achieve these goals leads to bad feelings. Functionally, the more or less implicit idea or representation of the challenging father is a component of the superego. In addition, one might assume that the internalized image of the supportive mother allows for relaxation. However, it is important to keep in mind that the internal images of parents include more than their normative and supportive components; and (3) there is the question of where and how parents (and/or significant others) are conceptualized in SOR or cognitive models. For example, the dysfunctional assumptions (“I am a loser”) are valid theoretical constructs in CBT, but this model raises as to their origin. To answer this question, it can be hypothesized that the parents’ recurrent comments about the person’s behavior are stored as basic reference values and as the core structure of the person’s idealized self-image. This explanation of the emergence of “dysfunctional assumptions” in the context of CBT is consistent with the psychoanalytic model regarding the functional role of the superego and the self-image, which must be viewed in relation to the image of significant persons in the environment. This is elegantly captured by object relations theory and the concept of ‘internal working models’[79,80]. This is currently being discussed in the context of the new theories of psychoanalysis.

**Toward an integrative theoretical model of the mind**

Consistent with the arguments presented earlier, we propose to design a psychological theory for psychiatry using conceptual building blocks from psychoanalysis. In his search for the origins of human knowledge and the role of memory in our ability to experience, Freud concluded that many of human mental processes although unconscious, influence our overt behavior. In his analysis of symptoms, he postulated that each of them makes sense when one considers that important processes of the mind operate below the level of conscious awareness. Drives associated with basic needs (*i.e.* basic biological instincts) and social demands directed to the individual by the environment and context shape the person’s personality, intrapsychic structures (ego, id, and superego), and mental capacities. Repeated interactions with the caregiver and other relevant persons in early childhood shape the way reality is experienced and perceived, leading to implicit knowledge and behavior patterns[81-84]. Emotional reactions that occur during interaction with others and the environment
are perceived and are more or less regulated from moment to moment. Thus, they can be transformed into signals that act as meaning markers and influence memory encoding and cognitive processing. The early social environment, such as relationships and interaction patterns with primary caregivers, influences behavior (e.g., exploratory urges, attachment patterns, expression of mental and affective states). Norms and values have to be integrated; commandments and prohibitions of the culture first are mediated by the respective others; submission to them is often ambiguous and has to be worked out first in order to be tolerated without doubt[82].

Within these reciprocal experiences with the reference person and the basic environment, adaptation to reality is one of the main goals of psychoanalysis, reality being the starting point of unconscious fantasies, indicating the distinction between reality and fantasies is important for mental health and illness[85]. Psychodynamic considerations take into account subliminal signals (e.g., through language, gestures, facial expressions) and latent perceptions of stimuli as well as unconscious and embodied knowledge; they aim at reshaping cognitive processing (e.g., perception, attention, memory) by creating new ways of seeing and experiencing in the analytic relationship.

**A framework for the processing subsystem**

Taking together, these psychoanalytic issues and various cognitivist concepts used to describe the addict’s internal processes, such as expectation, cognition, memory, decision, etc., and also considering the importance of affective internal representations, an integrative view of addiction-related human information processing yields a systems model of the mind. This systemic model must first correspond with the symptom checklists of clinical psychiatry, which explore perception, thinking, memory, emotions, drives, etc. as functional elements of the mind, here linked in a basic conceptual network of the mind. Also, the concept of a self with an operational foundation of psychoanalysis when Sigmund Freud developed his influential conceptual framework of the Oedipus conflict that, today of course, requires more subtle interpretations and clarifications. In psychoanalysis, the patient’s relationships (more or less explicit) image of his parents but also in the reality as seen by the therapist (Figure 1).[86,87]. Since the first of these modeling attempts, several similar but partial models have been proposed in addiction theory[88,89].

**The structure of representations as a reference system**

In the proposed systemic model of the mind, the level of representations (or images or affective-cognitive schemas) serves as an internal reference system for further information processing, and it is also modified by this processing (EI and SI in Figure 1). It consists of three subsystems: Representations of the self, of the environment (“objects”), and of the interrelations. These representations are processed through thinking and emotional appraisal and stored in different layers of memory, so that they can function as a reference system. They are products and contents of mental processing.

In this context, it is important to emphasize that in the behaviorist view, the environment is generally conceived as the world of “stimuli”: Looking more closely at the world of “stimuli”, drugs (such as a bottle of beer), for example, are “salient” stimuli in the addicted person’s environment, and they have an “incentive” quality for consumption. From a psycho-ecological perspective, they have a positive valence. However, the many efforts to interpret the concept of “stimulus” in the concrete “environment” of the addicted person are not fundamentally related to the trivial fact that the family (and its drug use) is the first type of social environment. The family as a system with its various elements such as mother and father and their relations (harmony, broken domestic situation etc.) and their relations with drugs is the framework for the child’s drug use. The family exists both as an external reality and as an image in the mind, and it influences the client’s experience even when parents are already dead. This issue of the contextual conditions of mental disorders was a foundation of psychoanalysis when Sigmund Freud developed his influential conceptual framework of the Oedipus conflict that, today of course, requires more subtle interpretations and clarifications. In psychoanalysis, the patient’s relationships with parents are a central theme of therapeutic reflection and intervention[90].

This leads us to the heuristic value of a wider concept of “environment” that describes the surrounding world of the individual as it acts as an initial frame for behavior, for example, the family of an adolescent drug user (e.g., the son): The tension caused by the incompatible relationship with the overdemanding father and the overprotective mother can lead to a chronic conflict. This conflict has two epistemic sides: It plays out not only in the psyche of the adolescent with the dysfunctional (more or less explicit) image of his parents but also in the reality as seen by the therapist (Figure 2). The experienced incoherent action structure of the parental dyad
Figure 1 Visualized systems model of the mind. The list of subsystems being checked in psychiatric examination are represented here as interconnected mental operators (subsystems) within an operationally closed network. Consciousness is represented here as a shaded circle, regarding the metaphor of light and dark. Not all relations, especially of the ego/self, are depicted here in detail[88]. Perc: Perception; Expe: Expectation; Thin: Thinking; Mem: Memory; Plan: Plans of behavior; Beh: Behavior; Des: Desires; Emo: Emotions; EI: Environmental image; SI: Self-image as affective-cognitive schemas.

Figure 2 Model of reality and affective-cognitive representation of the adolescent drug using person (P; D, drug) and his or her family as the environment (E): The demanding father (F, with line with transom) and the supporting mother (M, with arrow) induce tension for the child that results in drug consumption for sedation and/or escape. In addition, the parents oppose each other (line with two transoms). The model depicts the reality (E, P, F, M, D) and the mental representation by the patient (E’, P’, F’, M’, D’).

may be an important trigger for drug use. Consequently, the family structure of the young drug user is one a very important topic for therapeutic exploration.

However, in line with this theoretical and practical insight, the term “environment” needs to be more differentiated, as the adolescent drug user is involved in other environmental subsystems: His peer system, the world of school, the leisure space, the community, the internet, etc. In other words, not only the family but the whole “ecology of the drug user” has to be explored by the therapist[91]. This complex relevance of the term “environment” (and relations) demands some deeper meta-
“ENVIRONMENT” – ORDINARY LINGUISTIC TERM OR PHILOSOPHICAL CONSTRUCT?

The theoretical importance of a solid conception of “environment” is not sufficiently considered in psychiatry. As for biological reductionism in psychiatry, the goals of environment-related research also exhibit some conceptual and epistemological simplifications that are misleading. Surprisingly for biologists, the use of the term “environment” in biological psychiatry usually is mostly based on a very vague meaning of everyday language: For example, in the extremely important gene-environment interaction studies 20 years ago by Caspi et al[92], “environment” is operationalized only by stressful situations, such as childhood maltreatment, mainly by the family[92]. These conceptual shortcomings can also be observed in RDoC as will be explained later. Interestingly, it neglects a very important epistemological distinction that emerged more than 100 years ago in the context of the history of biology (and ecology/environmental science): (1) The objective environment as seen by the remote researcher; and (2) The subjective environment as experienced by the immediate living being.

If we briefly concede at this point that “objective” is shorthand for the sophisticated scientific observation reported through data and “subjective” is the phenomenal appearance reported verbally, we think that this categorical dichotomy is at least workable in practical psychiatry, although “objective” is not an epistemologically grounded category from a constructivist point of view, since scientists are also subjects and can only represent a methodologically grounded “intersubjectivity”. From a pragmatic point of view and for a clinical psychiatrist, a very important challenge in the diagnostic phase is to compare intersubjectively, together with the patient, the objectively observable (and possibly data-based) situation with the subjectively experienced situation, in order to identify a paranoid processing of the surrounding world.

Interestingly, a pragmatic dichotomy of objective/subjective is also important in human biology as it partly reflects a difference between natural science (molecular biology) and sciences of the mind or phenomenal self-awareness (phenomenology). This problem of a double methodology corresponds to the philosophical qualia problem beautifully formulated by Thomas Nagel: “What is it like to be a bat?”[93]. This question deserves closer consideration here.

Biology and ecology – the physical environment and the sensorimotor loop

In ecology as the science of the relationships between environment and organism, environment is a crucial concept. Consequently, it must to be emphasized that Ernst Haeckel, the founder of academic ecology, defined environment as the “surrounding external world” (or “the surrounding external space”) of an organism, referring mainly to the physical-chemical environment[94]. From a philosophical point of view, this position is associated with a positivist, empiricist, objectivistic, reality-based view. In contrast, but independently, Jakob von Uexküll’s environmental theory (“Umweltlehre”) focused on the informational realm, namely the perceptual and effectual world surrounding an animal, and emphasized that organisms only selectively receive environmental stimuli, a subset which he called the “Merkwelt”, and that organisms can also only selectively act on the environment, a subset which he called the “Wirkwelt”[95]. With his analysis, Uexküll elaborated the importance of subject-centered attribution of meaning to the external world and so is thus considered the founder of biosemiotics. In this respect, this view was also called “phenomenological biology” or “subjective biology”, which means that the term environment only makes a functional analytical sense, when it is related to information processing properties of the respective living being. In line with this view, Uexküll completed the picture by connecting “Merkwelt” and “Wirkwelt” through the model of a “functional circle” constituting organism-environment interactions. This kind of concept emerges in phenomenology and ecological psychology reducing the gap between the “outside” and the “inside”, leading to a kind of circular psychology by emphasizing the operational closedness of information processing and generation.

Psychology – the informational environment /world

The aforementioned dichotomies of “internal” and the “external” or “subjective” and

theoretical consideration.
Tretter F et al. Ecology of the person

the “objective” in early ecological psychology (or psychological ecology) were essentially reflected in critical controversies between Kurt Lewin[96] and Egon Brunswik[96,97]. Brunswik[97] emphasized the external world as it is perceptually screened by ecological sampling, while Lewin[96] focused on the topological structure of internal representations of the world that he called “habitat”. Certain subsets of the habitat exert a field-like “valence” on the person, but barriers may oppose this appetitive disposition and behavior. Later, James Gibson[98,99] proposed a hybrid conception that assumes that perception of the world depends on action, and action depends on perception of the world, a circular model similar to Uexküll’s functional cycle[98-100]. In his dynamic view, the perceived image of an object is the still frame of a movie created by movements (e.g., of the eyes and/or the person). The cues of the physical world as its significant stimuli during movement are vectors that are processed by the sensory system, looking for invariants in the low of stimuli, resulting in images such as “gestalts” as discussed by Gestalt psychologists. Gibson[98,99] also understood the external visual world as a “stimulus array” that acts on the human subject as “affordances” by and for intentional actions of the person: The structure of a chair becomes visible through the by exploration of this object and then it “offers” the possibility to sit down.

Gibson’s work can serve as a reference for new phenomenology that emphasizes the intertwining of the individual with the world as proposed in the anthropological concept of the person as an embodied, embedded, extended, and lived mindful being[101,102]. Interestingly, the conceptual bridge from the physiological sensorimotor level to social psychology has a long tradition of discussion in the context of ecological psychology[103].

Sociology – the textual and communicative world as the social environment of actors?

The “social environment” is only implicitly touched in RDoC through the category of “social function”, a domain exemplified by social dominance, affiliation/separation, identification of facial expressions, self, etc. It should be measured by self-disclosure and behavioral analysis but also at the genetic, molecular, cellular, circuit, and physiological levels. The social conditions of mental states and mental processing-family integration, socioeconomic class, educational level, social support, etc. are important, however, because social functioning requires a certain level of social integration and vice versa[104]. In addition, urbanization, migration, neighborhoods, unemployment, group membership, ethnic diversification, gender issues, etc. are also issues that are rarely explicitly considered by the biological perspective of RDoC.

Obviously, these categories are not essential in RDoC, and it seems that social theories are not covered by RDoC. For example, social anomie, lack of social resonance, and similar aspects of social order and social relations are not explicitly included for study. By construction, RDoC seems to underestimate the top-down causality (and reciprocal bottom-up causality) of individual behavior through social rules[105], the influence of social structure on action[106,107], etc., which are at least mediator or even moderator variables of mental health.

But what does “social” mean? It is important to note that in modern sociological theory, the social as an epistemic object is more than the population; it is a set of operations (actions) and rules that relate to interpersonal and interinstitutional domains[107]. “Social” also means a culturally specific definition of norms, values, beliefs, etc. by which someone functions. In turn “functioning” means performing a social role, especially in the context of organizing a social system, institution, social space, or behavioral setting (see below). In the context of psychiatry, the term “social” in most cases refers to the micro-level of interpersonal interaction and communication as the object of knowledge, while on a mesosocial level, groups and organizations are the object of social scientific investigation that finally, on a macro-level, also analyzes society and its social subsystems such as science, law, religion, economy, state, politics, etc. In other words, one must admit that individuals exist in the context of multiple social systems (e.g., economics, law, science) and are confronted with institutions through rules of action: To pay or not to pay, to be right or wrong, to be true or false in relation to socially defined “reality”, etc. The consecutive relations between the individual and the social environment mean integration or exclusion: If I cannot pay I cannot participate, if I am wrong I must accept punishment otherwise I will be excluded by imprisonment, etc. A society based on the rule that everyone is responsible for his own destiny would not organize public support systems for marginalized people (e.g., the chronically mentally ill), who more and more exhibit a multiple downward spiral to homelessness. In general, the top-down modulation of
individual behavior by macrosocial concerns is seen as a systemic colonization of the individual lifeworld[108].

As a methodological consequence, the social domain can be explored analytically but only on a multivariate correlational basis and not seriously on a causal level. In accordance with this methodological limitation, social neuroscience should compare electroencephalograms and other biomarkers of subjects belonging to society with a strongly state-controlled economic order with the question: Is there a correlated difference (covariance/variance analysis)? And if so, can it be ruled out that there are other factors such as biological or psychological (and micro-social) variables that cause this difference?

There is another more profound methodological problem in social science that has already been addressed: Categories of the social can be measured objectively to some extent, as empirical social science does, but for behavior the subjective side is more important: How do I experience the world? This subject-centered view was already fundamentally emphasized by the philosopher Husserl[109] who constructed the subject-centered concept of a world view he called “lifeworld”[109], which was further developed especially by the Austrian sociologists Schütz and Luckmann[110–111], Berger and Luckmann[112], and also Habermas[113]. Another influential approach in sociology, namely the work of Bourdieu[114–116], must be considered. Bourdieu[114–116] conducted empirical studies in Algeria and developed new theoretical categories to describe the context of social behavior (e.g., concepts such as “social space” or “social fields”). Interestingly, Bourdieu[114–116] acknowledges an objective social reality and not only its constructedness by the subject and society, as constructivists often see it. In relation to psychiatry, cultural psychiatrists also noted, that “culture is in the head/brain/mind and in the environment”[117]. The “social” in psychiatry is a social reality.

However, there is another side to the “social” in psychiatry: practical clinical work requires close collaboration with social workers to optimize service to patients. Consequently, knowing how social workers think, i.e. the theoretical background of social work, is useful. Social work theory is mainly based on aforementioned concept of the “lifeworld” model derived by Husserl[109,118,119] and is often associated with Bronfenbrenner’s[120] approach to psychology[120]. This is another good reason to differentiate the concept of “environment” in the context of psychiatry.

Towards an integrative taxonomy of the term “environment”

Looking back at this section on the different aspects of the concept of social environment and in terms of its essential importance for mental health, the distinction between the physical environment facing climate change and the digital environment with its virtuality and the machine-based technical environment (e.g., robots) should also be made clear. Moreover, the “natural” environment (and its hazards) may be an interesting field for psychiatric research. In other words, an explicit taxonomy of the term “environment” is useful, similar to that used in ecological (and environmental) psychology. These semantic distinctions imply a holistic concept of environment that also covers the heterogeneity of topics mentioned earlier. Interestingly, almost no biological psychiatrist refers to ecology, which is a biological field of research and had a strong influence on sociology and (psychiatric) epidemiology in the first half of the 20th century. Perhaps in the context of studies of “urbanity” and mental health, there is a revival of the ecological view[121]. With the aim integrating these different meanings of the term environment, we present an outline of a comprehensive conceptual framework for a more precise use of this term, which is already useful in the context of ecology and sustainability science.

First, it is important to recognize that the general term environment also includes semantically similar terms in the scientific context, such as environment, “context”, “setting”, “behavioral setting”, “social space”, etc. These terms have specific meanings, but they share a common semantic stem, namely the “external” (or “surrounding”) world. From a systemic perspective, the external world is a system of elements that are related to each other and to the living being as observer and actor. These elements are often referred to as “factors”, “dimensions”, “areas”, “levels”, etc. However, there is no universally accepted convention on the taxonomy of these concepts. Moreover, there are no interdisciplinary accepted definitions and/or conceptual explanations of these terms. Here are some semantic diversifications that can help clarify meaning when “environment” is used in interdisciplinary communication[122].

Basically, it has to be taken into account that the distinction between system and its environment is a (physical) distinction with a more or less concrete boundary structure (skin, membrane, entry rules of a social system) that constitutes the system. At the very least, the observer of the system must construct a conceptual distinction. This
issue has been widely discussed in systemic family therapy\[123\] and sociology\[124\]: A system constitutes itself by constituting and/or enforcing or extending its boundary structures.

Consequently, the categorical and real distinction between system and environment implies the notion of a “relation” (e.g., a boundary relation such as inclusion or exclusion mechanisms or rules) between a social system and its environment. From a cell biological point of view, this difference is established up by the membrane with its ion channels and its gating and ion pumps that “establish” the difference between the external and internal environments. Psychologically, objects of the environment have an attraction, a valence, or an affordance (or a repulsion) for the person. This disposition is often meant when one uses the very general term “relation”. However, this term deserves further analytical consideration, which can only be touched upon in the final section of this paper.

If we summarize these semantic aspects of the term “environment” in a preliminary taxonomy, we can highlight some points (Figure 3): (1) The distinction between “objective environment” and “subjective environment” is necessary because it has already been figured out (e.g., measured and/or experienced air temperature). It is an everyday experience of any self-critical person acting in a social context. The history of scientific discussion in psychology and sociology also shows that this distinction is irresolvable. Philosophers distinguish the objective “third person perspective” and the subjective “first person perspective” and emphasize that the subjective experience precedes the subjective view of the external world; (2) The temporal dimension is evident in the heuristic usability of the concept of “past”, “present”, or “future” environment. The developmental ecological psychologist Bronfenbrenner\[120\] proposed the concept of a chronosystem representing the significant ecological transitions when a child becomes an adolescent, an adolescent becomes an adult, a student becomes a researcher, etc.; (3) As for the objective side, the external environment is primarily a spatio-temporal world, that can be described in terms of physics and other natural sciences, but perceptual Gestalt psychology has shown for decades that the experienced world is also the result of constructive processes of the subject and the society. Construction is not possible without elements, so for the construction of an image (Gestalt) of an object of the world, the constitution of relations of its elements is essential. In line with this spatial dimension and in a systemic view, the distinction between a micro-level (family, work, and school), a meso-level (community), and a macro-level (society) has been influentially emphasized by Bronfenbrenner\[121\]. The microsystem is primarily understood to be the family, but the school or workplace, the recreational sphere, and other sub-areas of individual’s living space are also addressed by this level. The meso-environment (or meso-system) is the summary of these micro-environments. The meso-level is embedded in the macro-environment of the society, which is referred to as the macro-system. Bronfenbrenner also introduced the notion of the exo-system, which influences the person but that in turn cannot be influenced by the individual; (4) Entities: With this term we propose to distinguish between material-energetic or physical vs immaterial (informational) environment. In addition, scientific ecology distinguishes other sub-entities, such as inanimate natural (abiotic; soil, water, air), animate natural (biotic; microbes, plants, animals), technical, personal, social, cultural, etc., environment. The everyday understanding of these terms is sufficient here, but a precise definition remains difficult, especially since there is currently no philosophical ontology (see above); (5) Spheres of life: Lifeworld, work world, etc., are environmental domains that are distinguished according to their functional importance for people. They are separate but partially overlapping spheres of life that are concerned with these issues (home office); (6) Qualities: Environmental characteristics can be classified as “good” or “bad” in relation to the environmental effects experienced by the person (attraction, repulsion); (7) Quantities: These environmental properties can be represented, for example, as the density of the respective environmental elements occurring in the space-time framework. A stimulus-rich world may trigger stress in a busy person but may also be perceived as pleasure by a bored individual; and (8) Effects and their directionality: Impacts, effects, and interactions are corresponding categories.

This taxonomy should not be understood as a rigid framework but as a reference for important specifications and distinctions in interdisciplinary communication; the communicator should define what he or she understands by the chosen term.

This taxonomy can also serve as a guide for the term relation. Although in the context of psychiatry one usually means “social relationships” by this term, relationships can be observed objectively/subjectively at certain times, at different levels (micro/macro), in terms of physical or interpersonal dimensions or domains, in
CONTOURS OF A_NEW_THEORETICAL_FRAMEWORK_–_“ECOLOGICAL PSYCHIATRY”

We now we try to put together the things that have been distinguished before and propose a sketch of an integrative conceptual framework of a human ecological perspective, which has some cross-disciplinary traditional roots but is not yet really established in psychiatry and is also only rudimentary in medicine: (1) The first section has shown that psychiatry lacks an area that can be called a “theoretical psychiatry” and that the methodological plurality needs knowledge integration as it was proposed by the bio-psycho-social model. It is believed that the best basis for a general theory of psychiatry could be the modern psychoanalytic theory, especially since it is currently elaborating its roots in neuroscience[125-127]. Moreover, the last fundamental theoretical developments in psychoanalysis that led to in modern object relations theory could be a fruitful basis, also for an ecological perspective; (2) The second chapter on addictions showed the usefulness of the behaviorist approach but also highlighted the heuristic relevance of psychoanalytic thinking and modeling, where “relationships” are the key issues, even if its therapeutic relevance sometimes only comes into play at a later stage of treatment. Using the clinical example of adolescent drug addicts, it was emphasized that the family, as the most important social environment, has always been a constitutive theme of psychoanalysis. Thus, psychoanalysis per se has always been a bio-psycho-social approach; (3) In the third section, the semantic complexity of the term “environment” was pointed out. This chapter also pointed out the theoretical disparity of the “environment” when considered as a set of stimuli and/or as the family; and (4) Taken together, these issues imply a conceptual framework based on a differentiated but integrated model of human-environment relations as a system that is provided by the human ecological approach. But what exactly is “ecology”, and what are constitutive concepts?

The human ecological framework in psychiatry – knowledge and theory integration

As mentioned earlier, ecology as a biological science was founded by Haeckel[94] in
1866 and was defined as the science of the relationships between organisms and their surrounding environment [95]. For a time in the 20th century, this approach, which aims to understand human living systems, had a major influence on disciplines such as sociology, geography, psychology, epidemiology, etc. Several very general calls for a systemic ecological perspective have also been published in psychiatry [128-130]. In a similar vein and in relation to clinical issues of addiction, an ecosystemic multilevel perspective was proposed in the 1990s under the name “Ecology of Addiction” [92]. Following this, a generalization as “ecology of the person” [131] has been published and recently further elaborated [122]. This corresponds with the call for a “neuroecology” [132,133] or even for an “ecology of the brain” [134].

Rooted in cultural psychiatry, Kirmayer and Crafa [21] recently proposed an “ecosocial psychiatry” [21]. Emphasizing the need for a “new science of psychiatry” he argued: “A multilevel, ecosocial approach to biobehavioral systems is needed both to guide relevant neuroscience research and to ensure the inclusion of social processes that may be fundamentally contribute to psychopathology and recovery” [135]. This approach has been used as a basis for the development of a new science of psychiatry.

As a result, we propose here a person-centered and environment-centered view in psychiatry that considers the patient as a “situated subject”. We also propose a multilevel systemic view of mental disorders that focuses on the mind as an information processing subsystem. This systemic ecopsychiatric perspective appears as the logical integration of the current counterproductive divergences between research and practice.

In this sense, and in relation to addiction, a first step of theoretical integration seems to be possible by linking the psychoanalytic model not only to CBT, but also to the ecological psychology of Lewin, Barker, Gibson, and Bronfenbrenner, by linking it to the “ecology of the mind” [136] and the phenomenology of the embodied, embedded, expanded, and performed subject [37]. These efforts could lead to an “ecology of the person” that offers more coherent options for understanding disorders and planning their treatment and prevention [131].

With this in mind, we want to move toward a human ecology view of mental disorders based on the previous sections of this paper: We will propose three conceptual building blocks – the anthropological concept of the person, the ecological concept of the environment, and the analytical concept of the term “relationship”.

The person
The person can be described in a multidimensional framework, for example, when it comes to the clinical context [122]. In clinical psychiatry, about eight basic dimensions are usually checked: Time (When were you born?), space (Where do you live?), physical state (Are you physically well?), mental state (How are your memory functions?), linguistic competence (How can you describe your feelings?), social domain (Do you have a relationship?), cultural aspects (Religious belief?), and economic status (Are you employed?). This information provides a treatment-relevant framework of the situated person and is not theoretically integrated.

The “environment”
The complex meaning of this term has already been discussed at length. Related to basic clinical practice, as highlighted above, five of the above questions deal with the person’s environment: The temporal, the spatial, the social, the cultural, and the economic. These are some essential dimensions of the surrounding world that determine the manner and quality of the persons’ existence, with implications for health status.

The relations
The categorical distinction between the person and her environment is counter-balanced by the concept of the relations: A relation is basically conceived here as a disposition between two elements that is recognized by an observer, such as a doctor in a clinic or the subject or the researcher: What are the subject’s relations with his or her parents? In this construct, too, the first essential distinction is the epistemological one, between the “objective” relationship as seen by the expert and the “subjective” relationship as seen by the patient. Although there are varying degrees of realism within the views, which may even be paranoid in nature, these dichotomous distinctions are pragmatically useful here. The relations of the internal and external psychosocial environmental relations between the person and their environment in terms of affective-motivational values must be optimized in therapy. Consequently, the occurrence of objects (persons) in the environment that are similarly structured to
the reference objects of the representation evokes old patterns of cognitive-affective processing.

Although a similarly detailed analysis of the concept of relations as shown in this paper for the concept of “environment” is necessary, the issue of relations is already addressed essentially in object relations theory [80,138,139]. It is the affective valence of relations (or the quality of affects) that causes the appearance of environmental objects. It is, of course, co-determined by the sociocultural context. The system of relations as a “structure” of the representational system of the object and the subject and their interrelations can be distinguished in a simple but more systematic way: The relations of the person to the environment can be classified as processes and thus with “fugal” and “petal” directionality: The giving and the taking. In the abstract mathematically oriented terminology, the taking is often preceded by a demanding, and if a giving is not accepted it is a refusing. The content of the relationships, the entity transferred, for example, information, matter/energy or love, support, etc., is central (Table 1). These framing concepts have not yet been translated in this form into the psychoanalytic terminology. Interestingly, they have some roots in the social psychological school of exchange structuralism [140,141]. Although developed in the context of economic studies, they can also be used to describe interpersonal interaction within the family: The father demands performance from the child and thus has a demanding image for the child, the mother supports the child and has a supportive image for the child. As a result, the polarized image of the “bad father” and the “good mother” emerges and solidifies as object representations with their respective emotional charges. In other words, the content of the actions or the (verbal) signals of the relationship can be categorized as acknowledgement, care, support, commitment, etc. Thus, intrafamily interaction and the structure as the essence of interactions can be described in a naturalistic way through ordinary language that can be easily documented in psychotherapy sessions (Table 1). The psychopathological relevance is that in every interaction, not only an intertemporal balance of the integral of interpersonal give and take but also of intrapersonal give and take must be established to produce harmony and avoid conflict, which in any mental disorder are at least accelerators of pathogenetic mechanisms.

This notion of balance and imbalance of interactions is in turn linked to the well-known and highly integrative and heuristically fruitful concept of stress.

**Stress as a central construct – an ecological conceptualization**

A relational theoretical framework for stress theory can be reconstructed through an imbalance of four relations between needs (desires) and competencies of the person and offers and barriers of the environment: (1) The person’s needs (desires, demands) in relation to their (2) competence to satisfy these needs depends on the (3) environment’s offers in relation to these needs, and in relation to the (4) environment’s barriers. For example, related to the person: I want roses in my room (desire), because roses are beautiful (offers), but they have thorns (barriers). If I have a tool (competence), I can cut them and put them in a vase at home.

Although this concept could be elaborated further, this example is meant to illustrate the ecopsychological perspective. A relational stress concept is also appropriate at the biological level: The probability of an infectious disease is determined by the relationship between the infectivity of the pathogen agent and the vulnerability of the host and their respective embeddedness in their environments.

Seen in this light, the fruitful bio-psycho-social conception of the stress paradigm can be combined with an ecological perspective, including for a relational/ecological understanding of addiction.

**CONCLUSION**

For a comprehensive picture of mental health, the constitution of a “systemic ecology of the person” is crucial. Within this theoretical project, the interdisciplinary reconstruction of the concept of “relations” seems to be essential. The focus is on the quality of the relation between the person and the (social) environment, as they are conceptualized by various theoretical approaches/constructs such as “vulnerability”, “resilience”, “salutogenesis”, “risk factors”, “protective factors”, “affordances”, “asymmetries”, “barriers”, “incongruences”, “attachment”, etc. Psychoanalysis, especially object relations theory, has a rich repertoire of descriptions of these relational terms [140]. Finally, the relationship between the subjective perspective (“I am observed by secret service”) and the objective perspective on the world (“there is
Table 1 Exchange forms and contents—some basic examples

<table>
<thead>
<tr>
<th>Give</th>
<th>Content</th>
<th>Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>Give out</td>
<td>Rejection</td>
<td>Accept</td>
</tr>
<tr>
<td>Spend</td>
<td>Money, time</td>
<td>Receive, consume</td>
</tr>
<tr>
<td>Give away, devote to</td>
<td>Burden</td>
<td>Accept</td>
</tr>
<tr>
<td>Offer</td>
<td>Care</td>
<td>Demand</td>
</tr>
<tr>
<td>Give</td>
<td>Money</td>
<td>Accept</td>
</tr>
<tr>
<td>Give away</td>
<td>Furniture</td>
<td>Accept</td>
</tr>
<tr>
<td>Send</td>
<td>Information</td>
<td>Receive</td>
</tr>
</tbody>
</table>

very likely no interest of secret service”) is a crucial issue for psychiatrists and psychotherapists to work on.

It is important to conclude this paper with the position of Kirmayer and Crafa[21]: “This ecosocial view of mind, brain, and culture calls for a shift in perspective from a psychiatry centered on brain circuitry and disorders toward one that recognizes social predicaments as the central focus of clinical concern and social systems or networks as a crucial site for explanation and intervention.”

ACKNOWLEDGEMENTS
We thank Dagmar Steinmair for her help in editing.

REFERENCES

1 Kuhn MH. Field theory in social science: Selected theoretical papers by Kurt Lewin. AAPSS 1951; 276: 146-147 [DOI: 10.1177/000271625127600135]
Debates held during the ICTALS 2019 conference.


Koob GF, Le Moal M. Drug addiction, dysregulation of reward, and allostatics. Neuropharmacology 2001; 40: 97-129 [PMID: 11120394 DOI: 10.1006/s0893-133x(00)00195-0]


51 Popper KR. Objective knowledge: An evolutionary approach. New York: Oxford University Press, 1972


73 Kanfer FH, Goldstein AP. Helping people change. New York: Pergamon Press, 1975


81 Ogden TH. The concept of internal object relations. *Int J Psychoanal* 1983; 64 (Pt 2): 227-241 [PMID: 6347926]
94 Haecckel E. Generelle Morphologie der Organismen. 1th ed. Berlin: Reimer, 1866
95 von Uexküll J. Umwelt und Innenwelt der Tiere. (Environment and internal world of animals). Berlin Heidelberg: Springer, 1921
98 Gibson JJ. The senses considered as perceptual systems. Boston, MA: Houghton Mifflin, 1966
113 Habermas J. Theory of communicative action, Volume 2: Lifeworld and system: A critique of


Fuchs T. Ecology of the Brain. Oxford: Oxford University Press, 2018


Foa UG, Foa EB. Societal structures of the mind. Springfield, IL: Charles C Thomas, 1974

Therapeutic termination of pregnancy and women’s mental health: Determinants and consequences

Ester di Giacomo, Rodolfo Pessina, Mario Santorelli, Daniele Rucco, Valeria Placenti, Francesca Aliberti, Fabrizia Colmegna, Massimo Clerici

ORCID number: Ester di Giacomo 0000-0001-5433-1268; Rodolfo Pessina 0000-0001-5561-2037; Mario Santorelli 0000-0002-3637-2354; Daniele Rucco 0000-0002-9922-9774; Valeria Placenti 0000-0001-7847-6760; Francesca Aliberti 0000-0002-6305-0406; Fabrizia Colmegna 0000-0002-9984-9908; Massimo Clerici 0000-0001-8769-6474.

Author contributions: Pessina R and di Giacomo E wrote the paper; Rucco D, Colmegna C, Santorelli M, Placenti V, Aliberti F and Clerici M collected data; All authors analyzed papers and approved the final version of the manuscript.

Conflict-of-interest statement: The authors declare that there are no conflicts of interest.

Country/Territory of origin: Italy

Specialty type: Psychiatry

Provenance and peer review: Unsolicited article; Externally peer reviewed

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

Abstract

The therapeutic termination of pregnancy (TToP) is an induced abortion following a diagnosis of medical necessity. TToP is applied to avoid the risk of substantial harm to the mother or in cases of fetal unviability. This type of induced abortion is provided after the second semester of gestation if fetal illness or the pregnancy cause physical danger or pathological mental distress to the mother. Socio-cultural and economic determinants could influence the desire for children and family planning in couples, as well as the use of effective contraception and the choice to perform an induced abortion. Furthermore, the TToP is a reproductive event with an important traumatic burden, but also with an intrinsic therapeutic effect and it can produce different psychological and psychopathological effects on women and couples. The aim of this review is to evaluate what demographic, reproductive and psychopathological determinants are involved in the choice of undergoing a TToP in women. Also, we will examine both positive and negative consequences of this procedure on women’s mental health, underlying which factors are related to a worse outcome in order to provide the best clinical support to vulnerable groups.
Core Tip: The choice of having a therapeutic termination of pregnancy (TToP) is strictly personal and several demographic and reproductive factors may contribute. Women who have a TToP are more commonly affected by a previous mental disorder, which influences their mental outcomes after it. There is a higher prevalence of any axis I psychiatric disorder in women who choose for a TToP, but the correlation with personality disorders has been less examined. TToP seems to produce both positive and negative psychopathological effects on women and only a minority of women show psychiatric disorders after it, especially those with a previous psychiatric history.

INTRODUCTION

The therapeutic termination of pregnancy (TToP) is an induced abortion following a diagnosis of medical necessity. TToP is carried out to avoid the risk of substantial harm to the mother or in cases of fetal unviability. The TToP, according to the Italian Law 194/78[1] and different legislations around the world[2], is provided after 90 d of gestation if fetal illness or the pregnancy may cause physical danger or pathological mental distress to the mother. Unlike voluntary termination of pregnancy (VToP), which is performed on women’s request, TToP requires a clinical evaluation by a physician.

The choice of undergoing a termination of pregnancy (both voluntary and therapeutic) is strictly personal and complex for the woman and the couple since numerous factors are involved in the decision-making process.

Furthermore, TToP is an obstetric event with an important traumatic burden, but also with an intrinsic therapeutic effect. This review aims at evaluating the demographic, reproductive, and psychopathological determinants and the psychiatric consequences of TToP in women and couples.

DEMOGRAPHIC AND REPRODUCTIVE DETERMINANTS OF TTOP

Several socio-cultural and economic determinants influence the desire for children and family planning in couples, as well as the use of effective contraception or the choice to have abortion in case of unwanted pregnancy or maternal physical or pathological mental distress. Many studies do not discriminate between therapeutic and voluntary ToP when analyzing possible demographic and reproductive determinants, thus introducing a bias in understanding decisional pathway of TToP.

Maternal age

In Western countries[3,4] the rate of TToP is higher in women aged between 20 and 34 years old. That timeframe correspond to the period of greatest female fertility rate, when it is more likely a woman gets pregnant and may wish to have an abortion in case of newborn illness or unwanted childbearing.

Instead, the relative rate of TToP compared to the total number of pregnancies by age group appears to be doubled in women aged under 20 years and over 34 years[3,4]. In fact, in extreme maternal ages, there’s a lower female fertility rate, but also an increased risk of genetic fetal diseases or congenital malformations, which could explain the higher risk of TToP.
Different but not conflicting results were observed in developing countries, such as Ethiopia[5] or India[6].

In the Third World, maternal age at first pregnancy is significantly lower than in Western countries and most of the female population requiring abortion is under the age of 34 (89.5%)[5]. Furthermore, most of the voluntary terminations of pregnancy are performed in women aged 20-24(34%) and 15-19 (24%)[5,6].

Several studies[3,7,8] show that adolescent women have abortion later in pregnancy than other women: this could be due to late pregnancy discovery and existing social, economic, and legal difficulties linked to that group.

**Marital status**

In Western countries, the frequency of TToP is 3 times higher in women who are divorced than in unmarried and 5 times higher than in married or widowed[3,9].

The choice to undergo an abortion seems to be strictly conditioned by the absence of a partner, a close marital bond, or a supportive family unit, also because divorced or unmarried women seem to be more inclined to unwanted pregnancy, VToP and TToP [5,9].

The couple relationship is often an important decisional determinant of TToP in women; as demonstrated by Chibber et al[10] the three most common partner-related reasons for having a TToP are: (1) A poor relationship between the two members of the couple; (2) Male partners unable or unwilling to support a child; and (3) Partner’s personal traits that make them unsuitable to be a father.

In developing countries, the majority of women who decide to have an induced abortion are married (46.4%) or single (41.6%)[5], probably because in many of these countries single-parent families and divorce are not culturally accepted, despite being legal.

**Women’s level of education**

Several studies[3,7] from the Mediterranean area show a direct correlation between women’s level of education and termination of pregnancy: women with a high level of education (secondary or university degree) show a rate of TToP 10-15 times higher than subjects without any school qualification.

However, in the Northern Europe[8,9] and Australia[11], higher rates of TToP were observed in women with a lower cultural level, which is more often associated with lower wages and a lower socio-economic level.

In developing countries, such as Ethiopia[5], the majority of abortions are performed on illiterate women (30%) or with education under 8 years (23%), while only 9% of TToP are performed on women with a diploma or university degree.

In India[6], instead, women with higher levels of education show a lower fertility index, because education strongly influences women’s attitudes towards pregnancy and family size, leading to a decline in the desire of numerous offspring. At the same time, educated Indian women tend to avoid sterilization practices and have an increased risk of unwanted pregnancy for more liberal sexual behavior. Additionally, educated women are more likely to accept reproductive control and induced abortion than poorly educated or illiterate women.

**Women’s employment**

The rate of female workforce affects the demand of induced abortion in a country, because working women, overall, have more control over their family planning and reproductive lives thanks to their economic contribution[6]. In fact, in Western countries, employed women have an absolute frequency of TToP almost two times higher than housewives and students. On the other hand, considering the relative frequency of induced abortion with respect to the fertility index, students show a double rate of abortion compared to both employed women and housewives[3,9,12]. This huge gap is due to the fact that female students do not have their own income and this leads them to decide they do not carry on an unwanted pregnancy.

Women with the lowest monthly income are those with the greatest use of abortion: in the United States[13] women with an income below 100% of the federal poverty level (FPL) have an abortion rate 5-6 times higher than women with a double income. A similar correlation was also observed in other European countries[7,9,14].

In developing countries[5,6], the use of abortion is greater in housewives (37%) or schoolgirls (27%), while only a minority of women requesting TToP are employed. Moreover, women with the lowest monthly income (< $ 26.3 per month) are the majority of requesting induced abortion (78%). Importantly, in those countries the effect of female occupation on reproductive behavior depends on women’s kind of job. Women employed in agricultural or industrial settings show a lower rate of induced
abortion, due to their possibility of carrying on pregnancy and raising children[6].

Number of children and family planning
There is a direct correlation between the number of children in the family and termination of pregnancy; women with several offspring (2 or more children) show an abortion rate 5 times higher than women who have no children or only one[3,12]. Furthermore, family planning (assessed as the woman's referral to a family planning clinic in the previous months) seems to be inversely related to the choice of having a TToP[12]. Uria and Mosquera[3] showed that 60% of women with an induced abortion did not refer to a family planning clinic in the previous 2 years. A similar rate of unplanned pregnancies ending in TToP or VToP was observed in developing countries[3].

Interestingly, the percentage of adolescent women who have an induced abortion without family planning rises to about 80% and this highlights the need for sex education programs to prevent unwanted pregnancy and other sexual issues in that fragile population[3].

Urbanization rate
In developing countries[5,6] and Australia[11], the frequency of TToP seems to be higher in urban than in rural areas, considering the widespread availability of health services in towns.

Possible reasons for correlation between urbanization rate and induced abortion are well explained by Bose et al[6] in their study from India. In India, urban residency appears to promote an anti-natal lifestyle and increased cost of having children. Children in urban areas are more difficult to supervise and are less economically productive than in rural places since they cannot be employed in agricultural work. This leads women from urban areas to wish for less children. Furthermore, the social norms that promote small family sizes and the government’s family planning messages are more easily spread in cities thanks to a better access to the media (TV, radio). Thus, women living in large cities are more likely to perform a termination of pregnancy not only for their desire for smaller families but also for an increased knowledge and access to legal abortion services.

Immigration
Rasch et al[9] showed that immigrant women are a vulnerable group with an increased risk of unwanted pregnancy and induced abortion due to their social and economic condition. Nevertheless, in a recent reanalysis of data from Mortensen[15], authors showed that immigrant women had a 2 times increased risk of induced abortion compared to non-immigrant women with the same monthly income. In addition to these findings, the study by Pérez et al[16] has shown that, in Spain, termination of pregnancy is more common in areas with a huger concentration of immigrants from outside the European Union. The authors suggest that the status of “immigrant citizen” is an independent variable that must be taken into account in evaluation why a woman may choose for abortion[15].

Women’s ethnic group
In the United States, women belonging to ethnic minorities may experience an increased need and reduced access to health services for induced abortion than white women, being thus exposed to negative medical and social consequences[13,17-21]. The disparities in abortion rates are well documented, with rates of 1.1% in non-Hispanic white women, 2.8% in Hispanic women, and 5% in black women. Furthermore, although pregnancy are unplanned in 51% of all American women, the percentage rises to 70% in African American women and 57% in Latin American ones, compared to 42% in white women[13,17]. Those differences are confirmed in multivariate analysis, net of age, and monthly income, suggesting that ethnicity is an independent risk factor for having a termination of pregnancy.

Data showed that Hispanic and Black women are also more likely to have termination of pregnancy at a later gestation age (from second trimester onwards) than white women. Authors justified[13,17] an increased abortion rate in ethnic minorities in the United States with lower socio-economic status, higher level of stress and precarious working/housing conditions, less common use of effective contraceptive devices, lack of health insurance coverage, and the absence, in many American counties, of accredited health care facilities for termination of pregnancy.

Although studies[17,19,20] have focused on African and Latin American women, a higher abortion rate has also been described in other minority ethnic groups such as
Religious beliefs
A strong religious belief seems correlated to a lower access to medical abortion in the general population. For example, the study by Pérez et al.\cite{16} showed that, in Spain, regions where a large part of the population is a practicing Catholic have a reduced access to gynecological services for abortion. Likewise, a study from Iran demonstrated an increased recourse to abortion in women with lower levels of Islamic religiosity and spirituality\cite{22}.

The fear of being judged unworthy by their religious community, as described in Poland by Zaręba et al.\cite{23}, often leads women with strong religious beliefs to have abortion in health centers far from their hometown and avoid to confess it to their ministers. However, an organized religious activity seems to allow women to carry on with pregnancy after the discovery of fetal malformations having a better psychological outcome, compared to those who decided to have a TToP\cite{24}.

Gender and intimate partner violence
Taft et al.\cite{11} showed that women who have experienced or are currently experiencing any form of gender-based violence (psychological, physical, or sexual) are at higher risk of TToP or VToP. Women with a recent episode of non-domestic gender violence reported a significant risk of having an induced abortion (OR = 1.46 for adolescent women and 1.85 for adults).

Women who reported at least one experience of intimate partner violence (IPV) in their life have a 3 times higher risk of having a termination of pregnancy. The risk for an induced abortion seems higher in women with a recent episode of physical or sexual violence perpetrated by their partner.

Since the association between any form of gender violence, miscarriage, and preterm birth\cite{11,25-27}, it is mandatory for clinicians to investigate such aspects in patients asking for induced abortion.

PSYCHOPATHOLOGICAL DETERMINANTS OF TTOP
Several studies explored possible adverse outcomes of abortion on women’s mental health, but, astonishingly, pre-existing mental health problems are often underestimated or neglected.

To evaluate the psychopathological determinants of choosing a termination of pregnancy is fundamental from an epidemiological and health economic points of view. Furthermore, the presence of pre-existing mental health issues could affect post-abortion mental health outcomes of those women\cite{25-30}.

There are already proofs that women who have had a TToP or a VToP may have higher level of psychopathology before abortion compared to other women who have never undergone one. Women who have a termination of pregnancy showed higher incidence rates of appointments with a psychiatrist or their general practitioner for mental issues, before and after abortion, than women who gave birth\cite{31,32}. Another research\cite{33} found that among women with psychiatric disorders and abortion, the majority of mental disorders occurred before ToP rather than after.

Termination of pregnancy in the female psychiatric population
van Ditzhuijzen et al.\cite{34} showed that the prevalence of axis I mental disorders was significantly higher in women who had a TToP or VToP (68.3\%) compared to those who had not (42.2\%). Compared to women who gave birth, those with an abortion (both VToP and TToP) have, after controlling for demographic factors, a 3 times higher risk of psychiatric history for any mental disorder\cite{34}. Furthermore, women who have a termination of pregnancy are more likely to have previous psychiatric history than the controls. Authors\cite{34} suggest two possible explanations: (1) An increased number of unwanted pregnancies in women with a history of mental disorders compared to healthy controls. This population may be more prone to other risky behaviors, such as unsatisfactory relationships, misusing of contraceptive devices, and substance abuse; and (2) A more frequent desire for TToP and VToP in women with psychiatric disorders. That wish could be mediated by factors related to their mental disorders and personality traits, such as low self-esteem, a pessimistic view about their children life or doubts about their parenting skills, antisocial behavior, aggression, delinquency, or criminal behavior. Women with mental health issues could show a higher frequency of socio-cultural and economic factors related to the choice of having an induced abortion.
abortion, such as a low income or indigence.

**Termination of pregnancy in women with anxiety disorders**

According to several studies[33,34] women who have an abortion more frequently show a previous history of any anxiety disorder than women who give birth.

Analyzing different anxiety disorders, women who have a termination of pregnancy report a double history of panic disorder (OR = 2.06), social phobia (OR = 1.94), or specific phobia (OR = 2.30) compared to women who decide to give birth[34]. They also show a 1.5-1.9 times higher risk for agoraphobia and generalized anxiety disorder [33,34].

Several authors[33,35] attributed the correlation between anxiety disorders and increased use of termination of pregnancy to women’s fear of gestation and childbirth (tocophobia), an aspect that has been described in women affected by anxiety, mood, and eating disorders.

According to Hofberg and Ward[35], although pregnancy is a normal and often desired condition, women may experience anxiety about physical and psychological consequences of pregnancy, sometimes with pathological connotations. Primary tocophobia[35], *i.e.*, not due to a previous negative gestational episode, appears to be due to 3 factors: (1) Social culture: It seems that the fear of gestation and childbirth can have an intergenerational transmission and, often, women’s attitude towards pregnancy is strongly conditioned by the experiences reported by their mothers and grandmothers; (2) Pre-gestational anxiety disorders; and (3) Previous trauma: A history of childhood sexual abuse appears associated with avoidance to gynecological examinations and obstetric care. Besides, the trauma of vaginal delivery (or even just contemplating it) can potentially cause the re-emergence of stressful emotions in women who have experienced a previous sexual abuse.

**Termination of pregnancy in women with affective disorders**

Mota *et al*[33] and van Ditzhuijzen *et al*[34] showed that women who have a termination of pregnancy show a 2 times higher risk of having a previous history of any mood disorder compared to those who give birth. In particular, those authors described a significantly increased history of major depression (OR = 1.51), dysthymia (OR = 1.79), and bipolar I disorder (OR = 2.8).

Furthermore, women with mood disorders who have abortion had more frequently a previous history of suicidal ideation (OR = 1.59) and attempted suicide (OR = 1.51) [34]. According to Mota *et al*[33], that evidence underlines that women with more severe forms of mood disorder are also those with the highest motivation to terminate pregnancy, probably because they feel incompetence towards motherhood duties. Several studies[36-38] highlighted more induced abortion in women with major depression and bipolar disorder under medications compared with women with the same disorder but drug-free when they have conceived. The greater access to TToP could be explained by the concern of teratogenic damage due to psychotropic drugs.

The request for abortion in women with affective disorders is higher, especially in those taking fluoxetine and mood stabilizers (especially sodium valproate). Concerns are linked to the evidence of increased risk of fetal malformations and obstetric complications in women who take those medications during pregnancy, but clinical experience with women who are affected by epilepsy suggests that these consequences are rare and manageable with adequate antenatal planning[39,40].

It is crucial to understand the reasons that push women to ask for abortion, to provide them information on the real teratogenic risk of their therapy or to suggest alternative strategies to abortion, such as a switch to safer drugs and careful antenatal care with ultrasound monitoring.

**Termination of pregnancy in women with borderline personality disorders**

Emotional and behavioral dysregulation experienced by patients with borderline personality disorders (BPD) may expose them at increased risk of unwanted pregnancy, compared to women without those symptoms[41]. Several studies[41,42] suggest that they may have earlier and unprotected intercourses than healthy women or women with different mental illnesses, due to their personality traits such as unstable attachment, impulsive behaviors, and compromised decision-making processes.

Furthermore, women with BPD show a greater predisposition to sexually transmitted infections (STDS) and associated consequences, such as pelvic inflammatory disease and infertility[41]. Other factors that can contribute to increase the risk of unwanted pregnancies and termination of pregnancy in those women are the higher prevalence of sexual abuse and comorbidity with substance use disorder (SUDS).
De Genna et al[42] found that the severity of BPD symptoms is directly correlated with the risk of having an unwanted pregnancy in adolescence, even controlling for socio-cultural determinants; in fact, in their sample, up to 18% of women affected by BPD becomes pregnant by the age of 16.

The effect of BPD severity on unwanted pregnancies and termination of pregnancy appears to vary depending on comorbidity with SUDs: women with severe symptoms but without a history of SUDs have more unwanted pregnancies with full-term births, while women who SUD have more induced abortion.

According to these authors, the risk of miscarriages, VToP and TToP in women with BPD would not be correlated with the severity of personality symptoms but, instead, it would be linked to demographic variables, such as low socioeconomic status, a young age when conceive, belonging to ethnic minorities or prostitution[42].

**Termination of pregnancy in women with post-traumatic stress disorder**

The relationship between termination of pregnancy and post-traumatic stress disorder (PTSD) seems bimodal. Even if PTSD has been widely analyzed as a possible consequence of the termination of pregnancy (see paragraph 4), it may also be a prior event and, accordingly, be a central determinant in the choice of having an abortion, especially in case of rape, IPV and separation of the couple.

Lundell et al[43] found that the prevalence of a previous traumatic event in the female population requiring termination of pregnancy was 43%: severe physical threats, severe psychological trauma and severe physical injury were the most reported traumatic experiences, while 12% of their sample experienced sexual assault or burglary.

Steinberg et al[44] showed a high frequency of childhood adversities or traumatic experiences with the current partner in a United States female population with medically assisted abortion. Authors reported that: (1) 50% of the sample claimed they suffered any childhood psychological abuse, 31.7% any childhood physical abuse, and 13% any childhood sexual abuse; (2) 25% of the sample reported episodes of domestic gender violence in the original family unit; (3) 34.6% of the sample declared they suffered any verbal violence, 7.4% any physical violence, and 4.8% any sexual violence by the current partner; and (4) 11.9% of the sample said they had been coerced into pregnancy and 7.4% had reported sabotage of their contraception devices by their current partner.

In a study by Mota et al[33] the prevalence of previous or current PTSD in women who had a termination of pregnancy was significantly increased compared to women who decided to give birth (OR = 1.91).

On the contrary, Wallin Lundell et al[43] showed a prevalence of PTSD in the female population requiring medically-assisted abortion around 7% and this value was not statistically different from that found in the general population. Although that finding, the authors described post-traumatic stress symptoms (PTSS) in 23% of women requesting a termination of pregnancy: It is possible that women with PTSS might be as susceptible as those with PTSD to traumatic consequences after an abortion and deserve appropriate treatment[43].

In general, these findings support the importance of investigating the presence of previous traumatic events and previous or current post-traumatic stress symptoms in women who have an induced abortion, also to limit the onset or the worsening of PTSD after ToP.

**Termination of pregnancy in women with psychotic disorders**

Deinstitutionalization and second-generation antipsychotics, with fewer endocrine side effects (e.g., hyperprolactinemia and amenorrhea), allowed women with schizophrenia spectrum disorders to be more sexually active, with a subsequent increase of pregnancies in this population. However, research focusing on the reproductive health of women with schizophrenia is sparse, and many studies are limited by small sample sizes[45]. Overall, women with psychotic disorders are more likely to have worse family planning knowledge than healthy women as personal relationships is one of the treatment areas with the most unmet needs[46].

Women with schizophrenia have more sexual partners in their lives, more rarely use contraception, and, consequently, have a higher risk of unwanted pregnancies and induced abortion than the general female population[46].

Simoila et al[45], with a 15-year follow-up on the reproductive health of women with psychotic disorders in Finland, found that 30.4% of women with schizophrenia and 29.9% of women with schizoaffective disorder had at least one termination of pregnancy (VToP or TToP) during the study period.
Patients with psychotic disorders showed significantly fewer births than healthy controls (0.65 birth per woman in the study group vs 1.72 births per woman in the control group): Up to 50% of affected women did not give birth in the study period, against 12% of healthy women.

According to these evidence, some of the reasons that led women with psychotic disorders to ask for abortion were younger age at conception and a lower chance of having a permanent partner at the time of the termination of pregnancy.

Examining the specific reasons behind the choice of termination of pregnancy, Simoila et al[45] found that: (1) VToP upon social indications (e.g., low socioeconomic status) was the most common, both in women with psychotic disorders and in the general female population, involving up to 94% of women in both groups; (2) TToP for maternal health problems was significantly more common among women with schizophrenia and schizoaffective disorder than in healthy controls (2.1% vs 0.4%), maybe for medical conditions related to the disease or antipsychotics intake. Conversely, TToP for fetus health issues was significantly more common among controls than patients (3.1% vs 1.3%). That finding is somewhat unexpected since maternal schizophrenia seems related to fetal abnormalities, such as low birth weight and cardiac malformations, probably due to a linkage disequilibrium in genes related to heart and brain development[46]. The authors question this finding is related to pharmacotherapy efficacy in reducing fetal abnormalities or to the younger age of pregnant women with psychotic disorders[45]; and (3) No significant differences were found in the frequency of termination of pregnancy for ethical indication (e.g., rape, IPV) between women with psychotic disorders and the general population. This reason was rare in both groups and that result was also unexpected because several studies[47,48] reported that women with schizophrenia experience higher rates of sexual abuse and gender-based violence.

**Termination of pregnancy in women with SUDs**

Women who have abortion are twice likely to have a current substance use disorder (OR = 2.16) than women who give birth. In particular, they show a 2 times higher risk of reporting a history of alcohol or drug abuse (OR = 2.29), and 4-5 times of alcohol (OR = 4.21) or drug addiction (OR = 4.96)[34]. According to van Ditzhuizen et al[34], women with SUDs are more vulnerable to unplanned pregnancy and induced abortion than the general population due to risky sexual behaviors, such as poor use of condoms or other types of contraceptives and a large number of sexual partners. The risk of abortion is even higher among injecting drug users due to worse health conditions and clinical comorbidities, such as hepatitis C, HIV, and other physical disorders that can force that woman to choose for the termination of pregnancy. Women with endovenous drug use may be more vulnerable to indigence and discrimination, socio-economic factors that are linked to termination of pregnancy[34].

Furthermore, Diehl et al[49] demonstrated that the risk of TToP and VToP is increased in women who smoke tobacco (OR = 1.6) and in women with alcohol dependence (OR = 2.4), with a positive correlation between the severity of addiction (assessed with different screening tests for drug, alcohol and tobacco addiction) and the risk of medically assisted abortion.

In all the studies[34,49-51] that evaluated the relationship between SUDs and induced abortion, the most used substances were, in order of frequency, alcohol and tobacco, cannabis, opioids, cocaine, and, to a lesser extent, stimulants such as amphetamines.

The women’s fear of gestational and fetal consequences due to drugs (e.g., intrauterine growth restriction, higher risk of premature birth, miscarriages, and fetal abnormalities) can play a role in women’s choice of termination of pregnancy. In fact, according to some authors[50,51], the recent policies of complete abstinence from drugs during pregnancy could lead to guilt feelings in women who used them during early gestation, that may lead them to decide to terminate the pregnancy even without evidence of any gestational or fetal consequence.

However, Roberts et al[51] showed that less than 5% of women who used alcohol, smoking, or other drugs in the first trimester of pregnancy considered substance use as the central determinant for VToP or TToP. Moderate alcohol consumption was not significantly associated with the decision to terminate the pregnancy, while 84% of women who recognized alcohol consumption as their reason to have an induced abortion were binge drinkers or patients who were previously diagnosed with SUD or alcohol-related physical problems.

It is therefore important to highlight that substance use, which is modest in the general healthy population, does not appear to be a determining factor in the choice of TToP in the majority of people, but is an important element in the decision-making
process of women with a SUD or with physical problems related to addiction[34,51].

**Termination of pregnancy in women with childhood psychiatric disorders**

A predictive association between a woman's childhood and adolescent mental health and her future choice to terminate a pregnancy, for social reasons and for maternal determinants, has been supposed.

A Finnish study[52] showed that there is an association between childhood and adolescent mental health and the risk of having a ToP when young (in women under 20 years) or having repeated induced abortions. Those evidence highlight the importance of implementing sex education for adolescents suffering from psychiatric and neuropsychiatric diseases, to limit their obstetric and reproductive issues.

van Ditzhuijzen et al[34] found that women whit a history of any childhood impulse-control disorder show a higher rate of induced abortion than healthy women (OR = 4.35). In particular, the most common childhood impulse-control disorders in women who had TToP and VToP were respectively conduct disorder (OR = 1.72)[33], oppositional defiant disorder (OR = 1.67-2.99)[33,34], antisocial personality disorder (OR = 3.87) and conduct disorder (OR = 6.97)[34]. Moreover, women with a diagnosis of attention deficit hyperactivity disorder (ADHD) in childhood (especially patients with greater impulsiveness) showed a 3 times higher risk to have an induced abortion [34].

Several studies conducted in Northern Europe[52,53] support the existence of a link between a history of childhood conduct disorder in women and medically assisted abortion. According to those authors, girls with conduct disorders can be involved in risky impulsive sexual behaviors with a higher risk of unwanted pregnancy, sexually transmitted diseases, and adolescent sexual abuse. In particular, Lehti et al[52] observed that a poor school performance during primary and secondary school can be considered as an independent predictor of subsequent terminations of pregnancy in that population, according to the fact that a low level of education is an important determinant of TToP and VToP (as underlined in paragraph 2 "Demographic and reproductive determinants of termination of pregnancy").

**Termination of pregnancy in women with eating disorders**

Several studies[34-56] documented a correlation between a history of eating disorder and various gynecological comorbidities, such as impaired menstrual function, infertility, increased risk of unwanted pregnancy, gestational complications (e.g., low birth weight, preterm birth, pregnancy bleeding, gestational diabetes, hyperemesis gravidarum) and a higher risk of oophorectomy and hysterectomy.

Bulik et al[55] showed that women with anorexia nervosa (AN) have a higher risk of teenage pregnancy and unwanted pregnancy (OR = 2.11) than women without any eating disorders, with more than 50% of subjects reporting unscheduled conception and a history of prior VToP. The risk of unwanted pregnancy and VToP seems higher in this population with a history of amenorrhea or oligomenorrhea in the year before conception, suggesting that menstrual irregularities may reduce patients' adherence to contraceptive use. Bulik et al[55] showed that, despite menstrual irregularities, women with AN and bulimia nervosa (BN) can still conceive and their gynecological comorbidities (infertility, amenorrhea) can give them a false sense of security in unprotected sexual intercourse. According to authors, a better sexual education for patients with eating disorders could make them more aware of the risk of pregnancy even during menstrual irregularities and thus reduce the use of medically assisted abortion.

O'Brien et al[54] noted an increased risk of miscarriage (OR = 1.19), VToP, and TToP (OR = 1.25) in patients with AN and BN, especially in women who had a later onset of the eating disorder (18-22 years old).

Similarly, Linna et al[56] showed a higher proportion of VToP and TToP in women with classical BN (OR = 1.85) and atypical BN (OR = 1.92) and in women with AN (approaching statistical significance in the latter, with OR = 1.62, \( P = 0.06 \)). Women with binge-eating disorder (BED) did not show an increased risk of termination of pregnancy.

**Termination of pregnancy in women with paraphilic disorders and sexually compulsive/behaviors**

Due to cultural background and social stigma, women may have difficulties in identifying themselves as affected by compulsive sexual behaviors or sexual addiction. Furthermore, some people show a sexual pattern of "binge-starvation", characterized by the alternation of periods of extensive sexual activity ("sexual binging") and periods
of complete abstinence (‘sexual anorexia’) [57,58]. In the context of paraphilias and sexual compulsions, women are more inclined to sexual submission or victimization compared to their male counterparts, although sexual dominance behaviors in women are also increasing [58].

Fear of rejection and feelings of shame, associated with compulsive or addictive sexual behavior or uncontrolled paraphilias, prevent women from developing intimate relationships: in fact, it is common for women with dysfunctional sexual behavior to avoid affective attachment with their partners and more often build unstable families [57,58]. All these elements can contribute to the choice of women with paraphilic disorders and sexually addictive/compulsive behaviors to terminate a pregnancy.

Roller [57] investigated the consequences of these disorders on the reproductive health of women, such as: (1) A higher risk of sexually transmitted infections, cervical and vaginal cancer (linked to chronic HPV infection), pelvic inflammatory disease, infertility, and ectopic pregnancies; (2) A higher rate of miscarriage, due to major obstetric complications, but also to gynecological trauma determined by violent sexual practices, sometimes practiced even during gestation; and (3) A higher frequency of unwanted pregnancy (40%-70%), VTToP and TTToP (36%), and feelings of discomfort and inadequacy during gestation and motherhood (60%).

**PSYCHOPATHOLOGICAL CONSEQUENCES OF TTOP**

A medically induced abortion can have important consequences on the mental and reproductive health of women because termination of pregnancy is a traumatic event for parents.

Pregnant woman and the baby’s father experience induced abortion differently, and woman reaction might also be influenced by her will to give birth and by the moment of the pregnancy when she has abortion.

Fergusson et al [59] showed that TTToP has a greater traumatic burden and psychiatric morbidity compared to VTToP or miscarriage, that normally occur at an earlier stage of the pregnancy.

According to authors [59-61], this finding could be explained with different reasons: (1) Taking place in the second trimester, TTToP is performed with a surgical procedure, unlike the medical abortion for VTToP or uncomplete miscarriages; (2) In case of TTToP, parental desire for a child clashes with emerging medical issues and it is performed later during the gestation, when the couple gets already used to their parental role; (3) In the first trimester, women do not perceive fetal movements and the relationship with the baby is weaker than in later stages of gestation; and (4) After the TTToP, parents can hold the child in their arms and can photograph him after the abortion procedure.

It is also important to highlight that TTToP is performed with a therapeutic aim and it can also have positive outcomes, especially for those with high levels of distress due to the diagnosis of fetal malformation. It is always important to evaluate the context and the factors that lead a woman to ask for TTToP to predict different outcomes she might experience.

*Positive mental consequences of TTToP*

Despite several studies concerning the negative consequences of TTToP, it is important to underline that TTToP is a medical act performed with a therapeutic purpose.

In this regard, several authors measured a significant improvement after TTToP of pre-abortion psychopathological distress, especially mood and anxiety symptoms [60, 62-64].

Interestingly, women’s mood seems to improve significantly in the long-term after TTToP: Major et al [65], measured depressive symptoms with the BDI in different moments before and after TTToP and discovered that, immediately after TTToP, women’s average score was under the cut off for mild depression and their mood significantly improved 3 wk later. A similar positive pattern in anxiety symptoms after TTToP has been shown by Cohen and Roth [64].

*Depression and anxiety symptoms after TTToP*

Usually, after TTToP women experienced a significant improvement in their mood compared to the period before the induced abortion, but in a minority of cases, women showed long-term depressive and anxiety symptoms and sense of guilt after the procedure [62,66,67].
Zolese and Blacker[68] found that the most common long-term consequences of TToP are depression and anxiety disorders, while psychosis is uncommon. These disorders chronize in around 10% of women. 

Di Febo et al[62] show that poor partner support and relational problems related to abortion are the most important risk factors connected with significantly worse mood outcomes after TToP. An increased risk of long-term depression and anxiety disorders was also observed in women with a previous psychiatric history, in young and pluriparous women with poor social support, and in those belonging to social or religious groups that do not accept abortion[66,67]. 

Furthermore, different authors[12,69] showed that women are more likely to be depressed before abortion than after it; as previously underlined, depression after TToP could highlight a pre-existing mood disorder. 

These findings support the importance of the assessment of socio-cultural and psychopathological determinants in all women who have a TToP to predict possible depressive or anxiety consequences.

**Post-abortive mania**

Several reproductive events (menarche, menstruation, childbirth) are associated with increased mood alterations in women with bipolar disorder; however, little is known about the impact of termination of pregnancy on the course of bipolar disorder[70]. 

Mahe et al[70] described the case of a woman with a history of 5 full-term pregnancies and 2 abortions (1 VToP and 1 TToP) in which each reproductive event was followed by an episode of psychotic mania. There are currently no studies that systematically examined the relationship between abortion and mania among women with bipolar disorder, probably because post-abortion mania is much less common than post-abortion psychosis[70]. A Danish study reported a similar risk of re-hospitalization in women with bipolar disorder before and after medically induced abortion, a possible sign that this reproductive event would not be associated with a clinical worsening of the disease[31]. 

Mahe et al[70] described several clinical cases of post-abortion mania and, according to them, the post-abortion period could be at increased risk of converting bipolar II disorder to bipolar I, due to a worsening of manic symptoms in women who have TToP or VToP. Furthermore, the termination of pregnancy could constitute a trigger for manic episodes in healthy women with a predisposition for affective disorders[70].

**Post-abortive psychosis**

Brockington and Guedeney[71] reported at least 30 cases of psychotic episodes following abortion, including one historical case described by Esquirol in 1819. The episodes occurred in women with schizophrenia as well as in women who had never previously suffered from a psychotic disorder. 

Post-abortion psychosis seems very rare: Brewer[72] found a 0.3/1000 hospitalization rate of post-abortion psychosis compared to 1.7/1000 post-partum psychosis after delivery. 

The clinical presentation of post-abortion psychosis is similar to puerperal psychosis and the onset typically occurs within 2 wk after TToP. Individual predisposition and estrogenic fluctuations are supposed as the major causal factors in post-abortion psychosis and puerperal psychosis[73]. Guillaume et al[74] suggest that the risk of psychosis may be proportional to the degree of estradiol levels drop associated with the end of pregnancy (e.g., following delivery, miscarriage, and induced abortion).

**PTSS and PTSD after TToP**

Retrospective studies[75,76] showed that TToP is a traumatic event that leads to severe post-traumatic stress responses and whose distress may be evident years later. A prospective study[75] showed that TToP for fetal abnormalities is associated with an increased risk of PTSD compared to VToP. 

However, research regarding the relationship between TToP and PTSD development had controversial results. Geerinck-Vercammen and Kanhai[75] studied 89 couples at 6 wk and 6 mo after TToP and they concluded that PTSS were experienced at the time of abortion and in the following weeks, but they disappeared after 6 mo, but, this survey was based only on a semi-structured interview. By contrast, in a Dutch sample[76] women showed PTSS and depression even years after a TToP for fetal malformations. 

Kersting et al[77] examined a sample of women who have had TToP, with different psychometric scales (MADRS, BDI, IES-R, SKID) over 14 wk and found that, despite the frequent development of intense PTSS in women 2 wk after the TToP (64.5% of the
sample), at 14 wk the prevalence of PTSD was much lower (22%). Accordingly, although post-traumatic symptoms can be very common in the short-term after TToP, only about 20% of women develop long-term PTSD.

**SUDs after TToP**

A connection between VToP and increased use of alcohol, tobacco, and illicit drugs is well described [78], but this relationship is rarely investigated in case of TToP.

Drower and Nash [79,80], in a 18 mo follow-up, showed a 2 times higher risk of alcohol, tobacco, or illicit drug abuse in women who had a TToP. In fact, women who have TToP are 1.5 times more likely to abuse tranquilizers compared to women who gave birth [79,80].

**Determinants of psychopathological outcomes of TToP**

Women who have an induced abortion (TToP and VToP), may have several factors linked to a higher risk of psychological distress in response [81].

**Medical or genetic indication:** Several studies [82,83] showed that the use of TToP for maternal or fetal health reasons (e.g., malformations or genetic issues) increases the risk of negative psychological reactions compared to VToP performed on ethical or social indications (e.g., rape, incest, low socio-economic status).

**Socio-cultural and demographic factors:** Women at greatest risk of negative psychological consequences are single, nulliparous, immigrant, or belonging to cultural minorities, young (< 30 years, especially adolescents) [81]. The relationship between strong religious feelings and psychological consequences is contradictory: Romans-Clarkson [81] pointed out that religious belief is not associated with increased guilt after TToP, but Payne et al. [84] claimed that a negative religious or cultural attitude towards abortion is one of the major risk factors for adverse psychological outcomes.

**Coercion of the procedure by partner or parents:** All the legislations emphasize that the choice of TToP/VToP is personal [2], but women can be pushed to make a decision against their desire for their pregnancy, with an increased risk of negative psychological outcomes [81]. Women who have independently chosen for a termination of pregnancy show lower risk of negative psychological effects, even if adolescent [85].

**Psychopathology and psychiatric history before the procedure:** Women who experienced greater psychological distress before the abortion were more likely to have low self-esteem and a high sense of alienation after the procedure [86]. Moreover, women with an internal locus of control and high levels of resilience showed positive outcomes after TToP/VToP [87].

**Presence of emotional support:** The extent of emotional support a woman perceives during abortion is positively correlated with a good psychological response [81]. On the other hand, Major et al. [65] found that if the partner is present during abortion, there is a higher risk of depressive symptoms and sense of guilt. Similarly, Robbins [88] found that single women who maintained a strong relationship with the father of the unborn, after abortion were much more at risk of negative psychopathological outcomes. Instead, Romans-Clarkson [81] showed a higher risk of a negative psychological outcome in women with absent or weak family bonds, fewer friends and excessive work pace, partner abandonment, and absence of family support for children care.

**Presence of conscious fantasies about the fetus:** the intensity of thoughts about the baby seems directly correlated with the risk of negative psychological consequences [81].

**Illegal abortion:** Diggory [89] suggested that illegal abortion is associated with depressive reactions and feelings of guilt after TToP and VToP.

**Absence of contraception:** Important sense of guilt and depressive reactions is described in women who do not use contraception [81]. Some authors highlighted the importance of providing adequate contraception education after an induced abortion, both to limit the future recourse to the procedure and their self-blame reactions [81].

**CONCLUSION**

Several studies evaluated which decisional factors (socio-demographic, reproductive,
psychopathological determinants, and fetal abnormalities) may lead women to induced abortion, but many studies do not discriminate between therapeutic and voluntary ToP, thus representing a possible bias in discriminating decisional determinants of TToP.

Women who have a TToP are usually over 35 years old[3,9], more often divorced or unmarried[3,9], with a high educational level[3,9], but low socio-economic status and a low monthly income. They usually have more children[3], live in large urban areas[11] and belong to ethnic minorities or are immigrants[17], do not have strong religious beliefs[16], and more often report any form of gender-based violence[27].

Women with TToP and VToP show common reasons for their decision, mostly linked to personal and social difficulties that may prevent adequate child care. That evidence stresses the importance of social assistance to support women with those factors.

Women who choose TToP are more commonly affected by a previous mental disorder[28,31,33], that probably influences their mental outcomes afterward.

Results by van Ditthuijzen et al[34] and Mota et al[33] showed higher prevalence of any axis I psychiatric disorder in those women, with increased rates of anxiety and mood disorders[33,34], SUDs[34], impulse control disorders[34], ADHD[34], PTSD[33,43], psychotic disorders[45], eating disorders[54], sexually compulsive/addictive behaviors and paraphilias[57].

Interestingly, current knowledge does not discriminate any specific major psychiatric disorder as more relevant in the decision to have abortion.

The correlation between a higher rate of TToP and personality disorders is less investigated and data are limited to women with antisocial personality disorder[34] and borderline personality disorder (BPD). De Genna et al[42] explained the link between BPD and TToP with worse reproductive wellness, a huge number of unplanned and early pregnancies (< 25 years), sexual-transmitted infections, unsafe sexual intercourse, and history of childhood trauma and sexual abuse. Women with some personality disorders are more keen to unwanted pregnancies and those affected by such disorders are rising more and more in the general and clinical population. Accordingly, an in depth analysis in those populations should be encouraged to identify specificities and offer better support.

Interestingly, TToP produces both positive and negative psychopathological effects in women.

Evidence[60,62,64] showed that women with a TToP experience a significant improvement in the psychological distress, mood, and anxiety symptoms they felt before. In fact, even if TToP involves strong negative feelings like guilt and sadness, the difficulties of raising a child with illness seems even worse for some women. Those women deserve support and respect in a very delicate moment that implicates ambivalent feelings and sufferance.

Furthermore, Zolese and Blacker[68] demonstrated that only a minority (10%) of women have psychological or psychiatric disorders after a TToP and several authors [67,68] found that negative psychic outcomes after TToP are more frequent in women with previous psychiatric history (e.g., depression, anxiety disorders, mania, psychosis, and SUDs)[33,70,79]. Therefore, in most cases psychiatric consequences of induced abortion are an exacerbation of pre-existing conditions and they are not an onset of psychopathology.

Nevertheless, concerning outcomes of TTOP are demonstrated by a higher rate of post-traumatic symptoms, especially in the short-term, with a persistent post-traumatic stress disorder in about 22% of the women[77].

In conclusion, we emphasize the importance of providing psychological and psychiatric support to all women who want a TToP, to predict and limit adverse outcomes on their mental and reproductive health. Moreover, further studies are needed to assess the mental history of the male partners and their possible outcomes after a TToP.

REFERENCES

252 [PMID: 18468755 DOI: 10.1016/j.socscimed.2008.03.033]
di Giacomo E et al. TToP and women’s mental health

10.1371/journal.pone.0181104


84 **Payne EC**, Kravitz AR, Notman MT, Anderson JV. Outcome following therapeutic abortion. *Arch Gen Psychiatry* 1976; **33**: 725-733 [PMID: 938192]


G-protein coupled receptors and synaptic plasticity in sleep deprivation

Shweta Parmar, Ramakrishna Tadavarty, Bhagavatula R Sastry

ORCID number: Shweta Parmar 0000-0002-9124-3619; Ramakrishna Tadavarty 0000-0001-7448-0367; Bhagavatula R Sastry 0000-0002-0795-0493.

Author contributions: Parmar S, Tadavarty R and Sastry BR contributed equally to the outline of this review; Parmar S performed 85% of the writing work, and Tadavarty R performed 15%; Illustrations were done by Parmar S; Suggestions and corrections were contributed by Sastry BR; All authors have read and approved the manuscript.

Conflict-of-interest statement: Authors have nothing to disclose.

Supported by Canadian Institutes of Health Research Grant, No. TGS-109219; and 4-Year Fellowship from the University of British Columbia.

Country/Territory of origin: Canada

Specialty type: Neurosciences

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report's scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C

Abstract

Insufficient sleep has been correlated to many physiological and psychoneurological disorders. Over the years, our understanding of the state of sleep has transcended from an inactive period of rest to a more active state involving important cellular and molecular processes. In addition, during sleep, electrophysiological changes also occur in pathways in specific regions of the mammalian central nervous system (CNS). Activity mediated synaptic plasticity in the CNS can lead to long-term and sometimes permanent strengthening and/or weakening synaptic strength affecting neuronal network behaviour. Memory consolidation and learning that take place during sleep cycles, can be affected by changes in synaptic plasticity during sleep disturbances. G-protein coupled receptors (GPCRs), with their versatile structural and functional attributes, can regulate synaptic plasticity in CNS and hence, may be potentially affected in sleep deprived conditions. In this review, we aim to discuss important functional changes that can take place in the CNS during sleep and sleep deprivation and how changes in GPCRs can lead to potential problems with therapeutics with pharmacological interventions.

Key Words: G-protein coupled receptors; Metabotropic glutamate receptors; Gamma-amino butyric acid-B receptor; Synaptic plasticity; Sleep deprivation; Memory consolidation

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Sleep hygiene is thought to be necessary for memory consolidation and learning while sleep disturbances can alter both synaptic plasticity and memory consolidation. Recent findings indicate that the expression of G-protein coupled receptors...
INTRODUCTION

Sleep is one of the fundamental needs of most conscious living beings. Earlier sleep was more analogous to just the idea of resting and having our mind and body recharge for the activities in our consecutive state of wakefulness. We have moved far along from this initial notion as we now know that even in the state of rest, our body is constantly working right from the molecular to the more visible physiological and psychological levels. The implications associated with sleep deprivation manifest themselves both behaviourally as well as physiologically. In a way, both these implications are correlated considering external stimuli enables an animal to act in a certain way depending on what kind of signals the brain has processed. The processing of such signals is mediated by cellular components like genes and proteins. In different neurological disorders like depression, loss of memory, psychosis, hallucination, anxiety, etc., certain biomolecules are differentially affected. Synaptic plasticity which is thought to be a cellular correlate involved in memory consolidation and learning, is also affected following sleep deprivation.

In this review, therefore, we aim to understand and highlight the fundamental properties of sleep and sleep deprivation by keeping memory consolidation and synaptic plasticity as representative functions which get affected by the different states of sleep. In addition, any changes in G-protein coupled receptors (GPCRs) caused by sleep disturbance are emphasised.

THE DYNAMIC STATE OF SLEEP

Sleep and its significance

Sleep is indispensable for most animals. Extensive studies in this area have now led us to believe that sleep is not a static phenomenon but it is heterogenous and dynamic even if its physical nature appears completely passive. There are interconnections between the thalamus, cortex and hippocampus regions of the brain and this interplay of networks is known to be operated by stage specific oscillations that take place while we are asleep[1,2]. Aserinsky was one of the first researchers whose studies overcame the concept of the cerebral cortex being dormant in sleep and gravitated the scientific community towards the reality of the brain being in an active state even while we are asleep[3]. There are stages or periods in sleep which transition alternatively from an active period, which is manifested by rapid eye movements (REM) to an inactive or quiescent period, also known as the slow-wave sleep (SWS)[3]. The oscillation patterns in the brain fluctuate during these two stages of sleep with high frequency and low amplitude waves being a characteristic of the REM/Late-night paradoxical sleep and the SWS portraying the contrary wave pattern[2,3]. Advancement in this area has revealed many more biological functions like respiration, thermoregulation, etc., along with the neurobiological processes of learning and memory formation in association with the alternating stages of REM sleep and SWS[4].

Disturbances in regular sleep cycles were assessed through the monitoring of EEG patterns. Waves of low amplitude and high frequency, associated with wakefulness, also occur during REM sleep. Deprivation of sleep is usually followed by a compe-
nsatory increase in REM sleep (REM rebound). This increase is interpreted to indicate an attempt in re-establishment of homeostasis in learning and memory as well as emotional balancing[5]. Non-REM sleep is associated with high amplitude and low frequency delta waves (stages 3 and 4) along with stage 2 spindle activity[2,3]. Sleep deprivation increases the amplitude of the waves associated with spindle activity but reduces spindle density[6-9]. Upon chronic sleep restriction, the power density of theta wave frequency increases[10,11]. Therefore, lack of sleep seems to result in changes in REM as well as non-REM sleep.

The functional significance of sleep guides us towards the negative implications of sleep deprivation. A balanced and sufficient sleep cycle acts as one of the major factors that determine the quality of human life. There are a wide range of environmental, psychological and physiological factors that lead to sleep deprivation. Environmental changes have influenced many aspects of our day-to-day life and sleep disturbances can arise due to an increase in surrounding noise as well as fluctuations in light and temperature[12]. These factors add on to the list of causes that negatively burden the state of mind. Mental health also plays a crucial role in maintaining a regular sleep pattern since the prevalence of stress, anxiety, depression, etc., affects regular sleep cycle which can translate into insomniac conditions. Psychosis disorders like schizophrenia or neurodegenerative diseases like Alzheimer’s have often been associated with issues related to sleep deprivation like a reduced REM sleep cycle as well as a lowered sleep spindle activity[13,14]. Apart from mental disorders, pathophysiological illnesses (e.g., cancer, diabetes, respiratory disorders etc.) also often result in sleep disturbances due to manifestations like pain or difficulty in breathing. Interestingly, lack of sleep can also increase the prevalence of such physiological disorders, since an important function of sleep is the regulation of the immune system[15]. Modern lifestyle changes like uneven working hours, over consumption of caffeine along with an increase in screen-time exposure have potentially interluded the quality and quantity of sleep. Homeostasis of the normal biological circadian rhythm is required for better cognition and task performance. For example, studies have implicated that night shift-workers, especially in chronic situations, experience varying levels of cognitive impairment and task performance[16,17]. Drug abuse and alcohol consumption which can lead to substance use disorder have also been recognized as a growing cause of sleep disruption. For instance, alcohol consumption is indicated to suppress REM sleep which in turn causes impairment in performing procedural tasks [18]. Conversely, lack of sleep affects the activity balance of important neurotransmitters like dopamine (DA) in the brain and this has been indicated to increase vulnerability to the use of drugs[19]. The above mentioned negative influences of sleep deprivation are just few of the many but it helps in bringing the importance of sleep deprivation into perspective.

Without sufficient amount of sleep, the brain cannot adequately perform the processes that take place during the state of sleep. Sleep deprivation affects the ability to concentrate, intake information and mediate that information through neuronal signalling, learn as well as process memories for consolidation. Furthermore, it is affected by the signalling and expression profiles of many biological molecules including the GPCRs[20,21]. These functions are related to the different stages of sleep and hence there are stage specific implications of sleep deprivation. For example, lack of REM sleep has been directed towards an impairment in the development and expression of emotional and spatial memories[22,23]. Additionally, disruption of SWS is suggested to reduce attention span, affect motor activity and task performance[24]. Since learning and task performance are dependent on the memory processing of various responsible stimuli, the concept of memory consolidation has been highlighted to understand the significance of sleep and its associated cellular machinery.

Sleep and memory consolidation

Memory is the ability of living beings to retain information that they have acquired through their various day to day experiences and activities. The brain consolidates memories in different stages starting first from acquiring the memory through learning experiences, encoding and then consolidating those memories to be recalled or retrieved upon stimulus[25]. Interestingly, even the processed memories can become transient overtime and a reconsolidation of those memories is required for which sleep is essential[26]. While we sleep, the brain is actively carrying out its functions and ‘offline-reprocessing’ of memories appears to be one of them[3]. The presence of the rapid and spontaneous oscillations in the cortical networks of the brain during wake and sleep periods have been known to get triggered upon sensory stimuli [27,28] which can get incorporated in the brain in form of a memory. This gradual incorporation of acquired memories into the different regions of the brain is more
crucial for long term memory than the short-term memory, which points towards a quicker mechanism of action[29,30]. REM sleep and SWS are generally associated with the consolidation of long-term implicit (non-declarative) and explicit (declarative) memories, respectively[31-33], both working in a double-step process. In support of the alternate nature of the REM sleep and the SWS, work done on the memory consolidation function of these stages has revealed that SWS also has its part to play in reactivation and redistribution of the spatio-temporal patterns of the neurons which are observed during the encoding stage of memory function when we are awake[31, 34-36]. When there is distraction or disruption of the periods immediately after the learning or teaching stimulus, the formation of memory is interrupted and consequently the task reperformance associated with that memory is affected[29,30]. Now, since memory is processed during the state of sleep, it appears that disruption of this state during sleep deprivation can also affect memory consolidation in the brain. For instance, there have been studies demonstrating how lack of sleep can cause weak recollection of visual stimuli or how motor skill learning improvements are dependent on a good night-time sleep[37-39].

The discussed oscillation and wave patterns during the state of sleep are at least partly a result of the electrical impulses that are transmitted between neurons via the synapses. This transmission modifies the synapses and hence synaptic plasticity is an important consideration while studying about the various functions of sleep including memory consolidation. To better understand this correlation, we will first discuss the regions and cells of the brain that are associated with sleep dependent functions and how these regions communicate via signal transmission.

The cortical hippocampal dialogue
As mentioned earlier, activities in the brain during the state of sleep are in conjunction with changes in the oscillation and wave-patterns in specific regions of the brain. These complex wave-sequences form a characteristic component of the neocortical and the thalamic system, which co-ordinate in order to exhibit the various functions of sleep, including memory consolidation. Although the mechanisms that are involved in the functions and regulation of memory consolidation are yet unclear, various cellular processes in brain regions have been studied as potential candidates. Amongst these, the hippocampus is one of the main centers for processing sleep dependent memories through neurons like granule cells and pyramidal cells[40]. The significance of the hippocampus and its sub-regions may be attributed to its function of independently giving the neuronal networks the ability to process multimodal information and building an integrated system that can represent the newly acquired information for its consequent storage in the higher centers of the brain in the form of ‘memories’[41]. Inspired by this finding, a two-staged memory consolidation concept was put forward which essentially highlighted the events wherein there occurred a transmission of information between the hippocampus (short-term storage) and the neocortex (long-term storage) through the exploratory theta burst stimulations of the granule cells and sharp waves of the pyramidal cells[42]. While the hippocampus is concerned with high frequency oscillations and sharp wave ripples in the CA1 and CA3 regions, the neocortex shows slow oscillations and spindles during SWS[43]. During SWS, the signal from the hippocampus to the neocortex is transmitted through the CA1 neurons whereas during REM sleep the information enters the hippocampus from the neocortex through the CA3 neurons; thus, causing the CA1 neurons to hyperpolarize the CA3 neurons through the entorhinal cortex[43]. This rhythmic communication of oscillations gives rise to coupling as well as consistent spike timing relationship between the neocortex and the hippocampus which further strengthens the information processed for learning and memory consolidation[43-45].

As discussed above, there seems to be a pattern of relay of information that is associated with sleep dependent functions between the hippocampus and the neocortex. Since the relayed signals go through synapses, any plasticity at the junctions can significantly alter the above mentioned patterns. Therefore, in this review, a discussion on the phenomenon is made.

THE ROLE OF SYNAPTIC PLASTICITY IN SLEEP AND LEARNING
The active nature of brain during the state of sleep so far has been advocated by multiple studies on the specific regions of the brain which primarily show distinct wave patterns and relay information through signal transmission. Transcending from the Hebbian[46] theory, the strength of these signals has been associated with the
continuous firing of action potentials from a pre-synaptic neuron to a post-synaptic neuron. It has appeared that a signal gets stronger when it is repeatedly transmitted from one synapse to another and this is where the phenomenon of synaptic plasticity comes into play. When there is continuous firing of signals between neurons, there are changes that take place at the synapses of these cells, which means that the cellular components at these junctions also undergo changes and consequently affect the strength of the signal being transmitted. This concept can be appreciated by looking at instances wherein it appears that by repeatedly performing a task, we get better and better at it. In a similar way, a repeated stimulus to the brain results in repeated signal transmission from one neuron to another and this can have the potential to strengthen that signal[46]. This takes place through synaptic plasticity which entails molecular as well as electrophysiological changes at the synapses. The electrophysiological changes can be characterized by long term depression (LTD) and long term potentiation (LTP), each relating to the persistent weakening and strengthening of the synapses, respectively. Initial work done in this area revealed that at the excitatory synapses, there occurred a use-dependent and long term strengthening of the synapse upon high frequency stimulation of the pre-synaptic fibres in the hippocampus, which was termed as LTP[47]. Analogous to the Hebbian[46] findings, experimental data at cellular level has indicated that synaptic plasticity in the form of LTP in the hippocampus is input specific and can be associative in the sense that if activity in inputs coincides with a depolarization of the post-synaptic neuron, the active synapse gets strengthened[48-50]. The post-synaptic depolarization is thought to remove a Mg²⁺ block of the N-methyl-D-aspartate (NMDA) receptor coupled ionic channel allowing a Ca²⁺ influx which stimulates protein kinases that sets up the expression of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors on the dendritic spines at synapses leading to an increase in response to the released glutamate, the transmitter[51]. The post-synaptic depolarization is also suggested to facilitate the release of a retrograde messenger that causes an increase in glutamate transmitter release and, thus, an increase in synaptic transmission[49]. Upon stimulation of synapses at lower frequencies than those that induce LTP, the induction of LTD also involves an influx of Ca²⁺ ions, although the surge of Ca²⁺ ions is comparatively lower than that in case of LTP[52]. The difference in the level of Ca²⁺ influx results in activation of phosphatases that are responsive to lower concentration of Ca²⁺ ions. This subsequently leads to a dephosphorylation of AMPA receptors and a reduction in their activity which decreases the response to the released glutamate and causes an overall reduction in synaptic efficacy[53]. Therefore, the phosphorylation of AMPA receptor through kinases in LTP is counteracted by the phosphatases during LTD which leads to an interference and reversal of LTP. This interplay promotes modulation of synaptic plasticity which as explained in the section below, appears to be crucial for memory processing and learning. The comparative details of LTP/LTD associated molecular events are summarized in Table 1 and are also discussed further in subsequent sections of this review[54-66].

The process of learning and memory consolidation that takes place while we are asleep can also be attributed to further molecular changes that can take place following the synaptic strengthening. A disruption of these due to sleep deprivation can, thus, have a negative consequence.

**LTP and LTD as cellular correlates for learning and memory formation**

In spite of having a breakthrough discovery about LTP and LTD causing long-lasting modification in the synapses in the 1970s[47,67], the functional implications of LTD and LTP in learning and memory processing are still under a lot of conjecture. While the initial studies relied more on the active state of synapses through LTP as the focal model for learning, subsequent studies also implicated the role of LTD by highlighting the fact that a consequent increase and decrease in signal strength is what allows the phenomenon of synaptic plasticity to function without getting worn out[52,58,59,68]. For instance, Tsunoto[65]’s work highlights LTD in cerebral cortex being responsible in ‘forgetting’ certain memories in order to make place for the new ones. Looking back at Hebbian[46] and the follow-up findings, the role of synapses and the existence of synaptic plasticity as cellular correlate of learning and memory consolidation become apparent. As discussed in previous sections, learning and memory consolidation are a result of collective cellular communications through the processing of the learning stimuli. The cortical-hippocampal dialogue is thought to drive these learning stimuli forward in the brain to encode, process and store the information acquired through those stimuli. Several lines of studies demonstrate the correlation of memory acquisition with the processing of spatial learning to LTP and LTD[65,66,69]. Furthermore, exploratory behaviours and learning are known to promote information
### Table 1 Comparative features of long term potentiation and long term depression and their implications in sleep and associated functions

<table>
<thead>
<tr>
<th></th>
<th>Prominent brain regions</th>
<th>Receptors involved</th>
<th>Induction</th>
<th>Maintenance</th>
<th>Association with sleep</th>
<th>Functional significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LTP:</strong> Strengthening of synaptic transmission due to a transient high frequency stimulation of the synapses</td>
<td>Cerebellum, hippocampus, cerebral cortex</td>
<td>Ionotropic: NMDARs, AMPARs</td>
<td>Requires activation of both pre- and post-synaptic neurons at the same time for glutamate to fully activate the NMDA receptors [51].</td>
<td>Secondary activation of receptors like AMPARs by glutamate needed for maintenance of LTP [51].</td>
<td>Active nature of LTP likely associated with active state of sleep.</td>
<td>Enhances synaptic response as well as neuronal excitability.</td>
</tr>
<tr>
<td></td>
<td>Cerebellum, hippocampus, cerebral cortex</td>
<td>Metabotropic: Group I mGluRs</td>
<td>Ionotropic induction requires activation of NMDARs by glutamate released from pre-synaptic neurons for higher Ca(^{2+}) influx in the post-synaptic neuron [51].</td>
<td>Requires activation of intermediate protein kinases like CaMKIV, protein kinase M-ζ, PKA etc. by high Ca(^{2+}) concentration to release membrane obscured AMPARs [56].</td>
<td>More commonly linked with REM sleep associated cellular and molecular modulation of synaptic plasticity [57].</td>
<td>Involved in memory consolidation and learning; Strengthened synapses promote long-term memory storage [58]; Promotes associative and spatial learning [59]; Involved in motor learning and task reperformance [60,61].</td>
</tr>
<tr>
<td><strong>LTD:</strong> Weakening of the synaptic strength due to a relatively low frequency stimulation of the synapses</td>
<td>Cerebellum, hippocampus, cerebral cortex</td>
<td>Ionotropic: NMDARs</td>
<td>Activation of pre-synaptic neuron sufficient to trigger a moderate response of NMDA receptors and does not require both synapses to be activated at the same time [61].</td>
<td>Deactivation of AMPARs involved in maintenance of LTD [131].</td>
<td>Suppressive nature of LTD more likely to be associated with quiescent state of sleep.</td>
<td>Interferes with LTP thereby providing a counteractive balance to prevent hyperexcitability of neurons [52].</td>
</tr>
<tr>
<td></td>
<td>Cerebellum, hippocampus, cerebral cortex</td>
<td>Metabotropic: GPCRs like mGlu4Rs, mGlu5Rs, GABA-B</td>
<td>Ionotropic induction requires moderate activation of NMDA receptors by glutamate for Ca(^{2+}) influx in post-synaptic neuron which is lower than that for LTP induction [52].</td>
<td>Requires activation of intermediate protein phosphatases like protein phosphatase 1, 2 and calcineurin by low Ca(^{2+}) concentration to inhibit the release of membrane obscured AMPARs [53].</td>
<td>More commonly linked with SWS associated cellular and molecular modulation of synaptic plasticity [61].</td>
<td>Involved in memory consolidation and learning; Weakened state of synapses implicated in forgetting old memories in order to make space for new ones [62]; Involved in novelty acquisition and spatial learning [66].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metabotropic induction through mGlu4Rs and mGlu5Rs causes endocytosis of the expressed AMPARs and/or a decrease in phospholipase C via reduced adenylyl cyclase activity [62].</td>
<td>Enhancement of GABA-B mediated inhibitory effects may result in LTD [63].</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GABA-B: Gamma-amino butyric acid-B; NMDA: N-methyl-D-aspartate; SWS: Slow wave sleep; AMPA: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; LTP: Long term potentiation; LTD: long term depression; mGlu: Metabotropic glutamate; 5-HT: 5-hydroxytryptamine.

acquisition [70] and experimental studies on rats that were exposed to hole board exploration resulted in events that reversed LTP upon exposure to a novel environment with inductive LTD mechanisms being activated. These rats showed habituation upon a second exposure to that environment indicating LTD’s possible
role in remembering the previously acquired information of that environment[66]. On
the contrary, work done on adult rats has also revealed the significance of LTD and
LTP by demonstrating that the preservation of LTP is hampered upon an increase to
LTD susceptibility and this also interrupts the CNS functions including learning and
information retention[71,72]. Therefore, there are debatable influences of LTD on
learning but its role in novelty acquisition and spatial learning[66,73] deserves further
investigation.

What exactly is harnessing this LTP and LTD in these regions of the brain and how
these activity dependent synaptic changes relate to memory consolidation and
learning? In this review, an attempt is made to find some answers. An early hypothesis
put forward by Lynch and Baudry[74] on the biochemistry of memory involved an
examination of calcium proteinase-receptor interaction. They demonstrated that in
the forebrain sub-synaptic membranes, an influx of Ca²⁺ causes a long lasting increase in
the number of glutamate receptors through its activation of the enzyme calpain which
is a proteinase. The activated form of this enzyme degrades the membrane-anchored
cytoskeleton protein fodrin and as a result, exposes the obstructed glutamate
receptors. The effects of this Ca²⁺ proteinase-receptor interaction could functionally
modify neuronal circuits as well as showed similar biochemical effects that are seen
post learning[74]. This and other related studies examining molecular mechanisms
were, therefore, examined to find correlations between synaptic strength and memory
consolidation.

Most work done on excitatory neurons like the hippocampal pyramidal cells reveal
that tetanic stimulation of inputs can readily induce LTP that lasts for hours, days and
can even be permanent, making it a good cellular correlate of memory and learning
[75-77]. In vivo and in vitro inductive of LTP and LTD through external theta burst
stimulations in model organisms[47,49,78] has revealed the various cellular
components that present themselves as functional entities in the process of memory
formation and learning. It is already well established that receptors like metabotropic
Glutamate receptors (mGluR), ionotropic NMDA and AMPA receptors are mediators
for LTP and LTD through the increase or decrease in the Ca²⁺ concentration,
respectively[51,79,80]. Both ionotropic (e.g., NMDA, AMPA receptors) and metab
otropic (e.g., mGluR) receptors are known to be involved in synaptic plasticity but
project different mechanisms of action which are employed for mediating excitatory
and/or inhibitory signals[81]. The ionotropic receptors are heteromeric compounds
that consist of a ligand binding site and pore forming channel that combines the
function of the receptor and ion channel into a single unit for the mediation of post-
synaptic potentials. On the other hand, metabotropic receptors do not comprise of an
ion channel but mediate their effects on other ion channels via intermediate effector
molecules and secondary messengers. Ionotopic receptors like NMDA and AMPA
receptors get activated upon binding of the neurotransmitter glutamate on the ligand-
bindin site and this leads to an opening of the ion channel subunit of the receptor
which is permeable to Na⁺ and K⁺ ions to induce an inward current causing depolarization
through the excitatory post-synaptic potential (EPSP) that lasts for a few
milliseconds[51]. NMDA channel is permeable to the above cations and also to Ca²⁺
which affects the downstream signalling cascade that regulates synaptic plasticity[82].
Glutamate can also bind to the mGluRs which leads to an activation of signal
transduction molecules like the receptor associated G-protein that gets detached from
the receptor and directly activates the nearby ion channels like NMDA receptors or
promotes the action of various effector molecules that indirectly activate other ion
channels. mGluRs can also regulate the intracellular Ca²⁺ levels through membrane
proteins and secondary messengers. The influx of ions through these receptors
therefore results in post-synaptic potentials that have a slower induction and longer
response time ranging from a few milliseconds to even much longer times[83]. LTP is
known to be predominantly mediated by the ionotropic receptor NMDA when high
frequency stimulations result in a build-up of EPSPs due to the activation of AMPA
receptors causing the removal of a Mg²⁺ block of the NMDA channel[51,58]. However,
metabotropic receptor mediated LTP has also been observed with mGlU1 and mGlU5
receptors wherein mGluR activation can lead to LTP induction through NMDA
receptors as well as via the modulation of intracellular Ca²⁺ levels[84,85]. As mentioned
before, LTD arises upon a lower frequency stimulation of the excitatory synapses than
needed to induce LTP and also involves NMDA receptors but with a resulting
deactivation of the AMPA receptors. LTD mediated by metabotropic receptors like
mGluRs usually involves secondary messengers that modulate intracellular Ca²⁺ levels
[62]. Other metabotropic GPCRs like gamma-amino butyric acid-B receptors (GABA-B
Rs) can also induce LTD due to their inherent nature to transmit inhibitory signals[63].
The details of these molecular and cellular features of synaptic plasticity will be
discussed more in the coming sections. Since activity dependent synaptic plasticity and memory consolidation occur sequentially, cellular receptors and other associated molecules that mediate LTP and LTD consequently find themselves to be associated with learning and memory processing[61,86].

Significance of synaptic plasticity in sleep and sleep deprivation

The active state of sleep has been associated with many functions, including synaptic plasticity (LTP and LTD), cognitive learning and memory processing[57]. In fact, it is also reasonable to suggest that learning paradigms that promote sleep and synaptic plasticity also enhances brain development[57,88].

The above notions have risen from earlier hypotheses that sleep affects the patterns of neuronal circuits in the brain[42,89]. REM sleep is associated with long-term memory consolidation and learning events promote synaptic plasticity. Moreover, REM is required for brain development and an increase in intensive learning activity significantly enhances the duration and number of REM sleep cycles and promotes synaptic plasticity[90,91]. Correlating sleep cycles and synaptic plasticity, LTP has been credited with many learning functions of the brain, especially in the hippocampus. It appears that sleep deprivation results in an inhibition or decrease in the level of LTP and an impairment in the associated learning[92,93]. This trend has also been observed in rats during sleep fragmentation where in there was a decrease in LTP through NMDA receptors in the CA1 neurons of the hippocampus[94]. Interestingly, the reduction in LTP is also associated with a downregulation of cortactin, which is a dendritic cytoskeleton protein, pointing towards the role of LTP in the structural modification of neurons and a disruption upon sleep deprivation[95].

While the impact of sleep deprivation on LTP seems clear, the fate of LTD has not been adequately identified. However, LTD is known to protect neurons by regulating neuronal hyperexcitability and preventing a saturation of network activity[39]. Additionally, LTD’s role in memory consolidation has been seen through events like novelty acquisition and spatial learning or even in making space for new memories through the concept of ‘forgetting’[65,66,73]. Since these aspects of memory processing are sleep dependent, it is logical to suggest that sleep deprivation can affect LTD. In agreement with this line of thinking, an increase in hippocampal LTD following sleep deprivation was observed in rats[96].

In conclusion, sleep seems to promote synaptic plasticity, especially LTP, and memory consolidation, while sleep disturbances appear to adversely affect both. LTD seems to be enhanced following sleep deprivation and, thus, may be associated with interference with memory consolidation.

SIGNIFICANT REGULATORY MOLECULES IN SLEEP AND SLEEP DEPRIVED CONDITIONS

Molecular events in association with sleep and synaptic plasticity

After a brief introduction of a few cellular components involved in the functioning of synaptic plasticity in section II, some detailed events in relation to the state of sleep are described here. Going back to the fundamentals of electrophysiology, an electrical signal is transmitted when a neurotransmitter gets released from the pre-synaptic cell and activates its specific receptor on the post-synaptic cell. This further invites a series of downstream reactions that ultimately give rise to the required effect. Neurotransmitters and receptors are cellularly synthesized molecules and if the signal pattern changes during stages of sleep and wakefulness, changes in the expression of genes and proteins, that are precursors for these molecules, can also occur. In sleep and sleep deprivation, glutamate, acetylcholine, DA, serotonin, etc., are important transmitters to pay attention to. Acetylcholine, for instance, is known to be correlated with the amplitude of theta burst oscillations that take place in the hippocampus. Events that lower the levels of acetylcholine can reduce the amplitude of such oscillations[97]. Since sleep sensitive processes of learning, like memory consolidation demonstrate oscillatory changes, neurotransmitters like acetylcholine are significant to such brain functions. Additionally, imbalances in the levels of neurotransmitters like DA and serotonin are associated with many neurogenic issues like depression, anxiety and psychosis; all of which are symptoms often seen in patients that have been sleep deprived[98-100].

Focusing more on the receptors involved in mediating synaptic plasticity in excitatory glutamatergic pathways, the glutamate receptors present themselves as...
frontline players and amongst these, the ionotropic NMDA and AMPA receptors are known to depolarize/excite the post-synaptic neuron through an influx and efflux of Na+ and K+ ions, respectively[83]. The main inhibitory neurotransmitter of the CNS is GABA and it acts upon the two types of GABA receptors: Ionotropic GABA-A/GABA-C receptors and metabotropic GABA-B Rs. GABA’s contribution to modulating LTP and LTD in excitatory pathways, owing to the strategic location of its receptors to regulate input and output signals in the pyramidal neurons, is well documented[101-103]. LTP and LTD are carried out by the long term excitatory and inhibitory actions of these receptors including GPCRs, which will be looked into in the next section. Ca²⁺ has been a crucial factor and its increase has shown to depend on the activation of NMDA receptors by glutamate, Ca²⁺ release from intracellular stores and its influx through the voltage gated calcium channels[104]. The binding of the released calcium to calmodulin and then further to two other proteins: Ca²⁺/calmodulin dependent protein kinase type 2 (CAMK II) and calcineurin, results into LTP and LTD, respectively[104].

In continuation with the molecular changes associated with synaptic plasticity, protein synthesis and post-synaptic biochemical changes have been implicated in different phases of LTP[105]. For instance, brain derived neurotrophic factor, protein kinase Mζ calcium/CAMK II, and activity-related cytoskeletal protein, have so far been indicated in the induction or maintenance of late phase LTP[106-109]. Whether these or other factors participate during sleep-induced consolidation of memory is unknown and needs future investigation. Another not so conventional molecule that is speculated to promote synaptic plasticity is nitric oxide (NO). What is intriguing about this molecule is that the enzyme which synthesizes it is activated by an increase in Ca²⁺ concentrations which again is dependent on receptors like NMDA which promotes synaptic plasticity[57,110]. More insight on such accessory molecules needs to be put forward through detailed experimental work and NO has so many other important physiological functions like decreasing vascular resistance and increasing cerebral blood flow and oxygenation rate[112] or playing a key role in mediation of an immune response against infectious diseases[112]. Thus, its definitive role in sleep deprivation through synaptic plasticity can be a major contribution to this field of research.

In summary, these are just few of many cellular components which have been discussed here to help understand and appreciate the intricate network of cellular molecules in driving synaptic plasticity which also play an important part in carrying out sleep associated functions. The functions carried out by these regulatory molecules have been known to get altered during the state of sleep and sleep deprivation. This means that the genes that transcribe the proteins for these receptor molecules also undergo changes. Differential gene expression has been observed during both the states of sleep and wakefulness. In fact, modulation in the gene expressions during these two states have shown that around 10% of the total cerebral cortex transcripts are differentially expressed[113]. For instance, the mRNAs for the receptors of the inhibitory neurotransmitter GABA, shows a higher expression during the state of sleep as compared to the state of wakefulness[113,114].

**GPCRs**

GPCRs by definition are receptors that mediate their effects after binding to G proteins which are heterotrimeric molecules made up of alpha, beta and gamma subunits that function by phosphorylating the nucleotide guanosine triphosphate to guanosine diphosphate. Structurally, GPCRs are transmembrane proteins with multiple transmembrane domains and the homology of these domains categorize them into 4 classes (A, B, C and F). About 90% of the non-sensory GPCRs are expressed in the brain and appear to play regulatory roles in various neurological processes. The expression profile of these receptors is extremely high in the hippocampus (around 300 GPCRs) and about 20 of these GPCRs display their potential role in synaptic plasticity[115]. The signalling pathways that are induced by the GPCRs regulate both pre and post-synaptic components which ultimately can affect synaptic plasticity as well as the release of pre-synaptic vesicular molecules[116]. They are also known to be involved in events of structural plasticity as well as cognitive development[117]. GPCRs can also alter NMDA receptors by a direct action at the CA1 synapses, thus indirectly affecting NMDA receptor-mediated synaptic plasticity[118], suggesting the possibility for receptor co-expression. Furthermore, GPCRs also seem to elicit their effects intracellularly while modulating synaptic plasticity[119]. The range of neuronal
functions that are attributed to GPCRs needs a more coherent and concrete investigation to be well defined in relation to sleep deprivation, synaptic plasticity and memory consolidation. For the purposes of this review, two major types of GPCRs known as the mGluRs and GABA-B Rs are discussed in more detail. Other GPCRs, especially those that are important for drugs used in psychiatry, are also discussed.

**mGluRs**

The mGluRs belong to the class C of GPCRs. They are categorized into three groups based on their signal transduction pathways. Group I (mGlu1R and mGlu5R), Group II (mGlu2R and mGlu3R) and Group III (mGlu4R, mGlu6R, mGlu7R, mGlu8R)[120]. The subdivisions are based on the difference in their physiological activity and structure. The group I receptors display a post-synaptic location and are known to act via the Gq protein through the activation of phospholipase C protein and the synthesis of the secondary messengers such as inositol-1,4,5-triphosphate (IP3) and diacylglycerol (DAG). The general pathway involving these secondary receptors results in the release of intracellular Ca\(^{2+}\) stores which is a prerequisite for the induction of synaptic plasticity[54,55]. These receptors are also involved in the modulation of neuronal excitability in the hippocampus pyramidal cells wherein they activate cationic conductance through the reduction in resting K\(^+\) current and inhibition of Ca\(^{2+}\) ion channels[121,122]. On the other hand, mGluRs were also seen to induce IP3-mediated release of Ca\(^{2+}\) from intracellular stores through stimulation of CA3 pyramidal neurons. Unlike the Group I mGluRs, Group II and III mediate their functions via the Gi/Go protein. They are negatively coupled to adenylyl cyclase and inhibit the production of cyclic adenosine monophosphate Cyclic adenosine monophosphate (cAMP)[123] which results into activation of K\(^+\) channels and inhibition of Ca\(^{2+}\) channels. These two groups have a very distinct pre-synaptic localization pattern with Group II being predominant in extra synaptic sites whereas Group III in the synaptic sites. Additionally, the Group III receptors also appeared to be segregated in correlation to their target post-synaptic neurons[124].

**GABA-B Rs**

The metabotropic GABA-B Rs are widely known for their inhibitory action in the CNS and like mGluRs, also belong to the class C GPCRs. Structurally, these are obligate heterodimers made up of GABA-B1 and GABA-B2 receptor subunits which are genetically co-expressed and are homologous to the structure of mGluRs[125,126]. Gene expression studies for this receptor have shown that heterodimerization is essential for the GABA-B Rs to elicit their function since the individual expression of the GABA-B1 subunit in the target mammalian cells results in immature and low functionality receptors as compared to the heterodimer[127]. These receptors show a wide cellular and subcellular localization pattern through their expression on pre-synaptic, post-synaptic as well as extra synaptic regions in the brain[128]. Based on work performed on rat hippocampus, the two subunits of GABA-B Rs: GABA-B1a/b and GABA-B2, appear to be more on the post-synaptic regions as compared to the pre-synaptic ones. Furthermore, they also displayed an abundant distribution of GABA-B1 on glutamatergic synapses on the spines and on dendritic shafts of pyramidal cells[128]. GABA-B2 on the other hand, has a dominant extra-synaptic localization. These localizations are indicative of their functions and considering the range of cellular and subcellular occurrences of these receptors, they show a promising involvement in a great number of neurogenic processes. For example, their presence at the glutamatergic terminals points towards their involvement in glutamatergic neurotransmission[128].

Functionally, GABA-B Rs induce their slow and long lasting action via the Gai/o proteins through the inhibition of adenylate cyclase and like the mGluRs, they too affect the Ca\(^{2+}\) and K\(^+\) conductances[129]. GABA-B Rs are known to control the calcium dependent neurological processes through the inhibition of voltage sensitive Ca\(^{2+}\) channels[85,130]. On the pre-synaptic sites, GABA-B Rs appear to be negatively coupled to Ca\(^{2+}\) channels and block the Cav2.2 (N-type) and Cav2.1 (P and Q type) voltage gates channels via the action of G-proteins. They also have the ability to affect K\(^+\) channel directly without the involvement of a G protein. This controls the release of neurotransmitters from the pre-synaptic neurons and thus GABA-B Rs have the potential to regulate the onset of a stimulatory signal[85,130]. At the post-synaptic sites, these receptors mediate their effects via the G proteins through a slow hyperpolarization by inward rectification of K\(^{+}\) channels such as the GIRK channels as well as Ca\(^{2+}\) channels[131]. GABA-B Rs are, therefore, diverse both in their structural as well as functional mediation of inhibitory signals and show immense potential to regulate synaptic plasticity in association with sleep deprivation.
Serotonin, DA and norepinephrine receptors

The serotonin, DA and epinephrine group of GPCRs are very important in psychiatry considering their ligands are responsible for a wide range of CNS functions like cognitive learning and sleep-wake behaviours and are implicated in many neurological disorders. Serotonin receptors, also known as the 5-hydroxytryptamine (5-HTA) receptors are classified into 7 families (5-HT1 to 5-HT7), all of which are GPCRs except the 5-HT3 receptor[132]. These are further divided into their respective 14 sub-types [e.g., 5-hydroxytryptamine subtype 1A (5-HT1A) to 5-HT1F] based on their function and location[132]. They exhibit their function mainly through the Gα, Gq/11, and Gas proteins followed by modulation of Ca²⁺ concentrations via the regulation of adenylyl cyclase and cAMP or activation of phospholipase C producing IP3 and DAG[133]. In the CNS, 5-HT1 and 5-HT3 are associated with inhibitory signal transduction and the others are excitatory in nature; with pre- and post-synaptic localizations. For instance, the 5-HT1A receptors, which are widely implicated in depression, anxiety and learning, are found in both pre-synaptic (auto-receptors) and post-synaptic (heteroreceptors) locations[134]. Additionally, structural significance in this class of receptor through dimerization with each other and other GPCRs has also been indicated as a functional prerequisite for the cellular trafficking and functioning of 5-HT receptors[135].

DA receptors are GPCRs divided into 5 subtypes (D1-D5) and exhibit their action through the Gs/olf and Gi/o proteins by activating or inactivating adenylyl cyclase, respectively[136]. These receptors have a wide pre and post-synaptic distribution pattern in the CNS which contributes to the functional significance of DA[137]. For instance, the D1 and D2 receptors are abundantly localized in the striatum and substantia nigra region of the brain and are involved in the nigrostriatal pathway which is responsible for the control of bodily movements[138]. These are also majorly distributed in the ventral tegmental area (VTA) of the brain which is connected to the ventral striatum and this connection promotes reward-associated behaviour via the meso-limbic pathway[139]. Furthermore, along with a localization pattern observed in both pre- and post-synaptic regions, co-expression is also a general trend seen with DA receptors[140]. Lastly, the GPCRs that are the target for the catecholamine norepinephrine/noradrenaline (NA) are classified into alpha-1, alpha-2, beta-1, beta-2 and beta-3 receptors and they are localized on both pre- and post- synaptic sites. Alpha-1 follows the Gq protein mediated elevation of calcium levels by activation of phospholipase C producing IP3 and DAG and the other three subtypes work via the modulation of adenylyl cyclase pathway[141]. Interestingly, co-localization patterns have also been observed between DA and NA receptors in the rat prefrontal cortex [142]. Whether such co-localizations and/or dimerization have implications for pharmacological actions of therapeutic agents need further investigation.

SIGNIFICANCE OF GPCRS IN SYNAPTIC PLASTICITY AND SLEEP DEPRIVED CONDITION

Role of GABA-B and mGluRs

As introduced above, GPCRs appear to be involved in processes that regulate the transduction of action potential especially in the cells of the hippocampus region of the brain. Both LTD and LTP have demonstrated a pattern of being modulated by such regulatory molecules. It is already well established that Type I mGluRs induce LTD through the synaptic endocytosis of AMPA receptors in the hippocampus[143]. Similar lines of studies have pointed towards mGluRs in the CA1 pyramidal cells being involved in the induction of LTD through post-synaptic elevation of Ca²⁺ levels[144]. Interestingly, studies on the cerebellar Purkinje fibres have also demonstrated that activation of GABA-B Rs enhances LTD through an association with mGluRs[63]. Even though the mechanism of action of these receptors may be related, many of their molecular machineries have their own distinctive features while promoting cellular LTD. For instance, in the Purkinje fibres, LTD mediated via GABA-B relied on the beta gamma subunits of the G-protein whereas mGluR mediated LTD depended on the alpha subunit[63]. So far, it appears that these GPCRs are associated with the induction of LTD and reviewing other related studies, we know that NMDARs are the main mediators of LTD. Contrary to these findings, recent experimental work has put forward a type of LTP that is induced in the absence of NMDARs and requires the activation of Type I mGluRs[145]. As for the role of GABA-B Rs in LTP, its auto receptors depress their own activity through negative feedback and promotes
NMDAR mediated LTP, thus maintaining a balance between LTP and LTD[146]. Considering how the hippocampus and the cortex communicate and manifest their functions of learning and memory consolidation in the state of sleep through synaptic plasticity, it is reasonable to imply that these GPCRs modulating LTD and LTP are significant for this purpose. It also seems that co-expression of GPCRs and their heteromeric subunits is a repeated trend seen in studies that show the association of GPCRs with LTD[63,125,126]. The functionality of GPCR dimerization has also been put forward in earlier demonstrative studies which showed that the GABA-B1 subunit needs the GABA-B2 subunit to reach the cell surface and that GABA-B1 is responsible for agonist interaction whereas the B2 subunit works towards the G-protein activation[147,148]. Furthermore, gene knock-out or elimination studies for these subunits have also indicated the dependence on dimerization for the functioning of this receptor and that in some regions of the brain, GABA-B subunit gene elimination does not affect the functionality of the receptor which implies that it is replaceable by other G-protein activating molecules[149] and thus may have the potential to associate with other GPCRs to mediate their actions. Progressing from these initial findings, recent work done on rats that were sleep deprived using gentle prodding and tapping has demonstrated that induced LTD of population EPSP in the hippocampus requires activation of mGluRs and GABA-B Rs along with an increase in Ca²⁺ released from intracellular stores[150]. In sleep deprived conditions, western blot analysis and co-immunoprecipitation studies revealed that there were expression levels of mGlu1R and GABA-B1 receptor subunit as well as enhanced co-expression and heterodimerization between mGlu1R and GABA-B1 subunit and mGlu1R and GABA-B2 subunit[150].

**Role of 5-HT and other GPCRs**

As introduced earlier, GPCRs for serotonin are diverse both in their localization and functions. Their effects are mediated by calcium levels, which again is crucial for synaptic plasticity. Furthermore, 5-HT1A receptors are known to decrease NMDA receptor activity whereas 5-HT2A tends to increase it[151,152], thereby playing an opposing role in regulating NMDA mediated LTP. These are also co-expressed in hippocampal pyramidal cells and may direct studies towards their reciprocal action in regulating synaptic plasticity. Similar cross-talks have been suggested between 5-HT and GABA-B receptors[153], which as already discussed, are involved in modulating LTD, especially in sleep deprived conditions. In relation to sleep associated studies, 5-HT1 and 5-HT2 receptors appear to be involved in the regulation of REM sleep and sleep-wake behaviours[154]. To further strengthen the significance of these receptors, experimental work has indicated that sleep deprivation resulted in a temporary increase in the expression profile of the serotonin receptor, 5-HT1A through an enhanced suppressive effect of this receptor on the EPSPs recorded from CA1 pyramidal neurons[155]. Studies also indicated a heterodimerization of 5-HT2A-mGluR2 which is implicated in psychosis[156]. However, in rat hippocampus, while there was a change in the expression profile of 5-HT1A receptor, there was no significant heterodimerization with mGluRs following sleep deprivation[155]. Therefore, depending on the 5-HT receptor sub-type, it appears that sleep deprivation induces differential changes in their expression, co-localization and heterodimerization.

DA and NA receptors demonstrate somewhat analogous modulations of synaptic plasticity. In rat prefrontal cortex, DA receptors appear to modulate both LTD and LTP through glutamatergic synapses[157]. In the hippocampus, opposing effect of D1/D5 and D2/D3/D4 receptors on NMDA receptor modulation, regulates LTD and LTP[158]. Furthermore, D1/D5 have also been implicated in novelty acquisition through hippocampal LTD and LTP[159]. The induction of synaptic plasticity via NA receptors is more prevalent through the beta group receptors and they do so by modulating both NMDA and AMPA receptors in the pyramidal cells and perirhinal cortex[158,160]. Consequently, the modulation of synaptic plasticity via DA and NA receptors can be extended towards sleep associated functions. Studies in rats have indicated that sleep deprivation leads to a differential expression of DA receptors[161] and chronic sleep restriction results in changes in the density of NA receptors[162]. Considering the range of these GPCRs and their functions, their cross-talk properties and effects during the state of sleep through synaptic plasticity require further detailed investigation.

**Allosteric modulation of GPCRs**

Another intriguing and related aspect of GPCRs that signifies its structural implications in synaptic plasticity is the allosteric modulation of these receptors. Binding at sites other than the active site of the GPCR, it has been suggested that allosteric
modulators can selectively stimulate the homomeric and heteromeric forms of the receptor. For instance, in the CNS, heterodimerization is seen between mGlu2R and mGlu4R with allosteric modulation being targeted to the homomeric forms relative to the heteromeric receptors[165,164]. Furthermore, LTP and LTD and spatial learning at the excitatory synapses in the hippocampus have appeared to be enhanced by the positive allosteric modulation of mGlu5Rs[165]. Considering even GABA-B receptors demonstrate their sleep dependent synaptic changes by forming such heteromeric complexes and associations with mGluRs[150], the importance of such structural modifications and associations needs more attention with respect to sleep dependent functions.

**DISCUSSIONS**

Sleep is important for normal physiological and neurological functions and sleep deprivation affects majority of these functions. The negative effects of sleep deprivation can be better managed when we are able to specifically target the cellular and molecular machinery that drives these functions and, in this review, GPCRs are examined as a pivotal molecular machinery that impacts sleep associated functions through synaptic plasticity. For example, memory consolidation and learning as a model involving hippocampal-neocortical dialogue was looked at. The dialogue, in turn brings forth distinct electrophysiological patterns that are active during the SWS and REM sleep stages. Activity mediated synaptic plasticity has been widely correlated to sleep associated functions of cognitive learning, memory processing and over all brain development. GPCRs like GABA-B Rs and mGluRs are types of receptors that have shown a potential to be involved in these processes through their actions on LTD and LTP. What is even more intriguing is the fluctuation in their distribution, heterodimerization and co-localization following sleep deprivation, suggesting that these receptors can exist in one condition during normal sleep and change with sleep deprivation. GPCRs are one of the major cellular targets for drug interaction and therefore changes in the receptor expression profile can affect drug action.

For instance, the antipsychotic drug clozapine is one of the common and effective drugs used in treating disorders like schizophrenia. The cellular targets for this drug are GPCRs like 5-HT2A and DA D2 receptors with a higher affinity for the former receptor than the latter. It acts as an antagonist of DA D2 receptors in the mesolimbic pathway and although initially classified as an antagonist, clozapine is also known to be an inverse agonist of the 5-HT2A receptor present in the prefrontal cortex of the brain[166-168]. Other antipsychotic drugs like olanzapine, risperidone, aripiprazole, etc., have also been indicated to elicit their effects through these receptors[166]. Interestingly, 5-HT2A and mGlu2R receptors are suggested to form heterodimers which raise the possibility of testing if clozapine and other such antipsychotic drugs modulate similar complex formation[156]. Moreover, drugs like clozapine are known to have sleep-inducing effects[169] and since disorders like schizophrenia and other psychiatric diseases are associated with sleep disturbances[170], potential interplay between the molecular events of sleep deprivation and actions of antipsychotic drugs needs to be investigated. Other examples of drugs that act on GPCRs include ropinirole for Parkinson’s disease which acts as an agonist on DA D2 receptors[171] and baclofen, a GABA-B receptor agonist suggested for treatment of depression and anxiety[172]. Whether their actions change with sleep deprivation and vice versa also needs to be tested. With regards to synaptic plasticity, antipsychotics and antidepressants are suggested to affect LTP upon both acute and chronic use and interestingly, these effects are different for each of these situations which may suggest differences in network behaviour with acute vs chronic exposure to the drugs (see Table 2[173-211]). Moreover, their effect on LTD has not been sufficiently explored and hence similar studies on the differential effects of chronic vs acute use of such drugs on network behaviour are required with regards to LTD. Table 2 provides a brief summary on antipsychotic, antidepressant and anxiolytic drugs with their mechanisms of action and association with synaptic plasticity, learning and sleep.

The relation between a drug’s mechanism of action and sleep can be directly observed in drugs that are used for treating sleeping disorders like insomnia. Characterized by a lack of both quality and quantity of sleep, insomniac conditions are treated with many drugs that target GPCRs. For example, melatonin is a hormone that has been known to play a key role in the sleep-wake cycle and promoting sleep. Agonist drugs like ramelteon act on the melatonin GPCRs type 1 and 2 (MT1 and MT2) which results in reduced sleep latency in chronic insomniac patients[212].
Table 2 Changes in synaptic plasticity, learning and sleep associated with antipsychotic, antidepressant and anxiolytic drug therapy

<table>
<thead>
<tr>
<th>Class of drugs</th>
<th>Drug examples</th>
<th>Target GPCRs</th>
<th>Mechanism of action</th>
<th>Therapeutic application</th>
<th>Possible association with synaptic plasticity and learning</th>
<th>Implications in sleep and associated functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antipsychotics</td>
<td>Chlorpromazine</td>
<td>DA D1/D2/D3, 5-HT2A, Histamine H1, α1-Adrenergic, Muscarinic receptors.</td>
<td>1 Acts as an antagonist by blocking postsynaptic DA D2 receptors.</td>
<td>1 In the treatment of psychotic disorders like schizophrenia, bipolar disorder, acute mania, dementia, etc.</td>
<td>1 Synaptic plasticity: Acute treatment with haloperidol can impair induction of LTP but not its maintenance in rabbit hippocampus. Chronic treatment with haloperidol suggested to enhance LTP in corticostriatal brain slices of rats[174,175].</td>
<td>1 Associated with sedation and often promote sleep inducing effects.</td>
</tr>
<tr>
<td></td>
<td>Haloperidol</td>
<td>2 Can also block serotoninergic, histaminergic, cholinergic and noradrenergic receptors[173].</td>
<td>2 Reduce psychosis symptoms like hallucinations and delusion.</td>
<td>3 Learning Functions: Haloperidol and Chlorpromazine can impair spatial learning in rats[176,177].</td>
<td>(1) Haloperidol, loxapine and mesoridazine increase REM sleep latency. Withdrawal of drug causes significant disruption of sleep with reduced REM and total sleep; (2) No significant effect on SWS[178-180]; (3) Chlorpromazine enhances SWS along with a dose dependent effect on REM sleep - lower doses increasing and higher doses reducing or having no effect on REM sleep[181].</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loxapine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mesoridazine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clozapine</td>
<td>DA D2, 5-HT2A, 2C receptors</td>
<td>1 Antagonistic effects by blocking 5-HT2A/2C receptors and agonistic effects by promoting 5-HT1A receptor mediated actions.</td>
<td>1 In the treatment of psychotic disorders like schizophrenia, bipolar disorder, acute mania, dementia etc.</td>
<td>1 Synaptic Plasticity: Without tetanization, acute clozapine administration induces LTP in rat prefrontal cortex[183]; Acute administration of risperidone impairs LTP in rabbit hippocampus[184]. Chronic treatment with clozapine impairs LTP in rat prefrontal cortex upon chronic treatment with olanzapine[186,187].</td>
<td>1 Associated with an increase in the total amount of sleep but show an uneven trend in their stage specific effects.</td>
</tr>
<tr>
<td></td>
<td>Risperidone</td>
<td></td>
<td></td>
<td></td>
<td>(1) Clozapine increases REM sleep density and improves NREM sleep. However, its effect on SWS is inconsistent[188];</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Olanzapine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aripiprazole</td>
<td>DA D2 receptors</td>
<td>1 Partial agonist of DA D2 and 5-HT1A</td>
<td></td>
<td>(2) Olanzapine improves SWS as</td>
<td></td>
</tr>
<tr>
<td>2nd generation / atypical antipsychotic drugs</td>
<td>Clozapine</td>
<td>DA D2, 5-HT2A, 2C receptors</td>
<td>1 Antagonistic effects by blocking 5-HT2A/2C receptors and agonistic effects by promoting 5-HT1A receptor mediated actions.</td>
<td>1 In the treatment of psychotic disorders like schizophrenia, bipolar disorder, acute mania, dementia etc.</td>
<td>1 Synaptic Plasticity: Without tetanization, acute clozapine administration induces LTP in rat prefrontal cortex[183]; Acute administration of risperidone impairs LTP in rabbit hippocampus[184]. Chronic treatment with clozapine impairs LTP in rat prefrontal cortex upon chronic treatment with olanzapine[186,187].</td>
<td>1 Associated with an increase in the total amount of sleep but show an uneven trend in their stage specific effects.</td>
</tr>
<tr>
<td>3rd Generation</td>
<td>Aripiprazole</td>
<td>DA D2 receptors</td>
<td>1 Partial agonist of DA D2 and 5-HT1A</td>
<td></td>
<td>(2) Olanzapine improves SWS as</td>
<td></td>
</tr>
</tbody>
</table>
Antidepressants and anxiolytics

<table>
<thead>
<tr>
<th>Drugs</th>
<th>Receptors/Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fluoxetine</strong></td>
<td>5-HT-1A receptors: 1 Blocks the serotonin reuptake and increases serotonin action via 5-HT1A receptors.</td>
</tr>
<tr>
<td></td>
<td>In the treatment of depression, obsessive control disorders, eating disorders, substance use disorders etc.</td>
</tr>
<tr>
<td></td>
<td>1 Synaptic Plasticity: (1) Acute treatment with fluoxetine before stressful stimuli indicated to suppress stress induced LTD in CA1 region of rat hippocampus[194];</td>
</tr>
<tr>
<td></td>
<td>(2) Chronic fluoxetine treatment enhances neurogenesis dependent LTP in mice hippocampal granule cells and also upregulates Brain-Derived Neurotrophic Factor-LTP associated genes. Acute treatment showed no such upregulation[195,196].</td>
</tr>
<tr>
<td><strong>Mirtazapine</strong></td>
<td>5-HT1 and 2 receptors, a 2-adrenergic receptors: 1 Blocks 5-HT2 receptors in order to increase 5-HT1 mediated transmission.</td>
</tr>
<tr>
<td></td>
<td>2 Antagonist of α2-adrenergic receptors to increase adrenergic neurotransmission [193].</td>
</tr>
<tr>
<td></td>
<td>(3) Acute treatment of vortioxetine enhances LTP in the CA1 region of rat hippocampus by increasing pyramidal cell output[199] of memory processing [207,208];</td>
</tr>
<tr>
<td><strong>Vortioxetine</strong></td>
<td>5-HT1A, 1B, 1D, 5-HT7 receptors: Antagonist of 5-HT1D and H1-HT7 receptors and a partial agonist of 5-HT1A and 1B receptors[198].</td>
</tr>
<tr>
<td></td>
<td>In the treatment of Major Depressive Disorder. (5) Acute treatment of baclofen induces GABA-B receptor mediated LTD enhancement of a glutamate-evoked current in the cerebellar cortical neurons. It also enhances mCluRI-coupled intracerebroventricular Ca²⁺ release[63].</td>
</tr>
<tr>
<td><strong>Vilazodone</strong></td>
<td>5-HT1A receptors: Blocks the serotonin reuptake inhibitor and also acts as a partial agonist of 5-HT1A receptor and stimulates id[200].</td>
</tr>
<tr>
<td></td>
<td>2 Learning Functions: (1) Fluoxetine reverses memory impairment in rats and enhances memory processing in mice[203,204]. Chronic use of mirtazapine and vortioxetine improve cognitive functions in humans[205,206,208,209]; (5) Buspirone increases REM sleep density and latency[210];</td>
</tr>
<tr>
<td><strong>Buspirone</strong></td>
<td>5-HT1A, DA D2 auto receptors: Partial agonist of 5-HT1A receptor and antagonist of DA D2 auto receptors with low affinity[201].</td>
</tr>
<tr>
<td></td>
<td>For treating of Generalized Anxiety Disorder and side effects of (2) Buspirone and baclofen are implicated in the impairment of memory processing [207,208]; (6) Baclofen prolongs total sleep time with an increase in duration for REM sleep latency. Risperidone increases SWS but reduces REM sleep[180,190,191].</td>
</tr>
</tbody>
</table>
Intriguingly, the MT1 and MT2 receptors have been known to form dimers among themselves as well as heterodimers with other GPCRs like the serotonin 5-HT2C receptor\cite{213,214}. Thus, differences in ligand selectivity for these homomeric and heteromeric forms of the target receptors can have important consequences if the expression profile of such GPCRs change upon sleep deprivation. Other drug targets for insomnia include neuropeptides like orexins which promote wakefulness and are synthesized by neurons in the hypothalamus region of the brain. Drugs like suvorexant act as antagonists of the orexin GPCRs type 1 and 2 (OX1R and OX2R) thereby blocking their wakefulness promoting effects\cite{215}. These GPCRs known to be associated with other GPCRs like the cannabinoid receptor type 1 (CB1) and GABA-B receptors. They form heterodimers with the CB1 receptors\cite{216} which have been implicated in affecting memory formation and maintenance of mood. Moreover, orexin and GABA-B receptor activity are indicated to have a balancing interplay wherein inhibitory GABA-B modulates the wakefulness promoting properties in the orexin producing neurons\cite{217}. If expression profiles of GABA-B receptors change with sleep deprivation, can there be consequences for actions of orexins? In summary, changes in receptor profiles associated with sleep deprivation can have consequences for drug action and need a thorough investigation to understand the mechanisms involved so that CNS disorders can be treated with more rationally based therapeutics.

Unlike receptor redistribution and co-localization, heterodimerization can have serious implications for drug-receptor interactions and consequently, its action. On the other hand, if co-localization of different receptors leads to common transduction pathways, it can significantly affect the actions of such drugs. These consequences can vary depending on whether pre- or post-synaptic receptors are changed with sleep deprivation. Since this process is dynamic and the receptors may be in different states, an understanding of the mechanisms involved in clearly needed. Moreover, since GPCRs among 5-HT, norepinephrine, DA, etc., may be affected, drug therapy involving antidepressants and antipsychotics should take the occurrence of receptor plasticity with sleep deprivation into account as drug-receptor interactions may be changed (Figure 1).
Figure 1 G-protein coupled receptors: Potential key-players in psychiatric therapeutics through modulation of synaptic plasticity in sleep deprivation. A correlation between sleep deprivation, G-protein coupled receptors (GPCRs) and synaptic plasticity. Memory consolidation and learning are strengthened during sleep and synaptic plasticity, through long term potentiation (LTP) and long term depression (LTD), can regulate the processing of these functions. Upon sleep deprivation, GPCRs like gamma-amino butyric acid-B, metabotropic glutamate and 5-hydroxytryptamine receptors experience changes in their expression profiles in terms of distribution, dimerization and co-localization with their own or different subtypes of receptors. Changes in these and the associated transduction pathways can in turn affect LTP and/or LTD, causing alterations in neuronal network activity. Since GPCRs are one of the most dynamic cell membrane proteins, they are the targets of many drugs used in various psychoneurological disorders. Therefore, a change in the expression profile of these receptors through sleep deprivation can reasonably alter the way in which these drugs exhibit their therapeutic action. Conversely, therapeutic strategies themselves can affect sleep causing a change in the GPCR expression profile and the downstream cascade. GPCR: G-protein coupled receptors; GABA-B: Gamma-amino butyric acid-B; LTD: Long term depression; LTP: Long term potentiation; mGlu: metabotropic glutamate; 5-HT: 5-hydroxytryptamine; cAMP: Cyclic Adenosine Monophosphate; DAG: Diacylglycerol; IP3: Inositol-1,4,5-triphosphate.

2-way possibility of effects

Can affect the action of drugs and therapeutic strategies

Changes in expression profile

Distribution pattern, dimerization and co-localization with same/different receptors

Affects downstream signalling molecules like cAMP, adenylyl cyclase, DAG, IP3, protein kinases etc.

Regulates intracellular calcium release

Psychoneurological disorders like: Parkinson’s disease, schizophrenia and other psychotic disorders anxiety, depression

Major cellular targets for drugs/therapeutic interventions

Affects memory consolidation and learning

Future scope for psychoneurological disorders and therapeutics

Apart from memory consolidation and learning, GPCRs are also implicated in other psychoneurological occurrences. Around 60% of drugs target membrane proteins and 30% of them are GPCRs[219]. They are especially important as targets for antipsychotic and antidepressant drugs, considering some of their ligands like serotonin, DA and norepinephrine are associated with pathways that are important in psychiatry. Additionally, mGluRs and GABA-B Rs have an almost ubiquitous involvement in major cellular pathways which need to be better understood since they are affected in neurological disorders like Alzheimer’s disease, Parkinson’s disease, stress related disorders, etc.[220,221]. Changes in sleep patterns are often observed in many neurological disorders and considering these major GPCRs are altered during the state of sleep, an understanding of this correlation can affect the way we approach therapeutics in this area. For example, even though there was a temporary elevation specific receptor types occurs with sleep deprivation[150]. Hence, the driving molecular factors which includes, specific genes and protein interactions, need a more coherent identification. For instance, the signalling linked with GPCRs are tightly controlled by a family of regulator of G protein signalling proteins in the brain[218]. Furthermore, since metabotropic receptors (e.g., mGluRs) affect the function of ionotropic receptors like NMDA and AMPA, similar studies can be extended for the metabotropic GABA-B receptor and its ionotropic counterparts like GABA-A and GABA-C in relation to sleep deprivation induced changes. Considering that GPCRs are involved in several CNS functions, there is a wide scope for studying the mechanisms involved in their plasticity. Another interesting area of research is the regulation of synaptic plasticity by GPCRs elicited intracellularly[119] and this could also be explored further in order to connect the transportation of GPCR subunits from intracellular compartments to the cell membrane and the processes that take place in these compartments which can give rise to heterodimers and co-localization.
observed in the expression of the serotonin 5-HT1A receptor upon sleep deprivation [155], it is universally correlated to disorders like depression, anxiety and psychosis and changes in its expression profile upon sleep deprivation does warrant for further studies on this GPCR and its other subtypes. Additionally, related GPCRs like the DA receptor D1 are also known to interact directly with LTP inducing NMDA receptors [118] which are important in many neurological disorders. Furthermore, since the D1 and D2 DA receptors are involved in the nigrostriatal and mesolimbic pathways, changes in their expression profiles affect movements and behavioural patterns that are also implicated in Parkinson’s disease and schizophrenia [222,223]. Therefore, DA GPCRs are highly potential drug targets and sleep associated changes in their expression profile needs more attention. Drugs used in psychiatry for such disorders, e.g., aripiprazole, are known to target 5-HTA and DA receptors [224] and if the expression profile of these receptors is changing upon sleep deprivation, could it also affect the action of such drugs? Disorders like schizophrenia and depression often encompass sleep deprivation and drugs used for these conditions which target such GPCRs, may get negatively affected if receptor plasticity (co-localization and/or heterodimerization with other receptor subunits, etc.) occurs with associated sleep disorders. For example, the antidepressant drug mirtazapine exhibits its action through GPCRs by blocking 5-HT2 and adrenergic α1 receptors, leading to an increase in the activation of 5-HT1A receptor mediated activity as well as NA release, respectively [193]. This drug is known to have sleep-promoting properties and has been associated with problems like day-time somnolence [225] potentially disrupting the normal sleep cycle. Interestingly, there are off-label therapies that employ antidepressants like mirtazapine in treating sleeping disorders due to its sedative effect [226]. Since antidepressants and antipsychotic drugs appear to directly or indirectly change expression profile of these receptors is changing upon sleep deprivation, could it also affect the action of such drugs? Disorders like schizophrenia and depression often encompass sleep deprivation and drugs used for these conditions which target such GPCRs, may get negatively affected if receptor plasticity (co-localization and/or heterodimerization with other receptor subunits, etc.) occurs with associated sleep disorders. For example, the antidepressant drug mirtazapine exhibits its action through GPCRs by blocking 5-HT2 and adrenergic α1 receptors, leading to an increase in the activation of 5-HT1A receptor mediated activity as well as NA release, respectively [193]. This drug is known to have sleep-promoting properties and has been associated with problems like day-time somnolence [225] potentially disrupting the normal sleep cycle. Interestingly, there are off-label therapies that employ antidepressants like mirtazapine in treating sleeping disorders due to its sedative effect [226]. Since antidepressants and antipsychotic drugs appear to directly or indirectly change activity at GPCRs, drug therapy for these disorders can affect sleep while sleep disturbances can also necessitate changes in that therapy. Therefore, sleep disturbance-induced plasticity and cross-talk between GPCRs can have consequences for drug therapy, the mechanisms for which need to be thoroughly examined. Lastly, allosteric modulation, which is another upcoming molecular interaction that is important for drug designing, can be applied to the homo and heterodimerization of GPCRs for various conditions. Hence, studies on structural, functional, receptor co-localization and remodelling, etc., will yield new insights into mechanisms involved and help improve therapeutics for a variety of CNS disorders, including in psychiatry.

CONCLUSION

Conclusion and future scope are included in the main text under heading "Introduction" above.

REFERENCES

Parmar S et al. GPCRs, synaptic plasticity and sleep deprivation

10.1016/s1388-2457(03)00238-4


18 Smith C, Smith D. Ingestion of ethanol just prior to sleep onset impairs memory for procedural but not declarative tasks. Sleep 2003; 26: 185-191 [PMID: 12683478]


29 Muller GE, Pilzecker A. Experimental contributions to the theory of memory. Z Psychol Z Angew Psychol 1900; 1: 1-300


34 Diekelmann S, Born J. The memory function of sleep. Nat Rev Neurosci 2010; 11: 114-126 [PMID: 20046194 DOI: 10.1038/nrn2762]


DOI: 10.1016/s1388-2457(03)00238-4


Fagni L, Malenka RC. NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD). Neuron 2009; 61: 587-596 [PMID: 19249278 DOI: 10.1016/j.neuron.2009.01.011]

Hebbian DO. The organization of behavior: A neuropsychological theory. New York: John Wiley and Sons, 1949


Lee HK, Barbarosie M, Kameyama K, Bear MF, Huganir RL. Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity. Nature 2000; 405: 955-959 [PMID: 10879537 DOI: 10.1038/35016089]


Bliss TV, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature 1993; 361: 31-39 [PMID: 8421494 DOI: 10.1038/36101a0]


Tsumoto T. Long-term depression in cerebral cortex: a possible substrate of “forgetting” that should


80 Harney SC, Rowan M, Anwyll R. Long-term depression of NMDA receptor-mediated synaptic transmission is dependent on activation of metabotropic glutamate receptors and is altered to long-term potentiation by low intracellular calcium buffering. *J Neurosci* 2006; 26: 1128-1132 [PMID: 16436598 DOI: 10.1523/JNEUROSCI.2753-05.2006]


89 Campbell IG, Guinan MJ, Horowitz JM. Sleep deprivation impairs long-term potentiation in rat hippocampal slices. *J Neurophysiol* 2002; 88: 1073-1076 [PMID: 12163556 DOI: 10.1152/jn.2002.88.2.1073]

90 McDermott CM, LalHoste GJ, Chen C, Musto A, Bazan NG, Magee JC. Sleep deprivation causes


105 **Scharf MT**, Woo NH, Lattal KM, Young JZ, Nguyen PV, Abel T. Protein synthesis is required for the enhancement of long-term potentiation and long-term memory by spaced training. *J Neurophysiol* 2002; 87: 2770-2777 [PMID: 12037179 DOI: 10.1152/jn.2002.87.6.2770]


113 **Cirelli C**, Tononi G. The search for the molecular correlates of sleep and wakefulness. *Sleep Med Rev* 2001; 5: 397-408 [PMID: 12531005 DOI: 10.1016/s1368-8140(01)00060-x]


Jones KA, Borowsky B, Tamm JA, Craig DA, Durkin MM, Mai D, Yao WJ, Johnson, Gunwaldsen C, Huang LY, Tang C, Shen Q, Salon JA, Morse K, Branchek TA, Gerald C. GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2. *Nature* 1998; 396: 674-679 [PMID: 9872313 DOI: 10.1038/25348]


Baik JH. Dopamine signaling in reward-related behaviors. *Front Neural Circuits* 2013; 7: 152 [PMID: 24130517 DOI: 10.3389/fncir.2013.00152]


Davies CH, Starkey SJ, Pozza MF, Collingridge GL. GABA autoreceptors regulate the induction of LTP. *Nature* 1991; 349: 609-611 [PMID: 1847993 DOI: 10.1038/349690a0]


Tadavarty R, Raijput PS, Wong JM, Kumar U, Sastry BR. Sleep-deprivation induces changes in GABA(B) and mGlur receptor expression and has consequences for synaptic long-term depression. *PLoS One* 2011; 6: e24933 [PMID: 21980366 DOI: 10.1371/journal.pone.0024933]


Parmar S et al. GPCRs, synaptic plasticity and sleep deprivation


170 Monti JM, Monti D. Sleep in schizophrenia patients and the effects of antipsychotic drugs. *Sleep Med Rev* 2004; 8: 133-148 [PMID: 15033152 DOI: 10.1016/S1087-0792(02)00158-2]


187 Shim SS, Hammond MS, Tatsuka C, Feng II. Effects of 4-weeks of treatment with lithium and
olanzapine on long-term potentiation in hippocampal area CA1. *Neurosci Lett* 2012; **524**: 5-9 [PMID: 22750162 DOI: 10.1016/j.neulet.2012.06.047]


194 Han H, Dai C, Dong Z. Single fluoxetine treatment before but not after stress prevents stress-induced hippocampal long-term depression and spatial memory retrieval impairment in rats. *Sci Rep* 2015; **5**: 12667 [PMID: 26218751 DOI: 10.1038/srep12667]


200 Cruz MP, Vilazodone HCI (Viibryd): A Serotonin Partial Agonist and Reuptake Inhibitor For the Treatment of Major Depressive Disorder. *P T* 2012; **37**: 28-31 [PMID: 22346333]


207 Ott GE, Rao U, Lin KM, Gertsik L, Poland RE. Effect of treatment with bupropion on EEG sleep:


Agmatine as a novel candidate for rapid-onset antidepressant response

Ana Paula Valverde, Anderson Camargo, Ana Lúcia S Rodrigues

ORCID number: Ana Paula Valverde 0000-0002-6493-516; Anderson Camargo 0000-0002-1632-4865; Ana Lúcia S Rodrigues 0000-0001-6285-8780.

Author contributions: All the authors performed the literature search and wrote the manuscript.

Conflict-of-interest statement: Authors declare no conflict of interests for this article.

Supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico, No. 310113/2017-2; and the CNPq Research Productivity Fellowship.

Country/Territory of origin: Brazil

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report’s scientific quality classification

Grade A (Excellent): A
Grade B (Very good): 0
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in

Abstract

Major depressive disorder (MDD) is a disabling and highly prevalent mood disorder as well as a common cause of suicide. Chronic stress, inflammation, and intestinal dysbiosis have all been shown to play crucial roles in the pathophysiology of MDD. Although conventional antidepressants are widely used in the clinic, they can take weeks to months to produce therapeutic effects. The discovery that ketamine promotes fast and sustaining antidepressant responses is one of the most important breakthroughs in the pharmacotherapy of MDD. However, the adverse psychomimetic/dissociative and neurotoxic effects of ketamine discourage its chronic use. Therefore, agmatine, an endogenous glutamatergic modulator, has been postulated to elicit fast behavioral and synaptogenic effects by stimulating the mechanistic target of rapamycin complex 1 signaling pathway, similar to ketamine. However, recent evidence has demonstrated that the modulation of the NLR family pyrin domain containing 3 inflammasome and gut microbiota, which have been shown to play a crucial role in the pathophysiology of MDD, may also participate in the antidepressant-like effects of both ketamine and agmatine. This review seeks to provide evidence about the mechanisms that may underlie the fast antidepressant-like responses of agmatine in preclinical studies. Considering the anti-inflammatory properties of agmatine, it may also be further investigated as a useful compound for the management of MDD associated with a pro-inflammatory state. Moreover, the fast antidepressant-like response of agmatine noted in animal models should be investigated in clinical studies.

Key Words: Agmatine; Fast-acting antidepressants; Ketamine; Major depressive disorder; Microbiota-gut-brain axis; Neuroinflammation

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.
Valverde AP et al. Agmatine as a rapid-onset antidepressant

Core Tip: One of the main challenges in the advancement of antidepressant therapy is the establishment of safe and effective fast-acting antidepressants. Ketamine is a prototype for rapid-onset antidepressant responses. Agmatine has been shown to produce fast antidepressant-like effect by stimulating mechanistic target of rapamycin complex 1 signaling pathway, similar to ketamine. Moreover, NLR family pyrin domain containing 3 and microbiota-gut-brain axis may be novel targets for fast antidepressant responses. These targets have also been postulated to play a role in the antidepressant effect of both ketamine and agmatine.

Citation: Valverde AP, Camargo A, Rodrigues ALS. Agmatine as a novel candidate for rapid-onset antidepressant response. World J Psychiatr 2021; 11(11): 981-996
DOI: https://dx.doi.org/10.5498/wjp.v11.i11.981

INTRODUCTION

Major depressive disorder (MDD) affects more than 300 million people and is a major cause of disability and suicidal deaths worldwide[1]. Despite the high prevalence of this psychiatric disorder, its neurobiological basis remains to be fully elucidated, and its treatment remains a challenge. Although the monoaminergic hypothesis of MDD has played a crucial role in its pharmacotherapy, it is now considered overly simplistic [2]. One of the limitations of this hypothesis is the fact that the drugs currently used to treat MDD exert their therapeutic effect only after 3-4 wk and many patients fail to respond to these drugs. Antidepressants also feature side effects that may include nausea, dizziness, insomnia, weight gain, sleep disturbances, and sexual dysfunction. This scenario underscores the strong demand for developing novel antidepressants with a fast antidepressant response, better efficacy, and fewer adverse effects[3,4].

In this regard, the year 2000 marked a turning point in the history of MDD pharmacotherapy, and Berman et al[5] showed for the first time that a single dose of ketamine, an N-methyl-d-aspartic acid (NMDA) receptor antagonist, produced a fast-acting antidepressant response in MDD patients. This discovery was reinforced by several subsequent studies that demonstrated that a single dose of ketamine elicited rapid and long-lasting antidepressant effects, even in treatment-resistant patients with suicidal ideation[6-8]. Although the discovery of ketamine is considered the major breakthrough in MDD pharmacotherapy and opened new perspectives to manage refractory patients at risk of suicide, ketamine has knock-on effects that limit its widespread clinical use[9]. For these reasons, ketamine has emerged as a prototype for screening novel fast-acting antidepressant agents. Agmatine, an endogenous neuromodulator, shares some common molecular mechanisms with ketamine and the ability to elicit fast antidepressant-like effects in preclinical studies[10]. Therefore, agmatine could be a novel candidate to elicit fast antidepressant responses.

Therefore, this narrative review presents evidence that agmatine has a rapid antidepressant effect and provides an overview of the possible mechanisms underlying this effect versus those already described for ketamine. The PubMed, SCOPUS, and SciSearch databases were searched for original manuscripts and contemporary reviews published in English.

KETAMINE AS A PROTOTYPE RAPID-ACTING ANTIDEPRESSANT AGENT

A milestone for the development of drugs with a rapid antidepressant effect has emerged over the past two decades[5,6]. Berman et al[5] demonstrated for the first time that a single subanesthetic dose of ketamine elicited fast (within 4 h) and long-lasting (for up to 3 d) antidepressant effects in MDD patients. These findings were reinforced and extended by Zarate et al[6], who showed that a single dose of ketamine effectively improved the MDD symptoms in treatment-refractory patients, as evidenced by an effect observed within 110 min that was sustained for up to 7 d. Importantly, a rapid resolution of suicidal ideation after a single infusion of ketamine in patients with treatment-resistant MDD was also observed, supporting the premise that ketamine has
rapid beneficial effects even in severely depressed individuals at risk of suicide[7,8]. Due to the ability of ketamine to alleviate MDD symptoms, several research groups have investigated its underlying molecular mechanisms.

Although ketamine has been well characterized as an NMDA receptor antagonist, its molecular effects extend far beyond this level. It is worth noting that this drug has a window of therapeutic efficacy that surpasses its short half-life of a few hours[11,12]. Experimental studies have provided evidence that the antidepressant-like effect of ketamine depends on mechanistic target of rapamycin complex 1 (mTORC1) activation, a key pathway required for protein synthesis-dependent synaptic plasticity [13-15]. It has been postulated that ketamine, by antagonizing NMDA receptors in GABAergic interneurons, attenuates the inhibitory action of this system on glutamatergic tonus. This blockade causes the disinhibition of pyramidal cells, which causes a burst of glutamatergic transmission[11]. In particular, the glutamate released under these conditions preferentially stimulates alpha-amino-3-hydroxy-methyl-5-4-isoxazole propionic acid (AMPA) receptors, promoting a transient sodium influx that depolarizes the cell and activates the voltage-dependent calcium channels (VDCC). This event causes the exocytosis of synaptic vesicles containing the brain-derived neurotrophic factor (BDNF) in the synaptic cleft, as a result of calcium influx by VDCC [16].

BDNF activates tropomyosin receptor kinase B (TrkB), which stimulates the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt)/mTORC1 signaling pathway[12]. In turn, mTORC1 controls the translation of proteins involved in new dendritic spine formation and synaptogenesis [e.g., postsynaptic density protein-95 kDa (PSD-95), glutamate AMPA receptor subunit 1 (GluA1), and synapsin] by activating the 70-kDa ribosomal protein S6 kinase (p70S6K) and inhibiting the eukaryotic initiation factor 4E-binding protein[13,14,17]. Although mTORC1-dependent synaptogenesis induced by ketamine was first demonstrated in the prefrontal cortex of rodents[13,15], it has been shown that similar events also occur in the hippocampus[18-20]. Therefore, these findings suggest that targeting the mTORC1-driven signaling pathway may produce rapid-onset and long-lasting antidepressant-like responses. The molecular mechanisms underlying ketamine’s antidepressant responses are shown in Figure 1.

Importantly, almost 20 years after the groundbreaking discovery that ketamine effectively produces rapid and sustained antidepressant effects, particularly in March 2019, the United States Food and Drug Administration approved the use of (S)-ketamine nasal spray (Spravato™) for treatment-resistant MDD. In the same year, (S)-ketamine nasal spray was approved for use in treatment-resistant depression in Europe[21]. Despite the fast and long-lasting antidepressant effects of ketamine, there is much concern about its abuse potential and serious adverse effects[9]. For this reason, ketamine is only available through a restricted distribution system, limiting its widespread clinical use. Although there are some drawbacks associated with its use, it may serve as a prototype for screening fast antidepressant agents. Given this scenario, the search for ketamine-like compounds has emerged as a promising therapeutic strategy. In this regard, our research group and others have shown that agmatine is also able to produce fast antidepressant responses and shares some mechanisms of action with ketamine[10].

**AGMATINE AS A NOVEL CANDIDATE FOR FAST ANTIDEPRESSANT RESPONSES**

Agmatine, a cationic amine produced from the L-arginine in a reaction catalyzed by the enzyme arginine decarboxylase, is widely distributed in human tissues, including the brain[22-24]. For almost a century, it was wrongly believed that agmatine was produced by bacteria, plants, and fish but not mammals[25]. Agmatine was “rediscovered” in 1994 during the search for an endogenous ligand for imidazoline binding sites[24]. In this study, a molecule was isolated from the mammalian brain and identified as agmatine[24]. This was the starting point for many studies that have evaluated the biological properties and possible beneficial effects of agmatine on a wide variety of diseases[26].

Soon after providing evidence of the presence of agmatine in mammalian nervous tissue, Piletz et al[27] documented its neuroprotective effects. The neuroprotective effects of agmatine reportedly involve protection mechanisms against excitotoxicity, since agmatine may block NMDA receptors and inhibit the increase in intracellular calcium concentrations in different neuronal cell cultures[28-30]. The NMDA receptor
Figure 1 Supposed intracellular signaling pathway implicated in ketamine’s antidepressant-like effects. A: Ketamine antagonizes N-methyl-D-aspartic acid receptors in GABAergic interneurons, which in turn attenuate the inhibitory action of this system on glutamatergic tonus. Subsequently, the disinhibition of pyramidal cells results in a burst of glutamatergic transmission; B: The glutamate released in the synaptic cleft preferentially stimulates alpha-amino-3-hydroxy-methyl-5-4-isoxazole propionic acid (AMPA) receptors, which promotes a transient sodium influx that depolarizes the neurons and activates voltage-dependent calcium channels. C: This event causes the exocytosis of synaptic vesicles containing brain-derived neurotrophic factor (BDNF) in the synaptic cleft[16,122]. BDNF culminates in protein kinase B activation that can phosphorylate and activate mechanistic target of rapamycin complex 1 (mTORC1). In turn, mTORC1 phosphorylates the 70-kDa ribosomal protein S6 kinase at Thr389, which regulates synaptic protein synthesis such as AMPA receptor subunit 1 and postsynaptic density-95, which contribute to dendritic spine formation and synaptogenesis[17,13]. NMDA: N-methyl-D-aspartic acid; AMPA: Alpha-amino-3-hydroxy-methyl-5-4-isoxazole propionic acid; BDNF: Brain-derived neurotrophic factor; VDCC: Voltage-dependent calcium channels; PI3K: Phosphatidylinositol 3-kinase; Akt: Protein kinase B; TrkB: Tropomyosin receptor kinase B; GluA1: Glutamate AMPA receptor subunit 1; mTORC1: mechanistic target of rapamycin complex 1; PSD-95: postsynaptic density protein-95 kDa.

is an ion channel controlled by glutamatergic excitation, which is essential for the normal functioning of the central nervous system (CNS), including cognitive function, locomotion, and breathing[31-33]. This type of receptor is located on the membranes of neuronal and glial cells[34], and is implicated in the development and maintenance of acute and chronic diseases of the CNS, such as stroke, Parkinson’s disease, Alzheimer’s disease, MDD, and schizophrenia[34]. Under excitotoxic conditions, the increase in extracellular glutamate intensifies NMDA receptor activation, causing an influx of Ca\(^{2+}\) and Na\(^{+}\)[34]. NMDA receptor stimulation also activates nitric oxide (NO) production pathways by activating nitric oxide synthase, thereby generating NO, one of the main mediators of cellular death[35,36]. The inhibition of NO synthesis is potentially beneficial in the treatment of brain disorders associated with its overproduction. Several studies have shown that the neuroprotective properties of agmatine in several neurodegenerative diseases are related to its ability to antagonize NMDA receptors and inhibit NO synthesis as well as its potential to counteract the effects of oxidative stress[37].

The first evidence of the antidepressant effects of agmatine was reported in a study that examined its impact on behavioral tests related to depression (immobility time in the tail suspension test and forced swimming test) in mice[38]. Since then, other studies have confirmed the antidepressant efficacy of agmatine in behavioral tests in rodents[39-41]. Subsequent studies implicated several molecular targets in the antidepressant effect of agmatine, namely the modulation of: (1) K\(^{+}\) channels[42]; (2) NO synthesis[43,44]; and (3) several neurotransmitter receptors including NMDA receptors[38,45], AMPA receptors[46,47], GABA receptors[48], serotonin receptors[49,50], and opioid system receptors[51].
In 2010, a human clinical trial showed the safety of oral agmatine[52]. In 2013, Shopsin et al[53] provided the first evidence that agmatine may effectively treat MDD, but this study included only three patients. None of these three patients treated with agmatine relapsed after the joint administration of a serotonin-depleting drug, indicating that the mechanism underlying the antidepressant action of agmatine is likely unrelated to the serotonergic system[53]. It was also reported that 8-wk treatment with the standard antidepressant bupropion normalized plasma agmatine levels[54]. In brain autopsies of chronically depressed patients, Bernstein et al[55] reported a significant increase in agmatinase immunoreactivity in hippocampal neurons, suggesting the role of the agmatinergic system in MDD pathophysiology. However, it was not possible to determine the exact reason why levels of this enzyme increased in hippocampal neurons due to the use of antidepressants.

In 2018, a gas chromatography–mass spectrometry study quantified agmatine levels in the brains of post-mortem humans who died by suicide and showed reduced agmatine levels in the suicide cortex regardless of these individuals formerly meeting the criteria for MDD versus controls[56].

Weiss et al[57] presented evidence of the activity of the agmatinergic system in habenular nuclei and investigated the actions of agmatine and agmatinase in the rat and human habenular systems. It is important to highlight that the role of habenular nuclei in mental disorders, including MDD, has already been considered[58,59]. In this study, agmatine was demonstrated responsible for the strong increase in the spontaneous action potential of medial habenular neurons by activating type II imidazoline receptors. It was also reported that increased activity of the agmatinergic system in habenular nuclei may strengthen the dopaminergic activity of the midbrain. This evidence suggests dysregulation in the habenular-interpeduncular axis in patients with MDD[57].

In summary, these results present the possible role of agmatine in the neurobiology of MDD and highlight the possible benefits of agmatine as antidepressant therapy.

Recent evidence has also demonstrated the possible fast antidepressant-like actions of agmatine (Figure 2)[10]. In this context, Neis et al[46] reported that the antidepressant-like effect of agmatine administered orally to mice subjected to the tail suspension test is dependent on the modulation of molecular targets associated with the fast antidepressant-like effect displayed by ketamine. In particular, the antidepressant-like effect elicited by the acute administration of agmatine in the tail suspension test appears to involve inhibition of NMDA receptors since it enhanced the antidepressant potency of MK-801 (an NMDA receptor antagonist) up to 100-fold[60]. Moreover, the antidepressant-like effect of agmatine in the tail suspension test is dependent on AMPA and TrkB receptor activation since the administration of 6,7-dinitroquinoxaline-2,3-dione (DNQX; an AMPA receptor antagonist) or K-252a (a TrkB receptor antagonist) completely abolished its antidepressant-like response[46]. A single dose of agmatine also increased BDNF levels in the prefrontal cortex of mice, and its antidepressant-like effect in the tail suspension test was abrogated by the administration of anti-BDNF antibody. Of note, the antidepressant-like effect of agmatine is also dependent on PI3K/Akt/glycogen synthase kinase-3β (GSK-3β)/mTOR signaling. In particular, the administration of LY294002 (a PI3K inhibitor) or rapamycin (a selective mTOR inhibitor) completely abrogated the behavioral responses of agmatine in the tail suspension test. Combined treatment with a sub-effective dose of agmatine and lithium chloride (a non-selective GSK-3β inhibitor) or AR-A014418 (a selective GSK-3β inhibitor) produced an antidepressant-like effect in the tail suspension test[46]. Importantly, these behavioral responses were accompanied by an increase in BDNF, GluA1, and PSD-95 immunoccontent in the prefrontal cortex of mice[46].

Supporting the assumption that agmatine could elicit a fast antidepressant-like effect, a study by Neis et al[61] demonstrated that a single dose of agmatine effectively reversed the depressive-like behavior induced by chronic unpredictable stress. In this study, mice were exposed to the stress protocol for 14 d and received a single oral dose of agmatine, ketamine, or fluoxetine. The results indicated that a single dose of agmatine or ketamine (after 24 h), but not fluoxetine, counteracted the depressive-like behavior induced by the stress protocol in the tail suspension test[61].

Expanding upon these findings, Neis et al[62] reinforced the ability of agmatine to rapidly reverse the depressive-like behavior induced by the 21-d administration of corticosterone, a pharmacological model of stress in mice. In the tail suspension test, a single dose of agmatine or ketamine abolished the depressive-like behavior of mice chronically exposed to corticosterone. In addition, treatment increased GluA1 immunoccontent in the hippocampus of control animals[62]. Notably, a single dose of fluoxetine did not produce the same effects as ketamine or agmatine[62].
Figure 2 Putative signaling pathways implicated in the rapid-acting antidepressant-like effects of agmatine. A: The antidepressant-like effect elicited by the acute administration of agmatine in the tail suspension test appears to involve inhibition of N-methyl-D-aspartic acid (NMDA) receptors since it enhanced the antidepressant potency of MK-801 (an NMDA receptor antagonist) by up to 100-fold [60]. Moreover, agmatine’s antidepressant-like effect in the tail suspension test is dependent on the activation of alpha-amino-3-hydroxy-methyl-5-4-isoxazole propionic acid (AMPA) and tropomyosin receptor kinase B (TrkB) receptors since the administration of 6,7-dinitroquinoxaline-2,3-dione (DNQX; an AMPA receptor antagonist) or K-252a (a TrkB receptor antagonist) completely abolished its antidepressant-like response. The antibody anti-brain-derived neurotrophic factor (BDNF) also abolished the antidepressant-like effect elicited by agmatine in the tail suspension test [46].

Of note, the antidepressant-like effect of agmatine is also dependent on phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt)/glycogen synthase kinase-3β/mechanistic target of rapamycin (mTOR) signaling. In particular, the administration of LY294002 (a PI3K inhibitor) or rapamycin (a selective mTOR inhibitor) completely abrogated the behavioral responses of agmatine in the tail suspension test. The combined treatment with a sub-effective dose of agmatine and lithium chloride (10 mg/kg, po; a non-selective GSK-3β inhibitor) or AR-A014418 (a selective GSK-3β inhibitor) produced an antidepressant-like effect in the tail suspension test [46]. Importantly, these behavioral responses were accompanied by an increase in the BDNF, glutamate AMPA receptor subunit 1 and postsynaptic density protein-95 kDa immunocontent in the prefrontal cortex of the mice [46].

B: Reinforcing the notion that ketamine and agmatine share common behavioral responses and molecular targets, the ability of agmatine to potentiate the antidepressant and synaptic actions of ketamine was also demonstrated [47]. In particular, the combined administration of subthreshold doses of agmatine and ketamine produced a fast (starting in 1 h) and sustained (lasting up to 7 d) antidepressant-like effect in the tail suspension test. These behavioral responses were associated with stimulation of the Akt/70-kDa ribosomal protein S6 kinase signaling pathway and increased synaptic protein synthesis in the prefrontal cortex in a time-dependent manner. More importantly, the combined administration of sub-effective doses of agmatine and ketamine raised the dendritic arbor and spine densities and effectively remodeled the dendritic spine architecture in the prefrontal cortex [47]. NMDA: N-methyl-D-aspartic acid; AMPA: Alpha-amino-3-hydroxy-methyl-5-4-isoxazole propionic acid; BDNF: Brain-derived neurotrophic factor; VDCC: Voltage-dependent calcium channels; PI3K: Phosphatidylinositol 3-kinase; Akt: Protein kinase B; TrkB: Tropomyosin receptor kinase B; GluA1: Glutamate AMPA receptor subunit 1; mTORC1: mechanistic target of rapamycin complex 1; PSD-95: postsynaptic density protein-95 kDa; GSK-3β: Glycogen synthase kinase-3β.
and agmatine share common molecular targets and could work in tandem to protect neuronal cells from the harmful effects of corticosterone by activating Akt/mTORC1/p70S6K signaling, resulting in synaptic protein expression[11,65].

Reinforcing these findings, Freitas et al[47] investigated the ability of agmatine to potentiate the effects of the antidepressant and synaptic actions of ketamine in mice. Of special interest, the combination of single subthreshold doses of ketamine and agmatine exerted antidepressant-like and pro-synaptogenic actions in a time-dependent manner. In particular, agmatine plus ketamine produced fast (1 and 24 h) and sustained (7-d) antidepressant-like effects in the tail suspension test[47]. Furthermore, this combined treatment increased p70S6K phosphorylation and GluA1 immunoreactivity in the prefrontal cortex 1 h after treatment. This same protocol increased the PSD-95 immunoreactivity, an effect that persisted for up to 7 d. The combined treatment also increased the complexity of the dendritic branches after 24 h, and this effect lasted up to 7 d. Likewise, ketamine plus agmatine treatment effectively increased the dendritic spinal density after 1 h later, a response that lasted up to 24 h [47]. These results reinforce the notion that agmatine and ketamine share common molecular targets and expand the findings regarding the ability of agmatine to enhance the antidepressant-like and synaptic actions of ketamine[47].

Taken together, these results support the hypothesis that agmatine can act as a ketamine-like compound, and further studies are crucial to investigate whether the rapid antidepressant effects of agmatine are reproducible in patients with MDD. Moreover, the use of agmatine in the clinic would be highly promising owing to its safety, even at high doses, without evident effects of toxicity[26,52].

### BEYOND mTORC1 HYPOTHESIS

In addition to focusing on the importance of the mTORC1-mediated signaling pathway for the antidepressant effect of agmatine, some studies investigated other signaling pathways that may play a role in its antidepressant effect. Understanding other pathways influenced by agmatine is important to its establishment as a therapeutic alternative in the clinical setting.

Regarding the factors that may influence the mechanisms associated with MDD symptoms, neuroinflammation has received much attention in recent years. Neuroinflammation reportedly plays an essential role in several neuropathologies, such as multiple sclerosis, Alzheimer’s disease, and MDD[66,67]. It was demonstrated in both humans and animals that immunological challenges may induce depressive behavior [68]. It is important to note that, since the 1990s, several studies reported a strong correlation between MDD and peripheral inflammatory markers[69,70].

In the last few years, three meta-analyses aimed to better understand the relationship between neuroinflammation and the development/maintenance of MDD. Kappelmann et al[71] published a meta-analysis in 2018 that analyzed data from four randomized controlled studies that examined the effects of pro-inflammatory cytokine inhibitors. In these studies, adalimumab and etanercept were used, and both treatments improved depressive symptoms in patients. In 2019, a randomized clinical study showed more pronounced antidepressant results in patients with higher high-sensitivity C-reactive protein levels[72]. In this study, the use of anti-inflammatory drugs improved clinical signs of depression, such as motor retardation, suicidal thoughts, and depressed mood[72]. Another meta-analysis evaluated 36 studies and assessed the effects of anti-inflammatory drugs in almost 10000 patients[73]. The findings suggested that the use of anti-inflammatory drugs sufficiently reversed depressive symptoms. The latest meta-analysis combined 26 randomized clinical trials of over 1500 patients[74]. This study also verified the improvement in depressive symptoms in patients with anti-inflammatory agent use[74].

Therefore, inflammatory pathways play an important role in the development and maintenance of depressive symptoms. Increasing evidence has shown that agmatine also acts on neuroinflammatory-related pathways that may participate in rapid-onset antidepressant responses.

The early administration of agmatine for 7 d prevented the depressive-like behavior caused by lipopolysaccharide (LPS) challenge in mice[75]. Agmatine pretreatment counteracted LPS-induced neuroinflammation by preventing increases in interleukin (IL)-1β and tumor necrosis factor (TNF)-α level in the murine brain. In addition, agmatine positively regulates BDNF levels in the hippocampus[75]. In another study, agmatine pretreatment also normalized LPS-induced sickness behavior in rats in addition to decreasing serum concentrations of IL-6 and TNF-α[76]. Zarifkar et al[77]
reported that agmatine prevented LPS-induced spatial memory impairment and hippocampal caspase-3 activation in LPS-treated rats. It is also noteworthy that agmatine effectively inhibited the LPS-induced production of nitrite and decreased body temperature in rats in a dose-dependent manner[78].

Notably, Neis et al[45] showed that agmatine effectively counteracted the depressive-like behavior induced by the pro-inflammatory cytokine TNF-α in mice. In this study, the combined treatment of sub-effective doses of agmatine and fluoxetine, imipramine, bupropion, MK-801, or 7-nitroindazole resulted in a synergistic antidepressant-like effect in mice subjected to TNF-α administration[45].

Agmatine also exhibits anti-inflammatory effects in other disease models. In particular, in a type 2 diabetes mellitus (T2DM) model induced by a high-fat diet for 12 wk, anxiety- and depressive-like behaviors were associated with an increase in pro-inflammatory cytokines, such as IL-6 and TNF-α, as well as a decrease in the BDNF immunoencontent in the rat hippocampus[79]. These parameters were inhibited by agmatine treatment in the last 4 wk of the protocol. In this study, agmatine levels in the hippocampus of rats subjected to the T2DM protocol were significantly lower than those in the control animals[79].

In an Alzheimer’s disease model, the administration of amyloid-β peptide (Aβ1-42) to mice caused depressive-like behavior in the forced swimming test, an effect parallel to an increase in the pro-inflammatory cytokines IL-6 and TNF-α in the hippocampus[80]. Both depressive-like behavior and pro-inflammatory markers were reversed by agmatine treatment, suggesting that the anti-inflammatory properties of agmatine may be related to its antidepressant effect. Notably, this study also detected lower concentrations of agmatine in the brains of animals injected with Aβ1-42[80]. These data point to the action of agmatine in neuroinflammatory processes, as a pharmacological strategy to decrease depressive-like behavior, including that associated with comorbid diseases, such as T2DM and Alzheimer’s disease.

It is important to note that the activation of various types of inflammasomes is a critical target in the inflammatory response. Inflammasomes are involved in the development of several neurological diseases, including MDD[81,82]. Among them, the NLR family pyrin domain containing 3 (NLRP3) inflammasome is the most closely related to MDD due to the exaggerated activation of inflammatory and immunological responses that contribute to the pathogenesis and progression of this disorder[83]. A compelling study reported that depressive-like behavior in mice subjected to LPS administration is related to NLRP3-dependent caspase-1 activation[84]. Accordingly, anxiety-like behavior reportedly occurs in rats exposed to neonatal inflammation or inflammatory stress early in life triggered by NLRP3 inflammasome activation in animals’ brains[85]. Altogether, evidence suggests that the NLRP3 inflammasome plays an essential role in the neurobiology of MDD and may be a potential target for antidepressant treatment.

In this regard, ketamine was shown to exert an antidepressant effect in the LPS-induced model via suppressing the NLRP3 inflammasome and upregulating AMPA receptors[86]. Importantly, in this study, the authors postulated that ketamine might increase AMPA receptor expression through the NLRP3 inflammasome, suggesting that NLRP3 could be a target in fast-acting antidepressant treatment[86].

The possibility that agmatine exerts antidepressant effects by modulating neuroinflammatory mechanisms has also been investigated. Sahin et al[43] investigated the effects of agmatine in a model of restraint stress-induced depressive-like behavior. The authors demonstrated that agmatine rescued anti-inflammatory cytokine IL-4 and IL-10 levels that were impaired by stress[43]. Moreover, the 6-wk treatment with agmatine counteracted the depressive-like behavior of animals exposed to chronic unpredictable stress by suppressing NLRP3 and IL-1β[43].

The exact role of the NLRP3 inflammasome–driven signaling pathway in MDD pathophysiology and antidepressant responses is still not well established. However, it has been proposed that the gut microbiota may influence activation of the NLRP3 inflammasome and neuroinflammatory processes through the microbiota-gut-brain axis.

The microbiome, a complex ecosystem in the human gut, includes bacteria, viruses, archaea, and fungi. The bacteria present in this system regulate aspects of the host’s health, mainly brain development and functioning[87,88]. The microbiome is a dynamic structure that is affected by delivery type, sex, age, nutrition, stress, and medications[89]. These interferences can compromise the balance between pathogenic and commensal bacteria[90], promoting the development of a process called dysbiosis, which can change the permeability of the intestinal wall, allowing bacteria and their products to leak into the sterile cavity and activate the immune response[91]. Immune response activation increases the levels of pro-inflammatory cytokines, which, together
with other toxic metabolites, damage the blood–brain barrier and trigger neuroinflammation[92].

The immune and brain mechanisms involved in intestinal dysbiosis may include microglial activation[93]. Microglia are responsible for releasing pro-inflammatory cytokines in the brain when activated by stress, a mechanism that is altered in MDD [94,95]. On the other hand, a balanced and healthy microbiota can regulate the activation of these stress response pathways through the synthesis of hormones and neurotransmitters, minimizing the effects of such stressors[96].

Studies have shown that there is a “microbiota of MDD” due to the difference in composition between depressed patients and healthy controls. In a microbiome study of patients with MDD and irritable bowel syndrome, less bacterial diversity, an effect associated with increased levels of bacteria from the phylum Bacteroidetes, and increased colon inflammation were noted in patients compared to healthy controls[97]. In a Chinese cohort, the microbiota of patients with MDD showed higher concentrations of Proteobacteria and decreased concentrations of Firmicutes[98].

Several studies have suggested the direct modulation of bacteria in the immune system. *Proteus mirabilis*, a proteobacterium, can activate the NLRP3 inflammasome and interleukin IL-1β production[99]. Other components of Proteobacteria, such as the LPS produced by *Pseudomonas*, are related to the development of MDD symptoms via activation of the NLRP3 inflammasome and pro-inflammatory immunoglobulins[93]. In patients with MDD, an increase in some Bacteroidetes species (*Parabacteroides* and *Alistipes*) reportedly converts tryptophan to indole, which can influence the availability of tryptophan in the body and affect serotonergic balance[100]. Other studies confirmed an increase in *Alistipes* bacteria in patients with MDD, chronic fatigue syndrome, irritable bowel syndrome, and stress models[101,102].

The transplantation of fecal microbiota from patients diagnosed with MDD to germ-free microbiota mice triggered anxious-, anhedonic-, and depressive-like behaviors in the animals[103,104]. This evidence suggests that the depressive phenotype may be transmitted by gut microbiota. These data show a close relationship between the composition of the gut microbiota and brain health, mainly in the pathological mechanisms involved in the development and maintenance of depressive symptoms. Furthermore, the immune system/NLRP3 inflammasome acts as an intermediary between gut dysbiosis and brain function.

Some studies have suggested that the ability of ketamine to elicit antidepressant effects may be mediated, at least in part, by modulation of the microbiota-gut-brain axis. Two studies that investigated the effects of ketamine administration in the gut microbiota of mice following the social defeat stress model reported that the treatment attenuated the alterations in *Bacteroidales*, *Clostridiales*, *Ruminococcaceae*, *Deltaproteobacteria*, and *Mollicutes* bacterial levels in their feces[105,106]. Moreover, ketamine prevented the increase in the *Clostridium* and *Butyricimonas* species induced by the stress model[105,106]. Other studies showed that ketamine significantly amplified the number of healthy bacteria and decreased the number of opportunistic pathogens in Wistar rats[107]. In an inflammatory model of LPS-induced depressive-like behavior, ketamine improved the diversity of the gut microbiota, positively regulating this microsystem[108]. Together, these data suggest that ketamine influences the composition of the microbiota, a response that may underlie its antidepressant-like effects.

The relationship between gut microbiota and agmatine levels has emerged and may play a role in the ability of gut microbiota to influence mental health. Agmatine is produced and released by gut bacteria of the human microbiome[109] and can be obtained from ingested food[110,111]. The composition of the intestinal microbiota influences agmatine availability in the gut lumen for absorption, and the majority of agmatine in humans is believed to be derived from bacterial sources[27]. Interestingly, agmatine may also be obtained from foodstuffs, particularly fermented foods such as alcoholic beverages (wine, beer, sake), which suggests the role of yeast in its production[109]. The filamentous fungus *Aspergillus oryzae*, which is widely used for the production of various Asian fermented foods, can enhance agmatine ingestion[112].

The consumption of fermented foods has beneficial effects on mental health[113]. The use of probiotics also reportedly exerts positive effects on depressive symptoms[114,115]. The possibility that agmatine is produced in the gut following the consumption of fermented foods and probiotics may account, at least in part, for its anti-inflammatory and antidepressant effects should be investigated in future studies.

Metformin, the mainstay therapy for T2DM, reportedly influences the diversity and composition of the gut microbiota[116]. This drug has recently been shown to act on *Escherichia coli*, elevating agmatine production and increases the longevity of
Valverde AP et al. Agmatine as a rapid-onset antidepressant

*Caenorhabditis elegans*[117]. Metformin has been shown to produce antidepressant effects in depressed patients with diabetes mellitus[118] and proposed as an adjunctive antidepressant approach in nondiabetic patients with MDD[119]. It remains to be determined whether agmatine levels are higher in individuals taking metformin and, if so, whether it contributes to the antidepressant effect observed with metformin treatment.

**CONCLUSION**

Agmatine, an endogenous cationic amine, exerted antidepressant effects in several preclinical studies[26,120]. Considering that the microbiota composition and consumption of fermented foods, or even some drugs such as metformin, may influence agmatine levels in the gut[27,109-119], it remains to be established whether agmatine derived from these sources may positively impact mood and exert antidepressant effects. Therefore, modulation of the microbiota and, consequently, gut agmatine levels may represent a novel approach to mood regulation.

In addition to the fact that agmatine may be synthesized by gut microbiota, several studies have indicated that it is safe even when administered at high doses as a nutraceutical. The sulfate salt of agmatine has been used for bodybuilding[27] and the management of neuropathic pain at doses as high as 2.6 g/day[121]. The fact that agmatine also exhibits several beneficial effects for a wide spectrum of diseases[27] suggests that it is a promising therapeutic strategy for the management of MDD and several comorbid diseases and inflammatory clinical conditions such as diabetes, obesity, pain, and neurodegenerative diseases. Of particular relevance, compelling preclinical evidence has indicated that agmatine has the ability to counteract several neuroinflammatory markers induced by models of depression and shares with ketamine the ability to elicit fast antidepressant responses[46,47,61,62,75-80]. The possibility that agmatine may afford a rapid antidepressant effect would give it an advantage over conventional antidepressants that require several weeks to alleviate depressive symptoms. In preclinical studies, agmatine elicited a synergistic effect with ketamine in mice subjected to animal models of depression as well as cell culture, and pharmacological evidence has pointed to similar molecular mechanisms of these drugs[46]. These properties of agmatine clearly warrant future clinical investigation of its beneficial effects for managing depressive symptoms as a monotherapy or adjunctive treatment. Therefore, clinical studies are warranted that investigate the possibility that agmatine may be combined with low doses of ketamine to diminish the side effects and provide synergistic antidepressant effects.

**REFERENCES**

Valverde AP et al. Agmatine as a rapid-onset antidepressant

Psychiatry 2010; 71: 1605-1611 [PMID: 20673547 DOI: 10.4088/JCP.09m05327bh]


Duman RS, Li N, Liu RJ, Duric V, Aghajanian G. Signaling pathways underlying the rapid antidepressant actions of ketamine. Neuropsychopharmacology 2012; 37: 35-41 [PMID: 21907221 DOI: 10.1016/j.neuropharm.2011.08.044]


Li N, Lee B, Liu RJ, Banasr M, Dwyer JM, Iwata M, Li XY, Aghajanian G, Duman RS. mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. Science 2010; 329: 959-964 [PMID: 20724638 DOI: 10.1126/science.1190287]


Aspartate (NMDA) Receptor Blockade Prevents Neuronal Death Induced by Zika Virus Infection. mBio 2017; 8 [PMID: 28442607 DOI: 10.1128/mBio.00350-17]


Zomkowski AD, Santos AR, Rodrigues ALS. Evidence for the involvement of the opioid system in


70 *Levine J*, Barak Y, Chengappa KN, Rapoport A, Rebay M, Barak V. Cerebrospinal cytokine levels
Valverde AP et al. Agmatine as a rapid-onset antidepressant


93 Maes M, Kubera M, Leunis JC. The gut-brain barrier in major depression: intestinal mucosal dysfunction with an increased translocation of LPS from gram negative enterobacteria (leaky gut) plays a role in the inflammatory pathophysiology of depression. *Neuro Endocrinol Lett* 2008; 29: 117-124 [PMID: 18283240]


95 Réus GZ, Fries GR, Stertz L, Badawy M, Passos IC, Barichello T, Kapczinski F, Quevedo J. The role of inflammation and microglial activation in the pathophysiology of psychiatric disorders. *Neuroscience* 2015; 300: 141-154 [PMID: 25981208 DOI: 10.1016/j.neuroscience.2015.05.018]


106 Yang C, Qu Y, Fujita Y, Ren Q, Ma M, Dong C, Hashimoto K. Possible role of the gut microbiota-brain axis in the antidepressant effects of (R)-ketamine in a social defeat stress model. *Transl Psychiatry* 2017; 7: 1294 [PMID: 29249803 DOI: 10.1038/s41398-017-0031-4]


112 Akasaka N, Fujiwara S. The therapeutic and nutraceutical potential of agmatine, and its enhanced...
Valverde AP et al. Agmatine as a rapid-onset antidepressant


117 MacNeil LT, Schertzer JD, Steinberg GR. Bacteria transmit metformin-associated lifespan extension. *Nat Rev Endocrinol* 2020; **16**: 9-10 [PMID: 31645681 DOI: 10.1038/s41574-019-0278-3]


Healthy diet, depression and quality of life: A narrative review of biological mechanisms and primary prevention opportunities

Octavio Pano, Elena H Martínez-Lapiscina, Carmen Sayón-Orea, Miguel Angel Martinez-Gonzalez, Jose Alfredo Martinez, Almudena Sanchez-Villegas

ORCID number: Octavio Pano 0000-0003-4848-9944; Elena H Martinez-Lapiscina 0000-0003-4272-0826; Carmen Sayón-Orea 0000-0002-4137-3263; Miguel Angel Martinez-Gonzalez 0000-0002-3917-9808; Jose Alfredo Martinez 0000-0003-4882-4044; Almudena Sanchez-Villegas 0000-0001-7733-9238.

Author contributions: Pano O provided supportive literature and wrote the manuscript; Sánchez-Villegas A provided supportive literature and gave expert advice and guidance on the topic; Martínez-Lapiscina EH, Sayón-Orea C, Martínez JA and Martínez-González MA gave their expert opinion and supervised the review; All authors have read and approved the final manuscript.

Conflict-of-interest statement: The authors have no conflicts of interest to declare.

Country/Territory of origin: Spain

Specialty type: Psychiatry

Provenance and peer review: Invited article;Externally peer reviewed

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B

Abstract
Unipolar depressive disorder (UDD) affects more than 264 million people worldwide and was projected well before the severe acute respiratory syndrome coronavirus 2 pandemic to be the leading cause of disability-adjusted life years
INTRODUCTION

More than 264 million people suffer from unipolar depressive disorder (UDD) worldwide, making it a leading cause of years of life lost as a result of disability[1]. In 2019, unipolar depressive disorder was ranked in 6th place amongst the top causes of burden of disease for the age groups between 25-49 years[2]. Previous projections had already speculated that depressive disorders would be the leading cause of disability-adjusted life years lost by the year 2030[3]. These estimations will inevitably rise as a
result of the COVID-19 pandemic, requiring the implementation of public health policies to contain this expansive global health issue[4].

Current efforts to prevent mental illness, including depression focus on early diagnosis with prompt medical attention and treatment, in addition to sufficient drug availability[5]; in other words, secondary prevention. Despite moderate success in some countries, current research is offering both a better understanding of this condition and the rationale to apply new preventive strategies based on this knowledge. Examples of such research have been carried out by Sanchez-Villegas et al [6,7] in a population of southern Europeans. In this study sample, poor dietary and sedentary behaviors were associated with a higher incidence of mental illnesses (including UDD), which are also determinants typically associated with cardiometabolic health[8]. Due to common biological pathways, nutritional epidemiology now offers further insights in the field of mental illnesses[9].

Regardless of promising reports which support the possibility of improving mental health through nutrition, evidence remains limited[10]. Some of the findings refer to the importance of maternal and early-life nutrition for prevention, but also the role of diet as a determinant of adequate symptom management and enhanced anti-depressant drug efficacy. Such findings have led to an exponential growth and interest in the prevention of UDD through primary prevention strategies[11-13]. Nonetheless, the implementation of such strategies requires solid, reproducible evidence which can only be obtained from nutritional intervention trials.

In this narrative review we summarize the available evidence on the biological mechanisms common to UDD and cardiometabolic diseases, in support of the implementation of dietary interventions as plausible preventive measures for these diseases. For this purpose, we provide an overview of the aetiological hypotheses of depression with a particular interest in the role of diet on these pathways. Finally, current epidemiological evidence on the best known healthful dietary pattern, the Mediterranean diet (MD), will be presented in support of dietary interventions to reduce depression incidence and aid overall prognosis.

**AETIOLOGICAL HYPOTHESES OF DEPRESSION**

A wide array of determinants are involved in the aetiology of depression, including biological, psychological, and environmental factors[5]. It is not unlikely that these mechanisms coexist in varying degrees with plausible synergistic activity; adding complexity and variability to the causes and clinical manifestation of UDD. Among the most accepted pathways is the dysregulation of hormonal signaling involving the Hypothalamus-Pituitary-Adrenal (HPA) axis, although monoamine activity, neurogenesis, neuronal activity and plasticity play fundamental roles in the aetiology of mental illnesses in general. Of note, oxidative stress and inflammation play a common role in most of these hypotheses[14].

**Oxidative stress and pro-inflammatory state**

Most of the associated biological mechanisms in depressive disorders are accompanied by the production and release of cytokines, such as interleukin (IL)-1β, IL-2, IL-6, interferon-γ (INT-γ) and tumor necrosis factor (TNF)-α. These disturbances have been confirmed in patients with UDD, among whom plasma levels of certain pro-inflammatory cytokines (TNF-α, IL-6, IL-8, and C-reactive protein among others) are significantly higher than those of healthy individuals[15-17]. Furthermore, effective pharmacological treatment decrease cytokine concentrations, particularly IL-6, C-reactive protein and INT-γ[17,18]. The basis for this aetiological hypothesis is the damage to the blood brain barrier (BBB) caused by the convergence of oxidative stress and pro-inflammatory signals, which impede an adequate brain function[19]. These alterations are particularly noticeable in the BBB tight junction protein, claudin 5 (Cldn5), increasing cytokine permeability to the brain of mice[19,20]. Additionally, altered BBB function promotes the infiltration of various immune cells into the brain, endothelial damage through hidac expression, and direct damage to astrocytes, glial cells, and neurons[5,21,22]. Further release of peripheral cytokines originate in tissues such as the adipose tissue and gastrointestinal tract, resulting in vagal nerve stimulation, which also reinforces this pro-inflammatory state[19,23]. Each of these signals converge in the central nervous system (CNS) and result in altered receptor expression, neurogenesis and plasticity[5,19].

Given these connections, inflammatory cytokines have been proposed as biomarkers for depression in association with cardiovascular disease. These markers
span various biological levels, including single-nucleotide polymorphisms identified in genome wide analyses, but also molecules such as the neutrophil gelatinase associated lipocalin[24]. Of particular interest are the associations found between the loci coding for the pro-inflammatory cytokine IL-6 and antidepressant responsiveness [25]. Oxidative stress and inflammation are also directly correlated to indoleamine 2,3-dioxygenase (a rate limiting enzyme in the metabolism of tryptophan), which leads to neurotoxicity and serotonin deficiency in the brain; both core features of the pathophysiology of depression[26].

These results are of great interest as UDD can be thought of as an inflammation-based disease, but also indicate potential anti-oxidant properties exerted by anti-depressant drug therapies. Overall, these findings offer new preventive strategies to the field of mood disorders. Regarding the robustness of these findings, a 2015 cumulative meta-analysis analyzing 58 studies evidenced that the role of the most commonly associated inflammatory cytokines on UDD was substantial[16]. Additionally, these results explain the diversity of clinical manifestations and responsiveness to treatment as a result of the interplay between biological pathways and individual factors. In this regard, we cannot disregard the potential effect of confounding as a plausible explanation for these differences, given the lack of clinical trials controlling for the numerous determinants of UDD.

**Hypothalamus-pituitary-adrenal axis**

Well-defined clinical and biochemical evidence support the role of glucocorticoid dysregulation as a cause of depression. The release of cortisol is the end product of the HPA-axis in response to external stress[5]. Moreover, circulating cortisol levels activate the sympathetic nervous system, and other immune responses that influence neuronal activity. Despite this being a natural response, disparities have been found in both animal and human models revealing individualized responses to external stressors, and thus differences in UDD risk[27]. With regard to this hypothesis of UDD, sustained induced stress will lead to the down regulation of glucocorticoid receptors in the CNS. Normally, cortisol inhibits the further release of corticotropin-releasing hormone and adrenocorticotropic hormone in a self-regulating cycle. However, this negative feedback is altered in patients with depression, leading to sustained elevated levels of cortisol, alterations of the cortisol circadian rhythm, as well as altered immunological responses[28]. Some of the long-term effects of these disturbances include not only a reduction in the total hippocampal volume and reduced rates of neurogenesis, but also an accumulation of visceral fat and obesity, a higher susceptibility to insulin resistance, and an increase in the production of inflammatory cytokines, which lead to cardiometabolic diseases[29].

To our knowledge a direct influence of dietary habits on the HPA-axis has not been reported, however some authors suggest that the gut-microbiome could mediate these effects[30]. Clinical and biochemical evidence has concluded that alterations in the diversity and particular strains of bacteria comprising the microbiome, also known as dysbiosis, modulate the negative effects of stress on the HPA axis at different stages of life[31]. A particular study on marital-related stress reported an increased release of lipopolysaccharides (LPS) originating in the gut microbiome, as a result of dysbiosis, with a dose-response relationship[32]. LPSs cause damage to the gut membrane and lead to major adverse cardiovascular events as demonstrated by Pastori et al[33] in 2017. Moreover, damage to the gut membrane results in the translocation of LPSs into the plasma, where they stimulate the differentiation of T-cell lymphocytes, further promoting inflammatory mechanisms[33,34]. Although the objective of the study was to assess the associations between circulating LPS and major adverse cardiovascular events, similar biological mechanisms have been identified in patients with depression[35]. Given the lack of clinical trials in this particular matter, incident depression has been inconsistently associated with cytokines such as LPS and others[35]. Hence, the true nature of this relationship has yet to be explained.

Dietary determinants of dysbiosis include fiber, fatty acid, and probiotic intakes[30,36,37]. In fact, dietary interventions have demonstrated that changes in the microbiome are possibly due to these food items, however incident depression has not been assessed concomitantly[38,39]. Results in a Korean study evaluating dietary fiber and prevalent depression demonstrated significant inverse associations between cereal, seaweed and mushroom fiber, and depressive symptoms[40]. Despite these encouraging findings, significant associations with prevalent depression were only found for the consumption of seaweed fiber (multivariable-adjusted OR 0.45 (95%CI: 0.23-0.88) for the comparison of quartile 4 vs quartile 1 of intake)[40]. Additionally, microbiome changes are highly dependent of the pre-existing diversity of bacteria in the gut. As stated by others, the microbiome demonstrates a resilience to change which
is not yet fully understood[41]. Hence, additional research is needed before general recommendations can be made.

**Monoamines**

Another widely accepted aetiological theory of depression focuses on the role of monoamine synthesis and activity. Indeed, lower concentrations of these monoamines (serotonin, dopamine and noradrenaline) have long been associated with depressive symptoms and ultimately with depressive disorder. Altered monoamine activity is driven by inadequate production and their rapid degradation in the pre-synaptic space [5]. However, some inconsistencies remain unanswered regarding this hypothesis. Primarily, the prolonged response time observed for typical antidepressants which reduce depressive symptoms weeks after initiating treatment despite a normalization of monoamine levels in just a few hours[42,43].

Monoamine production is primarily dependent of vitamin B-6, folate, and vitamin B-12 Levels as precursors[44]. These micronutrients are essential in the metabolism of S-adenosyl methionine and methionine, both of which are critical for monoamine production and other methylation processes within the brain[44]. Sub-optimal levels of folate and B-group vitamins have been associated with higher homocysteine concentrations, which in turn are associated with depressive disorders, Alzheimer’s disease, and altered cognitive disfunction[45]. The role of homocysteine in the hypothesized biological mechanisms of UDD is based on the direct vascular damage and excitotoxicity exerted by its metabolites (homocysteinic acid or cysteine sulfenic acid).

Specific examples of the effects of individual foods on monoamines are scarce, however increased serum levels of dopamine and serotonin have been documented in response to dietary restrictions or supplementation with cocoa extracts[46,47]. Monoamine activity is also dependent on its enzymatic degradation by the monoamine oxidase enzyme (MAO) in the pre-synaptic space; a staple target for antidepressant drug therapy. In particular, the MAO-B variant of this enzyme is susceptible to dietary interventions as demonstrated by Zellner et al[45]. This study demonstrated that an increase in vitamin B12 was correlated with a lower MAO-B expression (median expression reduction of 26%, adjusted P value < 0.05). In this three-week dietary intervention, a significant reduction of MAO-B was achieved through a high protein diet (approximately 3.0 g of protein/kg per day) primarily consisting of meat products (55% of total protein intake). Based on these observations, B-group vitamin supplements are theorized to have an antidepressant activity as noted in some *in vitro* and *in vivo* studies[48,49]. B-group vitamins are crucial for the transformation of homocysteine into methionine through the 5-methyltetrahydrofolate pathway [44], however vitamin B12 supplementation has not been uniformly associated with reduced UDD risk nor symptom improvements. Some clinical trials have noted elevated concentrations of homocysteine and low levels of B-group vitamins in depressed patients at baseline[50-52]. However, two interventions with B-vitamin supplements (the shortest consisting of two years, in a sample of *n* = 2919 geriatric patients with major depressive disorder) observed non-significant improvements in depressive symptoms[51,53]. Positive reports have been found under particular conditions, yet strong and reproducible evidence is warranted for the widespread use of vitamin B12 supplements as an inexpensive aid to pharmacological treatments[47,54]. Furthermore, a comparative analysis on the effect of folates and B-group vitamins should be conducted controlling for homocysteine reduction and increased monoamine activity through MAO-B suppression. Lastly, when considering the effects of dietary interventions on depressive symptoms, the overall antioxidant properties of a healthy dietary intervention should be considered.

**Neurotrophism and neuronal activity**

Although few reports exist on the effects of neurogenesis in the development of UDD, initial studies highlight the central role of the brain derived neurotrophic factor (BDNF). BDNF is synthetized and secreted primarily by endothelial cells[55] and is necessary for neuronal development as demonstrated in animal models[56]. The initial transcription and translation of the BDNF gene produces an immature molecule (prepro-BDNF) which is then converted into pro-BDNF and eventually transformed into mature BDNF through a series of protease enzymes[57,58]. Although prepro-BDNF has not been attributed biological activity, pro-BDNF can bind to the p75 neurotrophin receptor (p75NTR) and induce cell senescence in the hippocampus and prefrontal cortex[57-59]. Mature BDNF, on the other hand, has been related to the maturation of neurons, establishment of synaptic connections (through axonal growth and neuronal survival), and interestingly neural plasticity[56,57,60]. Such mechanisms, however, are highly dependent on the site of the brain in which they occur[59]. BDNF
secreted into the amygdala and nucleus accumbens has been found to worsen depressive symptoms, whereas the opposite is true in the prefrontal cortex and hippocampus, both in human and animal models[57,58]. The site-specific activity of BDNF was initially thought to be the cause of reduced hippocampal volume, as described in post mortem studies in patients with depression[59,60]. These findings have since been attributed to a single nucleotide polymorphism resulting in the translation of Methionine and not Valine-66 (Val°Met), which hinders the transformation of proBDNF to BDNF[59]. However, the lower hippocampal volume hypothesis has been thoroughly criticized in a recent meta-analysis which highlights the low statistical power of the available evidence, plausible publication bias, and tendency towards the null value of current evidence[61]. Post-mortem studies on patients with depression have also found lower BDNF levels, lower expression of the BDNF receptor (TrkB), and an increased expression of p75NTR in the hippocampus (particularly in suicide patients), all of which correlate to negative effects on depressive symptoms[60,62]. Finally, BDNF has also been suggested to be a mediator of certain antidepressant therapies, such as electroconvulsive therapy and anti-depressant drugs[56,58]. In the particular case of fluoxetine, mice with the knock-in Val°Met variant were completely void from the antidepressant effects of the drug[58].

A major debate in regard to antidepressants has been their delayed effects, leading some researcher to believe that BDNF activity is responsible[58]. This theory is based on the effects of antidepressants on BDNF levels, which stimulate neuronal plasticity and improve neuronal synapse in the hippocampus after a period of time; corresponding with the cessation of depressive symptoms[56].

In relation to diet, post mortem studies report lower levels of retinoic acid receptor-α (RARα) in the prefrontal cortex along with disturbances in the TrkB pathway[62]. Qi et al[62] found that retinoic acid and vitamin A sub-species could stimulate the TrkB promoter in an in vitro study and potentially improve depression management. Similar to BDNF, vitamin B12 has shown to selectively promote the expression of the BDNF-receptor gene (Ntrk-2) in the prefrontal cortex[48]. Despite observing no significant increase of these receptors in the hippocampus, an acute vitamin B12 supplementation counteracted the stress induced alterations of Ntrk-2 expression and depressive behavior in mice[48].

Given the novelty of these findings, few reports exist on the human effects of BDNF, and fewer which attribute its activity to dietary interventions. To our knowledge, only the PREDIMED (PREvenció n con Dieta MEDiterranea) trial has reported on the changes of BDNF as a result of a dietary intervention. In the PREDIMED trial, subjects allocated to the MD pattern and supplementation with nuts were less likely to have lower BDNF levels[63]. Moreover, after three years of intervention, a statistically significant increase in BDNF was observed among participants with prevalent depression and allocated to the MD supplemented with nuts. The results suggest that in patients with prevalent depression, BDNF concentrations are influenced by dietary interventions[63]. Despite the success observed in the dietary intervention of the PREDIMED trial, the complexities, regulatory mechanisms, and site-specific neuronal activity of this neurotrophin demand an in depth understanding of its implication in UDD.

**Health related quality of life**

One of the main reasons for UDD treatment is to improve patients' health related quality of life (HRQoL), however, this goal has proven difficult to achieve and understand. The perception of health has been broadly studied under two main physical and mental domains since the 70's[64]. In addition to overall health, external factors contribute greatly to this subjective concept. Bidirectional relationships between HRQoL and UDD have been reported, however, cause-effect relationships remain inconsistent. Some reports include reduced HRQoL at baseline for patients with UDD, in whom perceived health improves after anti-depressant treatment[65,66]. It should be mentioned that not all pharmaco-therapies improve well-being, possibly due to secondary effects or treatment related stress[66]. Hence the need for additional research. Moreover, if HRQoL does not improve, persistent perceived health reduction could worsen UDD and even increase suicide risk[5,67]. Various physical, mental and socioeconomic factors have been identified as determinants of increased suicide risk, all of which contribute varying effects[68]. In these matters, psychotherapy, particularly behavioral cognitive therapy, has proven to be an effective measure to improve HRQoL[69,70]. Unfortunately, these improvements have been limited to mental domains, which could indicate that additional factors are not being considered. On the other hand, diet has been cross-sectionally associated with both physical and mental domains of HRQoL[71-73]. Furthermore, nutrient deficiencies have been related to
worsened perceived health, whereas healthy dietary habits seem to improve HRQoL across various age-groups[74,75]. Research in this particular field is scarce, but initial results are beginning to shed light on these associations. A report by Gariballa et al[76] in 2013 found that in addition to changes in homocysteine levels, a significant improvement in HRQoL was observed. The added value of improving HRQoL is not only an effective measure of patient health, but also of the quality of medical care. Clinical trials and prospective studies should focus on the potential mediator effect of nutritional status on the association between depression-related clinical outcomes and perceived health (UDD - Diet - HRQoL).

In the following table (Table 1) a summary of the primary aetiological hypothesis and mechanisms involved in depression are summarized. Furthermore, one-to-one associations and evidence of the potential effects of diet on each of these mechanisms is presented.

## LINKS TO OTHER CARDIO-METABOLIC DISEASES

Along with UDD, cardiometabolic diseases are amongst the top causes of global burden of disease (GBD)[2]. Based on the mechanisms previously described, more and more evidence supports the role of diet and lifestyle as key determinants of mental health, quality of life, and cardiometabolic disorders[3]. Around the world, unbalanced diets, consumption of sugary beverages, and fast foods have been associated with and increased risk and presence of metabolic disturbances such as elevated fasting plasma glucose, and high blood pressure[77]. Sedentary behavior, a major cause of obesity and metabolic syndrome, has been associated with an increased risk for various mental disorders[7]. These observations further reflect the implications of (poor) lifestyle patterns habits on physical and mental health, both of which flourish under similar biological conditions. Likewise, the chronic pro-inflammatory state present in metabolic syndrome is the ideal scenario for glucose intolerance, lower insulin secretion, and the accompanying glucocorticoid dysregulation that precedes the onset of diabetes[78-81]. The cycle of damage persists once these changes have been established due to the detrimental effects of pro-inflammatory cytokines across various organs and systems, including pancreatic β-cells, endothelial cells, various strains of immune cells, among others. Unfortunately, advancements have been limited, however, biomarkers such as the myokine irisin are beginning to emerge as bridging biomarkers both in UDD and cardiometabolic diseases. This myokine, also produced in the adipose tissue, has been found to be a potent predictor of post-stroke depression in addition to solid associations with physical exercise and adiposity[82]. Considering these interactions, it is not outlandish to argue that the actual impact of depression has been underestimated, as prior researchers have suggested[83,84].

## WHY A MD PATTERN FOR THE PREVENTION OF DEPRESSION?

The literature surrounding the MD and its characteristics is vast and well documented. When referring to the MD, researchers address a pattern of dietary characteristics and lifestyle found in countries of the Mediterranean basin, specifically in the 1960s[85]. In this geographical and temporal period, diets were characterized for their high consumption of olive oil (as primary source of dietary fats), nuts, fatty fish, local seasonal fruits and vegetables, in addition to a low consumption of meat (with a preference for poultry rather than red meats), dairy products, sugared beverages and confectionary/baked goods, and a moderate consumption of red wine. Furthermore, traditional culinary practices such as the “sofrito” (a tomato and olive oil-based sauce with various vegetables) enhance not only the flavors in this dietary pattern but also their innate nutrients. For the purpose of nutritional epidemiological studies, dietary questionnaires have been developed to assess the consumption of key foods and culinary techniques of this dietary pattern. Pioneering to this field was the development of the MD score (MDS) by A. Trichopoulou and collaborators, which considers the consumption of 9 key foods characteristics of the MD[86,87]. More recently, the 14-item questionnaire developed for the PREDIMED trial is the most wide-spread questionnaire evaluating the adherence to the MD pattern and its health benefits[88]. A novel, more precise approach is presently being developed with the inclusion of metabonomic assessment in order to objectivelyascertain adherence to the MD. These analyses are based on the identification of key end-products of human
Table 1 Summary of aetiological hypotheses and nutritional determinants of depression

<table>
<thead>
<tr>
<th>Aetiological Hypotheses</th>
<th>Main mechanisms</th>
<th>Biological disturbances in UDD</th>
<th>Relationship with Diet</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidative stress and inflammation</td>
<td>Higher cytokine levels in patients with UDD induce BBB damage and increased permeability to the brain, mainly through claudin 5 dysfunction. Furthermore, oxidative stress and cytokines stimulate the vagal nerve, further promoting a pro-inflammatory in the CNS.</td>
<td>Increased IL-6, TNF-α, C-reactive protein, INT-γ, and other inflammatory cytokines. BBB damage and translocation of cytokines and immune cells into the brain.</td>
<td>Adequate antioxidant intake and lifestyle habits prevent the development of a pro-inflammatory state.</td>
<td>[16,17,19]</td>
</tr>
<tr>
<td>Glucocorticoids and HPA-axis</td>
<td>Chronic stress leads to the constant release of cortisol and eventually a down regulation of cortisol receptors in the hypothalamus. In addition, altered immune responses promote further neuronal damage and other metabolic diseases.</td>
<td>Increased basal cortisol and LPS.</td>
<td>Healthy gut microbiomes prevent the release of LPS and other cytokines into the plasma, limiting the potential damage of stress and theoretically decreasing UDD risk.</td>
<td></td>
</tr>
<tr>
<td>Monoamine activity</td>
<td>Inadequate monoamine production and an increased degradation rate, mediated by the MAO enzyme, are related to depressive symptoms. Although current therapy is based on improving monoamine availability in the synapsis, uncertainties remain regarding this hypothesis.</td>
<td>Reduced monoamine formation and activity.</td>
<td>Vit-B12 and folate are necessary for monoamine production.</td>
<td>[28,29,33,34]</td>
</tr>
<tr>
<td>Neuronal development and activity</td>
<td>Altered BDNF function is caused by oxidative stress and inadequate endothelial function. This neurotrophin is essential for neuronal development, synapse formation and cerebral plasticity, in addition to having anti-depressant effects. However, its antidepressant effects are dependent on brain region and therefore not fully understood</td>
<td>Altered BDNF activity and other markers of endothelial function (IL-6, TNF-α, ICAM-1 and VCAM-1).</td>
<td>Antioxidants, MUFA, and PUFA concentrations regulate endothelial function.</td>
<td>[44–46]</td>
</tr>
<tr>
<td>The role of Health Related Quality of Life</td>
<td>Lower HRQoL has been associated with depression relapse and increased risk of suicide.</td>
<td>Decreased post-mortem hippocampal and prefrontal cortex volumes.</td>
<td>Vit-B12 promotes the expression of genes that code for BDNF-receptors (Ntrk-2).</td>
<td>[55,57,62,63]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HRQoL is generally lower in patients with depression. Furthermore, pharmacotherapy improves mental components of HRQoL.</td>
<td>Quality of diet has been direct and cross-sectionally associated with HRQoL scores. The effects of diet on homocysteine level influence HRQoL and thus, potentially improve UDD patient status.</td>
<td></td>
</tr>
</tbody>
</table>

UDD: Unipolar depressive disorder; IL-6: Interleukin 6; TNF-α: Tumoral necrosis factor-α; INT-γ: Interferon-γ; MAO: Monoamine oxidase; BDNF: Brain derived neurotrophic factor; ICAM: Intercellular adhesion molecule; VCAM: Vascular adhesion molecule; MUFA: Monounsaturated fatty acids; PUFA: Polyunsaturated fatty acids; HRQoL: Health related quality of life.

metabolism that prevail when an individual adheres to this dietary pattern; also referred to as the “metabolic signature” of the Mediterranean diet[89].

The cardio-protective effects of the MD are mainly attributed to its anti-inflammatory and antioxidant properties. For this and other reasons, the American College of Cardiology and the American Heart Association amply recommend this dietary pattern in their 2019 guidelines[90], in addition to its inclusion in the Dietary Guidelines for Americans 2020-2025. All three institutions advocate for dietary patterns which emphasize the consumption of seafood, vegetables, fruits and nuts for the prevention of cardiovascular disease. However, in relation to UDD, the 2015 American dietary guidelines advisory committee deemed the evidence on the risk reduction of depression to be limited in its actual state[91]. When compared with other dietary patterns, the MD provides adequate concentrations of most micronutrients, particularly antioxidant compounds[92-94].

Based on these characteristics and reflecting on the biological hypotheses previously described, we could argue that the MD could also be beneficial for UDD prevention.
Epidemiological studies have found that these compounds exert positive effects on endothelial function and inflammation based on biomarkers such as TGF-β1 and IL-6, ICAM-1 VCAM-1, among others. As evidenced in the study by Sánchez-Villegas et al. in 2011, a dietary intervention with a Mediterranean diet supplemented with nuts in patients improved BDNF concentrations in patients with prevalent depression. Additional benefits to the endothelium of cardiovascular and neural cells originate from the adequate ratio of “healthy fats” in this dietary pattern [reflected in the mono- and polyunsaturated fat (MUFA/PUFA) and mono- and saturated fat (MUFA/SFA) ratios] and the exclusion of behavioral therapy yields no significant improvement of symptoms. The Mediterranean diet (MD) is more importantly, the exclusion of behavioral therapy yields no significant improvement of symptoms. The Mediterranean diet (MD) is more beneficially associated with the prevention of UDD [95]. Among the most researched effects of fatty acids on membrane function are the effects of Omega-3 fatty acids (also n-3). N-3 fatty acids include are polyunsaturated fatty acids (PUFA) that derive from alpha linolenic acid. The two primary forms of n-3 are eicosapentaenoic (EPA) acid and docosahexaenoic (DHA) acid, both of which are found in seafoods. EPA in particular plays a critical role in the adequate functioning of dendritic and synaptic membranes, hence its advised supplementation for patients with UDD. This line, DHA is an integral component of macular and dendritic cell membranes, further supporting its supplementation. Consumption of fish and nuts in the MD has demonstrated to improve overall lipid composition (n-3 and other unsaturated fatty acids instead of trans-fatty acids or saturated fatty acids) which ensures adequate membrane functioning.

Folate and vitamin B12 intake are a major concern for subjects following plant-based diets. Nonetheless, nutrient deficiencies were scarce in a sample of free-roaming individuals adhering to the MD [95]. These micronutrients are critical for BDNF function but also necessary for an adequate production and activity of monoamines. Green leafy vegetables, legumes, grains, nuts and fruits within the MD have a high nutritional value for their folate and vitamin B12 content as well as other nutrients, such as fiber, vitamins, and minerals. Most of these foods are responsible for the positive effects of the MD on gut microbiome, further evidenced by improvements in insulin sensitivity and diabetes risk [100,101]. In this study, it was estimated that the prevalence of vitamin B12 deficiency among the 4th and 5th quintiles of adherence to the MD was 1%. Hence, the prevention of UDD by a MD pattern has been partially attributed to this dietary pattern’s adequate intake of vitamin B12 [102]. This was further supported by a comparison between categories of MD adherence and UDD incidence risk; compared to the lowest adherence to the MDS, the top 4 categories were associated with a risk reduction for incident depression of 26%, 34%, 51% and 42%, respectively [8], observing a significant dose-response relationship [8,103]. Although subject to debate, caloric restriction has been demonstrated to play a role in UDD and aids in the progression of non-alcoholic fatty liver disease [104]. Hunger-related stress induced anti-oxidant effects in the brain through increased blood flow and the removal of damaged cellular organelles. Interestingly, the hippocampus is particularly benefited from the increased perfusion, resulting in an increase in neurogenesis mediated by BDNF and a decrease in apoptosis rates [105,106]. However, this theory would imply that the release of glucocorticoids and the subsequent HPA-axis activation, as a result of stress, is also capable of promoting neurogenesis despite its impairment in patients with depression. In a meta-analysis, the mean weighted effects of a very low calorie diet on depressive symptom scores revealed a Hedges’ g = −0.728 (95%CI: −1.204, −0.253) [107]. These results were further analyzed in order to understand the mediator effects of exercise, duration of the caloric restriction, and cognitive behavioral therapy. It was concluded that exercise and longer duration of caloric restrictions partially explain the improvements in depressive symptoms, but more importantly, the exclusion of behavioral therapy yields no significant improvement of symptoms [107]. These mechanisms and the UDD hypotheses to which they are linked, have been summarized in Figure 1.

These promising reports signal an under-developed field for UDD prevention; however, the long-term psychiatric effects that could arise from the severe acute respiratory syndrome coronavirus 2 pandemic demand for an exponential development of this knowledge. Front line workers should be considered at a double risk for COVID-19 and psychiatric disorders as suggested by the high prevalence of anxiety and burn-out in this group, with serious implications for the health and well-being of the population in general [108,109].

Some of the most noteworthy results from observational studies and clinical trials regarding the role of the MD, individual nutrients, and foods on incident depression are summarized in Table 2. This summary includes studies conducted in the Seguimiento Universidad de Navarra (The SUN study), studying the associations between diet and lifestyle with incident depression.
<table>
<thead>
<tr>
<th>Dietary components</th>
<th>Total/number of cases</th>
<th>mean follow-up t (yr)</th>
<th>Main Results' [comparisons between quantiles (q) or the designated categories]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort studies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nutrients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omega-3 fatty acids[36]</td>
<td>7903/512</td>
<td>2.4</td>
<td>q4 vs q1: OR (95%CI) 0.65 (0.47-0.90). P for trend = 0.376</td>
</tr>
<tr>
<td>Trans fatty acids[37]</td>
<td>12059/657</td>
<td>6.1</td>
<td>q5 vs q1: 1.42 (1.09-1.84); P for trend = 0.003</td>
</tr>
<tr>
<td>MUFA[37]</td>
<td>12059/657</td>
<td>6.1</td>
<td>q5 vs q1: 0.80 (0.62-1.03); P for trend = 0.05</td>
</tr>
<tr>
<td>PUFA[37]</td>
<td>12059/657</td>
<td>6.1</td>
<td>q5 vs q1: 0.76 (0.59-0.99; P for trend = 0.03</td>
</tr>
<tr>
<td>Magnesium[123]</td>
<td>12939/737</td>
<td>6.3</td>
<td>q5 vs q1: 1.11 (0.77-1.59) No association.</td>
</tr>
<tr>
<td><strong>Nutritional adequacy</strong>[124]</td>
<td>13983/953</td>
<td>8.5</td>
<td>≥ 4 nutritional deficiencies vs ≤ 1: 1.37 (1.01-1.85)</td>
</tr>
<tr>
<td><strong>Total Fiber</strong></td>
<td>548/198</td>
<td>Cross-sectional</td>
<td>q4 vs q1: OR = 0.54 (95%CI: 0.11–2.63); Non significant association</td>
</tr>
<tr>
<td><strong>Foods and food groups</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fast food[125]</td>
<td>8964/493</td>
<td>6.2</td>
<td>q5 vs q1: 1.36 (1.02-1.81); P for trend = 0.03</td>
</tr>
<tr>
<td>Ultra-processed foods[126]</td>
<td>14907/774</td>
<td>10.3</td>
<td>q4 vs q1: 1.33 (1.07, 1.64); P for trend = 0.004</td>
</tr>
<tr>
<td>Baked goods[125]</td>
<td>8964/493</td>
<td>6.2</td>
<td>q2-5 vs q1: 1.38 (1.06-1.80); P for trend = 0.06</td>
</tr>
<tr>
<td>Olive oil[37]</td>
<td>12059/657</td>
<td>6.1</td>
<td>q5 vs q1: 0.80 (0.62-1.04); P for trend = 0.06</td>
</tr>
<tr>
<td>Seafoods[127]</td>
<td>6587/1367</td>
<td>Cross-sectional</td>
<td>q3 vs q1 of fatty fish consumption: OR 0.72 (0.59–0.88). Higher intakes were non-significant</td>
</tr>
<tr>
<td><strong>Coffee</strong>[128]</td>
<td>14413/199</td>
<td>10</td>
<td>≥ 4 cups/d vs &lt; 1 cup/d: 0.37 (0.15–0.95); P for trend = 0.220</td>
</tr>
<tr>
<td><strong>Yoghurt</strong>[36]</td>
<td>14542/727</td>
<td>9.3</td>
<td>Low fat: max vs min: 1.32 (1.06–1.65); P for trend = 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Whole fat: max vs min: 0.78 (0.63–0.98); P for trend = 0.02</td>
</tr>
<tr>
<td><strong>Dietary patterns</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD score[8]</td>
<td>10094/480</td>
<td>4.4</td>
<td>3pts vs 0-2pts: 0.74 (0.57-0.98)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4pts vs 0-2pts: 0.66 (0.50-0.86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5pts vs 0-2pts: 0.49 (0.36-0.67)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6-9pts vs 0-2pts: 0.58 (0.44-0.77); P for trend &lt; 0.001</td>
</tr>
<tr>
<td>MD score (updated)[103]</td>
<td>15093/1051</td>
<td>8.5</td>
<td>q5 vs q1: 0.84 (0.69-1.02); P for trend = 0.001</td>
</tr>
<tr>
<td>Pro-vegetarian[103]</td>
<td>15093/1051</td>
<td>8.5</td>
<td>q5 vs q1: 0.74 (0.61-0.89); P for trend &lt; 0.001</td>
</tr>
<tr>
<td>AHEI 2010[103]</td>
<td>15093/1051</td>
<td>8.5</td>
<td>q5 vs q1: 0.60 (0.49-0.72); P for trend &lt; 0.001</td>
</tr>
<tr>
<td>DASH-Fung diet[102]</td>
<td>14051/410</td>
<td>8</td>
<td>&lt; 2pts vs 3-9pts: 0.76 (0.61-0.94); P for trend = 0.010</td>
</tr>
<tr>
<td>Pro-inflammatory pattern[129]</td>
<td>15093/1051</td>
<td>8.5</td>
<td>q5 vs q1: 1.47 (1.17, 1.85); P for trend = 0.010</td>
</tr>
<tr>
<td><strong>Lifestyle</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediterranean lifestyle[130]</td>
<td>11800/806</td>
<td>8.5</td>
<td>High vs low : 0.50 (0.32-0.81)</td>
</tr>
<tr>
<td>Healthy lifestyle score[131]</td>
<td>14908/774</td>
<td>10.4</td>
<td>Max vs min: 0.68 (0.49-0.95); P for trend 0.010</td>
</tr>
<tr>
<td>Alcohol[132]</td>
<td>13619/459</td>
<td>8</td>
<td>Moderate vs no consumption: 0.62 (0.43-0.89) in female subjects</td>
</tr>
<tr>
<td>CV risk equation[133]</td>
<td>16739/927</td>
<td>6</td>
<td>q5 vs q1: 3.45 (2.40-4.95) starting at 40 yr</td>
</tr>
<tr>
<td>Physical activity[7]</td>
<td>10381/923</td>
<td>6</td>
<td>q5 vs q1: OR (95% CI) 0.81 (0.65-1.00) P for trend &lt; 0.01</td>
</tr>
<tr>
<td>Leisure time physical activity[134]</td>
<td>6343/311</td>
<td>10.5</td>
<td>&gt; 20 METs-h/wk vs &lt; 10 METs-h/wk: 0.84 (0.72-0.99); P for trend 0.046</td>
</tr>
<tr>
<td>Tobacco[135]</td>
<td>8556/190</td>
<td>4</td>
<td>Smokers vs non-smokers: 1.30 (0.97-1.73)</td>
</tr>
<tr>
<td>Childhood obesity (5 yr)[136]</td>
<td>11825/651</td>
<td>6.1</td>
<td>Obese vs low weight: 1.50 (1.06-2.12); P for trend = 0.012</td>
</tr>
</tbody>
</table>
Youth obesity (20 yr)[136] 11825/651 6.1 Obese vs low weight: 2.22 (1.20-4.08); P for trend = 0.005

Clinical trials
The PREDIMED study [119] 3923/224 5.4 MD + EVOO2 vs nut supplements vs Low fat diet: multivariable adjusted 0.85 (0.64 to 1.13); Non significant associations
The MoodFOOD Study [117] 512/54 1 Multi-nutrient supplement vs placebo groups: OR = 1.06 (0.87-1.29); Non significant associations.

1Estimates are presented as HR (95% CI) for incident depression unless stated otherwise.
2Extra virgin olive oil. MUFA: Monounsaturated fatty acids; PUFA: Polyunsaturated fatty acids; MD: Mediterranean diet.

Figure 1 Summary of interactions between the Mediterranean diet and the unipolar depressive disorder hypotheses. On the left side of the panel are the main food groups and their bioactive compounds, with known interactions with unipolar depressive disorder pathways. Each bioactive compound has a designated figure and color. The right side of the panel summarizes the aetiological hypotheses of depression and the main contributors to the disease. Black arrows designate the sequence of biological events in each mechanism, whereas "T" shaped lines designate an inhibitor effect. The figures designated for each compound are located within the sequence of biological events to designate the site of interaction with these mechanisms. See the main text for a detailed description of these interactions and associations. MD: Mediterranean diet; UDD: Unipolar depressive disorder; IL-6: Interleukin-6; CRP: C-reactive protein; TNF-α: Tumor necrosis factor-α; BDNF: Brain-derived neurotrophic factor.

Epidemiological evidence supporting MD
In recent years epidemiological evidence on the benefits of the MD on UDD include several studies both within and outside the Mediterranean area. In the InChianti cohort study including 1362 participants aged 18-102 years, adherence to a dietary pattern rich in vegetables, olive oil, grains, fruit, fish and moderate in wine and red and processed meat, labelled as the "typical Tuscan dietary pattern", was associated with a reduction in depressive symptoms over a 9-year follow-up period[99]. In the Greek center of the EPIC cohort, a total of 732 men and women 60 years or older were analyzed to determine the association between risk factors for cognitive function and depressive symptoms in the elderly. Adherence to the Mediterranean Diet, as well as olive oil consumption were associated with better scores in both outcomes[110]. Some non-Mediterranean reports include the study by Rienks et al[111]. In this study, the authors found that a Mediterranean-style diet was inversely associated with both prevalence (cross-sectional analysis) and incidence of depressive symptoms (longitudinal analysis) with an overall risk reduction of 17% among a sample of mid-aged women from the Australian Longitudinal Study on Women’s Health. In the Chicago
Health and Aging Project, MD adherence was associated with a reduced number of new depressive symptoms over 7.2 years of follow-up: 98.6% fewer symptoms among subjects in the highest tertile of adherence compared with those in the lowest[112].

So far, most of the evidence on diet and UDD derives from cross-sectional studies which could lead to false inferences of causal relationships. Furthermore, methodological analyses must address potential confounding, such as the adherence to a particular dietary pattern as a result of a recent diagnosis of chronic diseases or the co-occurrence of lifestyle-related and sociodemographic factors. These issues could be solved by carrying out large randomized trials with interventions based on changes in the overall food pattern. As of today, only small intervention trials have been carried out attempting to prove these and other hypotheses. The SMILES study was an intervention trial that promoted a healthy diet based on Australian and Greek dietary guidelines. In this trial, a dietary intervention with MD significantly improved depressive symptoms as an aid for standard acute depression treatment[113,114].

Focused on an elderly sample, the MedLey study aimed to determine the effect of an intervention with a Mediterranean dietary pattern on this population’s cognitive function, depressive symptoms, and quality of life. Despite significant improvements in endothelial function and MD adherence, the study showed no improvements in depressive symptoms nor quality of life after six months of intervention[115].

Examples of multi-country intervention trials include the MooDFOOD prevention trial which examined the feasibility and effectiveness of two different nutritional strategies (multi-nutrient supplementation and food-related behavioral change therapy) to prevent depression. The trial included individuals who were overweight, highly burdened with depressive symptoms, but did not meet clinical criteria for an episode of major depressive disorder in the last 6 mo[116]. In comparison to the placebo, micronutrient supplementation and behavioral therapy failed to prevent incident depression[117]. In contrast to these results, the HELFIMED trial has shown promising results in patients with self-reported depression. In this trial, a MD intervention supplemented with fish oil for 6 mo significantly improved depressive symptoms and quality of life[118]. To this day, the PREDIMED study is the longest trial (mean follow-up of 5.4 years) evaluating the effects of a dietary intervention on the risk of incident cases of clinical depression. In this trial the relative risk (RR) estimates for an intervention with MD supplemented with extra-virgin olive oil (9% risk reduction) or a Mediterranean Diet supplemented with mixed tree nuts (22% risk reduction) suggested an inverse association. However, the confidence intervals for both estimates were wide and they showed that the results were compatible with a null result. Even when both Mediterranean diets were merged together and analyzed as a single group, the results were not statistically significant. Only when the analysis was limited to participants with type 2 diabetes (approximately 50% of the sample), a significantly reduced risk of depression was observed only for participants in the group with nut supplementation (significant risk reduction: 41%)[119].

Future research of great interest to this field includes the upcoming results from the PREDI-DEP trial, which was designed to study the impact of the MD on UDD. In this trial, patients over 18 years with a record of a major depressive episode, but with total or partial remission in the previous 6 mo, were recruited in four centers across Spain. With a target sample size of 250 participants per arm, subjects were randomly allocated to a MD intervention with supplemental olive oil, or to a control group who received no dietary advice nor supplements. Over the course of 24 mo participants were assessed every 3 mo to evaluate MD adherence (in the group of intervention) and every 4 mo to ascertain changes in depressive symptoms (in the overall sample)[120].

Food frequency questionnaires, HRQoL questionnaires (Short-Form 36), and physical activity questionnaires were administered in addition to blood sample collections, which were collected at baseline, and yearly during two years[121]. The upcoming results of this trial will aid in the endeavor to elucidate the mechanisms behind diet and UDD, in addition to discerning the particular effects of a MD diet supplemented with olive oil on the incidence and overall prognosis of patients at risk for this disease.

**INITIAL DIETARY RECOMMENDATIONS FOR THE PREVENTION OF DEPRESSION**

Although there are a number of important gaps in the scientific literature to date, existing evidence suggests that a combination of healthful dietary practices (including the adherence to the Mediterranean Diet) may reduce the risk of developing depression. In 2015, a group of experts with several backgrounds (epidemiology,
psychiatry and nutrition) constructed the first Dietary Recommendations for the Prevention of Depression[122]. These dietary recommendations also provide additional and/or concurrent benefits for obesity, cardiovascular disease, diabetes and metabolic syndrome, and essentially pose no risk of harm. (1) “Follow ‘traditional’ Mediterranean Diet.”; (2) “Increase your consumption of fruits, vegetables, legumes, whole grain cereals, nuts and seeds.”; (3) “Include a high consumption of foods rich in omega-3 polyunsaturated fatty acids (PUFAs).”; (4) “Limit your intake of processed-foods, ‘fast’ foods, commercial bakery goods and sweets.”; and (5) “Replace unhealthy foods with wholesome nutritious foods.”[123]. Recommendations such as these are crucial steps towards a widespread implementation of diets as preventive measures for diseases, such as UDD, with substantial evidence of its benefits on cardiovascular and metabolic diseases.

CONCLUSION

UDD is and will remain a global health issue that researchers should strive to further understand and prevent. To this day, the hypothesized biological mechanisms that drive this disease include oxidative stress, HPA-axis dysregulation, monoamine and overall neuronal function deregulation. Based on analogous mechanisms, the MD could be a safe and inexpensive contributor to primary prevention strategies for depression, as it is characterized by various advantageous cardiovascular health qualities. Some of these properties include adequate B-group vitamin and folate content, promotion of a healthy gut microbiome, and high levels of various antioxidant compounds. Initial observational studies and trials, such as the PREDIMED study, suggest that implementing nutritional measures has an influence on disease risk and symptoms control. However, dedicated clinical trials are required to confirm the initial observations in order to fully support and understand the impact of diet on unipolar depressive disorder.

ACKNOWLEDGEMENTS

The group would like to acknowledge Ms. Maria Soledad Hershey, PhD candidate at the Department of Preventive Medicine and Public Health at the University of Navarre for providing the English language editing certificate and proofreading this manuscript.

REFERENCES

8 Sánchez-Villegas A, Delgado-Rodríguez M, Alonso A, Schlatter J, Lahortiga F, Serra Majem L,


18 Dahl J, Ornstad H, Aas HC, Malt UF, Bendz LT, Sandvik L, Brundin L, Andreaassen OA. The plasma levels of various cytokines are increased during ongoing depression and are reduced to normal levels after recovery. *Psychoneuroendocrinology* 2014; 45: 77-86 [PMID: 24845179 DOI: 10.1016/j.psyneuen.2014.03.019]


Randomized Controlled Trial. Depressive Symptoms in Overweight or Obese Adults on an Energy Restricted Diet in a Homovanillic Acid with Cocoa Extract Consumption Is Associated with the Alleviation of monoamines and anxiety symptoms in obese subjects with metabolic syndrome. Perez-Cornago A, Ramirez MJ, Zulet MA, Martínez-González MA. Intake of High-Fat Yogurt, but Not of Low-Fat Yogurt or Prebiotics, Is depression. A. C-reactive protein, interleukin-6, soluble tumor necrosis factor α receptor 2 and incident clinical depression. J Affect Disord 2014; 163: 25-32 [PMID: 24836084 DOI: 10.1016/j.jad.2014.03.023]


84 Dhar AK, Barton DA. Depression and the Link with Cardiovascular Disease. *Front Psychiatry* 2016; 7: 33 [PMID: 27047396 DOI: 10.3389/fpsyg.2016.00033]


Pano O et al. Healthy diet, depression and HRQoL

WJP | https://www.wjgnet.com

November 19, 2021 | Volume 11 | Issue 11

Pano O et al. Healthy diet, depression and HRQoL

Moore MN. Liver disease (NAFLD) prevention? Percentage of food restriction and liver histology: which alternative is the best for non-alcoholic fatty liver disease (NAFLD) prevention? [PMID: 30894319 DOI: 10.1016/j.jacc.2019.03.009]

Makovicky P, Tumova E, Volek Z, Makovicky P, Sedlacek R. Relationships between variable time, regression and depressive symptoms over time: the Invecchiare in Chianti (InCHIANTI) study. [PMID: 19594961 DOI: 10.1017/S0007114509990559]


Perez-Cornago A, Sanchez-Villegas A, Bes-Rastrollo M, Deaton A, Moleo P, Lahortiga-Ramos F, Martinez-González MA. Relationship between adherence to Dietary Approaches to Stop Hypertension (DASH) diet indices and incidence of depression during up to 8 years of follow-up. Public Health Nutr 2017; 20: 2383-2392 [PMID: 27335121 DOI: 10.1017/S1368946216001531]


Moore MN. Autophagy as a second level protective process in conferring resistance to...


123 Derom ML, Martínez-González MA, Sayón-Orea Mdel C, Bes-Rastrollo M, Beunza JJ, Sánchez-
Villegas A. Magnesium intake is not related to depression risk in Spanish university graduates. *J Nutr* 2012; 142: 1053-1059 [PMID: 22513990 DOI: 10.3945/jn.111.155572]


Sánchez-Villegas A, Pimenta AM, Beunza JJ, Guillen-Grima F, Toledo E, Martínez-Gonzalez MA. Childhood and young adult overweight/obesity and incidence of depression in the SUN project. *Obesity (Silver Spring)* 2010; 18: 1443-1448 [PMID: 19875985 DOI: 10.1038/oby.2009.375]
Neurodevelopmental disorders: An innovative perspective via the response to intervention model

Celestino Rodríguez, Debora Areces, Trinidad García, Marisol Cueli, Paloma Gonzalez-Castro

Author contributions: Areces D contributed conceptualization; Rodríguez C and Areces D contributed methodology, software; Areces D contributed formal analysis; Rodríguez C contributed investigation; Rodríguez C and Cueli M contributed resources; Areces D and García T contributed data curation; Areces D contributed original draft preparation; Rodríguez C and Gonzalez-Castro P contributed review and editing; Areces D contributed visualization; Rodríguez C contributed supervision, and project administration; all authors have read and agreed to the published version of the manuscript.

Conflict-of-interest statement: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Supported by Ministry of Sciences and Innovation I+D+i Project, No.

Abstract

Neurodevelopmental disorders are a group of conditions classified together by the most recent edition of the Diagnostic and Statistical Manual of Mental Disorders which include intellectual disability, communication disorders, autism spectrum disorder, attention-deficit/hyperactivity disorder, specific learning disorder (SLD), and motor disorders. SLD is present in many students, who exhibit significant difficulties in the acquisition of reading, written expression, and mathematics, mostly due to problems with executive functions (EF). The present study is a review of the current situation of neurodevelopmental disorders and SLD focusing on the benefits of the response to intervention model (RtI), which allows the combination of evaluation and intervention processes. It also addresses the key role of EF. The importance of adapting RtI to new possibilities such as the use of virtual reality is discussed and a theoretical framework for carrying that out is provided.

Key Words: Neurodevelopmental disorders; Specific learning disorder; Response to intervention model; Virtual reality

Core Tip: The present study aims to look deeply into detection of and intervention for students with neurodevelopmental disorders. It includes a description of the current situation of neurodevelopmental disorders focusing on specific learning disorder and executive functions as key aspects in students with problems in reading, writing, and mathematics. The study also presents the evaluation and intervention of neurodevelopmental disorders using an approach via the Response to Intervention model and provides a novel avenue for implementation related to the use of virtual reality.
INTRODUCTION

In recent years, one of the most pressing concerns at both national and international level has been achieving more effective education of students with neurodevelopmental disorders. This is evident in educational legislation and current social policies and has led education professionals to develop multiple activities in order to enhance student performance. This has meant dealing with diversity along with growing interest in the field of neurodevelopmental disorders and the students who suffer from them in one way or another[1]. Aware of this reality, the research field has produced a huge number of initiatives to help easily identify students who may be at risk of suffering from problems with their learning or their future behaviour at the earliest possible ages. At the same time, these initiatives have attempted to provide teachers with guidance and strategies so that they can provide timely instruction in pursuit of improved student academic progress and satisfactory school adjustment[2].

In this context, the present study aims to make a detailed examination of the detection of and intervention for students with neurodevelopmental disorders via a response to intervention model (RtI) framework. First, we describe the situation of neurodevelopmental disorders, focusing on specific learning disorders (SLD). Subsequently, we introduce the RtI model and provide an innovative approach for implementing it.

NEURODEVELOPMENTAL DISORDERS

Neurodevelopmental disorders are a group of conditions classified together by the latest edition of the Diagnostic and Statistical Manual of Mental Disorders[3] due to their common onset during childhood, high comorbidity rate of 20%-80%, and essential overlap of contributing factors across specific diagnoses[4]. This category includes intellectual disability, communication disorders, autism spectrum disorder, attention-deficit/hyperactivity disorder, SLD, and motor disorders.

SLD is a general concern in modern societies, where reading, writing, and mathematics are necessary skills for daily life[5]. However, students with SLD exhibit significant difficulties in the acquisition and use of these key skills (reading, written expression, and mathematics) despite having intact senses, normal intelligence, proper motivation, and adequate sociocultural opportunities.

Children and adolescents with SLD make up the largest group of students with educational needs in most countries[6]. According to the American Psychiatry Association[3] the prevalence of SLD in children across different languages and cultures is 5%-15% with rates of between 4 and 9% for reading deficits, and between 3 and 7% for deficits in mathematics. Bearing in mind the percentages of children exhibiting SLD, it represents one of the biggest focuses of interest in the educational community.

In recent years, research has begun to examine the common relationships and deficits in SLD (reading, writing, and mathematics) in order to better understand how they overlap rather than focusing on a single deficit, and also because findings suggest that children with a deficit in learning frequently exhibit deficits in other domains[7]. In addition, some genetic studies such as Kovas et al[8] have produced evidence that difficulties in reading, writing, and mathematics share genetic variations. More specifically, Kovas et al[8] found a genetic correlation of .67 between mathematics and reading difficulties which suggests a strong genetic overlap between them. In a meta-analytical review, Daucourt et al[9] indicated higher magnitude phenotypic and etiological overlaps between reading and mathematics difficulties. The consideration of common genetic risk factors linking SLD falls under the generalist genes hypothesis, which posits that the same genes underlie all cognitive abilities and disabilities both within and between academic domains[10].
Along these lines, the aetiology of SLD has been studied from general and specific domains. For example, at the specific level, low reading and writing skills have been shown to be related to problems in representation of sounds or phonological understanding, or decoding, whereas low numerical skills have been associated with a deficit in processing quantities or in the mapping of numerical symbols with mental representations of magnitude[11,12]. However, at the general level, researchers have looked at the role of attention, speed of processing, working memory, and executive functions (EF) in general.

The term EF refers to a mix of cognitive and behavioural processes such as planning, problem solving, sequencing, attention, inhibition, cognitive flexibility, and working memory, among others[13]. All of these processes allow good behavioural and emotional regulation, and have been shown to play an important role in learning in different areas such as reading, writing, and mathematics from very early ages.

While there are many different models which have attempted to give an account of the organization or structure of EF, one of the most influential models nowadays is from Miyake et al[14]. This model groups the different EF into 3 fundamental components: inhibitory control (also known as impulse control), cognitive flexibility (a range of attentional focus or “set shifting”), and working memory (or the ability to maintain information in mind while using it). According to this model, the different executive components would develop linked to the maturity of the individual. Miyake et al[14] indicated that although it is difficult to differentiate between EF at early ages, they seem to emerge separately in preschool ages (from 3-4 years old), and follow different development paths. At these ages, the children are gaining growing levels of voluntary control over attention, while at the same time being capable of keeping and representing certain information in mind, inhibiting certain responses using mental rules, and of responding and changing their focus of attention flexibly[15,16].

This developmental aspect has led to the study of EF as abilities related to school learning, generally in areas of reading, writing and mathematics[17]. Other longitudinal studies have shown the predictive value of EF measured in infancy on academic success in later educational stages[18-21]. The study by Morgan et al[19] with a sample of 18080 pre-school children showed how deficits in working memory and cognitive flexibility in pre-school ages posed a risk for the appearance of reading problems and difficulties with mathematics in the first year of elementary education. Similarly Clark et al[18] examined a group of 3-4 years old children and found a significant relationship between working memory and inhibitory control at these ages and performance in mathematics at 5 years old. As for reading, Birgisdóttir et al[22] found a relationship between various EF measures taken in preschool and reading understanding in first grade. These studies agree on the importance of EF as essential variables in learning, as well as the need to provide validated evaluation measures and intervention programs allowing work with students from early ages. In this context, the RtI model is one of the approaches which has begun to successfully spread in schools, specifically in early years schooling[23,24].

**RTI: A POWERFUL NEW MODEL**

The RtI model combines educational evaluation and intervention processes for all students within a classroom action system with 3 levels: Level 1- all students receive high quality, evidence-based instruction from teachers starting from an initial evaluation; level 2- those students who do not respond adequately in level 1 are given more explicit instruction with more frequent follow up; and level 3 - which is a supplement to levels 1 and 2 applied to those students who need more intensive and explicit teaching to meet their learning needs[25]. Throughout the three levels, students are evaluated in order to determine their progress and needs. Because of this, RtI is recognised as an extremely valuable model, allowing evaluation and intervention to be carried out with reference to a normative group, and given the starting point, evaluating improvement against the students’ own performance[26]. Figure 1 shows an example for a design following the RtI model considering the evaluation and intervention points and levels.

Considering the potential of the RtI model, it also represents a substantive change in the conceptualisation of SLD (a broader vision of how it progresses) in lockstep with social change and changes in the guidance this emerging practice refers to. The RtI model acts as both a preventive and palliative model[27], unlike the traditional (ability-achievement discrepancy) model in which it often takes years before children with SLD are treated[28]. The RtI model offers a solution to the main problem of
previous approaches, as children are detected early and receive appropriate intervention, leading to better prognoses than those whose problems are detected later [1,29]. For instance, in Spain, like in many other countries, an approach based on the ability-achievement discrepancy model is still used to detect students with SLD. These students are usually evaluated and diagnosed from the second year of primary school onwards, when they are already exhibiting significant delays in learning and it is often too late[30]. RtI has the potential to solve numerous problems compared to the ability-achievement discrepancy model (e.g., it provides poorly performing students with individualized instruction, it allows differentiation of students who have true disabilities from students who perform poorly due to lack of proper education)[31].

There is strong empirical evidence of the benefits of using the RtI model in classes in improvement in learning results for all students. Some of those studies focused on achievement in basic academic skills such as writing[32,33], reading[34-36], and mathematics[37]. For example, de León et al[38] observed the benefits of implementing the RtI model for intervention with students at risk of failing math. In addition, Jiménez et al[39] analysed the efficacy of an intervention for reading and math within the context of the RtI model. Their results showed that at-risk children in the intervention condition appeared to benefit more than at-risk children in the control condition (who did not receive intervention following the RtI model).

Other studies have focused on promoting RtI to help highly capable students to work to the best of their potential in the classroom[40], as well as students with emotional, social and behavioural needs[41]. In all of those cases, RtI was highly successful as it overlays an approach which includes combined educational evaluation and intervention. The evaluation process covers both initial and final student evaluation, along with all the intermediate progress measures in the skills being worked on. In this way, evaluation and intervention form part of the same continuum in the classrooms. It is only in those cases where the student does not improve, after participating in high quality interventions, that a student proceeds to a more exhaustive evaluation and diagnosis by specialists[42].

In early ages the model allows differentiation between students whose performance is below expectations because of a problem in instruction and those students who really have a neurodevelopmental disorder such as SLD, thus reducing the number of false positives. RtI is considered to be an effective means of determining whether students have received adequate instruction and thereby if their academic skills have improved without needing to be seen by special education services. Furthermore, RtI helps to differentiate as early as possible those students who have not improved despite specific instruction who might not be identified by other models that are less sensitive to intervention, in other words, false negatives[27].

In short, RtI is a promising model allowing the evaluation and identification of individuals with neurodevelopmental disorders such as SLD[43] and intervention according to their levels of need in an inclusive manner.

Considering the benefits of the model, the next question is how to implement the RtI model in relation to EF.
EF AND RTI MODEL

While the RtI paradigm is fundamentally based on the most common causes of difficulties, such as poor phonological awareness or basic mathematical skills, the literature about SLD indicates a wide variety of components that contribute to its appearance and progression, from genetic to cognitive and environmental factors. The literature also shows that focusing on a single component can be a significant restriction to the effective identification and treatment of neurodevelopmental disorders[44]. For this reason many authors have advocated a combined approach which includes evaluation and intervention addressing the cognitive processing areas or EF that may underlie the learning processes in the different areas of reading, writing, and mathematics[45]. Howard et al[45] argued that there is sufficient research on the relationship between EF and academic skills and that the evaluation of EF would provide information that would be useful for the design of individualised instruction which is the trademark of special education.

From this combined perspective, Hale and Fiorello[46] proposed what they called the Cognitive Hypothesis Testing Model, which is based on 4 main premises: (1) Academic performance depends on a large number of complex neuropsychological and cognitive processes; (2) Children normally have unique cognitive and learning profiles, based on different strengths and weaknesses; (3) These learning profiles should be evaluated via direct evaluation of the associated cognitive processes as well as via the analysis of the response to ecologically valid treatments; and (4) Academic difficulties should be remediated or counterbalanced according to the underlying cognitive strengths and weaknesses. The authors recommended that SLD should be approached from a broad evaluation-intervention perspective, based on problem-solving, and forming part of a continual process which includes RtI and a comprehensive evaluation of cognitive processes.

Despite advances at the theoretical level in this combined perspective of RtI-cognitive (executive) processes, it has not been a well-tried approach at the empirical level to date. The few studies carried out using this approach have so far given mixed results[47]. Furthermore, the design and implementation of new evaluation and treatment models for neurodevelopmental disorders must consider advances in the development of instruments, such as using virtual reality (VR).

VR, A NEW ENVIRONMENT

Using VR improves on the evaluation and intervention systems for disorders such as SLD, offering better ecological validity which leads directly to better specificity and sensitivity than current evaluation tests (thus reducing false positives). VR gives students the perception of being in a virtual environment that is similar to their everyday surroundings, and therefore the (visual, auditory and haptic) sensations and feelings produced will be similar to those in the real classroom[48]. Ultimately, VR is an innovative alternative for performing neuropsychological evaluations of various cognitive processes involved in learning[49,50]. More specifically, in terms of the usefulness of tests in VR, previous research indicates high validity in the evaluation of memory[51], attention[52-55], and other components such as planning and impulse control[56-59]. Numerous studies have demonstrated the advantages of VR for a variety of clinical conditions and groups of healthy participants[58,60-62].

Another great advantage of including VR in current evaluation and intervention systems is that it would allow the student to progress via a set virtual experience that adapts to the student’s own pace[63]. Various authors have stated that applications using VR have produced higher levels of motivation for learning[64,65]. These findings have even been reported in infant populations (aged between 3 and 4 years old), with not only increased motivation being observed, but also the use of this technology producing better concentration on the relevant stimuli[66].

VR has also been shown to be effective as an intervention tool. Various studies have indicated the advantages of using VR in intervention and rehabilitation programs in infant populations with intellectual disability[67], cerebral palsy as well as in treating learning difficulties and attentional problems[67-70]. One of the advantages indicated by the users themselves is that VR allows them to train different skills and abilities in a similar way to in the classroom without worrying about what others think of their mistakes and without embarrassment[68].

In short, VR offers the possibility of objectively analysing cognitive variables that are difficult to observe directly in the school environment, and at the same time it
allows tasks to be performed (both evaluation and intervention) with better ecological validity. Researchers often state that VR is more effective than evaluating children in laboratory settings where they are assessed individually, without classmates, under the supervision of an evaluator they do not know[70,71].

To adapt evaluation and intervention tasks to VR environments, one possibility is the innovative hardware from PICO interactive headsets which includes tools that allow the users to move, walk, crouch and do all manner of movements, without degrading the interaction with the virtual environment as there is no physical connection with any other device. With this hardware, a 3D context is created in which the participant is immersed in a natural, dynamic computer-generated environment which allows the creation of similar situations to those that happen in everyday life. In these environments, the type, amount and complexity of the presented stimuli are easily controlled, allowing the design of useful settings to gather objective indicators related to the participants’ execution. To achieve this objective, the designers use various technologies such as head mounted displays, tracking systems, head phones, movement sensors, and joysticks. The possibility of controlling all of these stimuli through a multi-sensorial experience, similar to real-life surroundings, increases ecological validity while at the same time maintaining methodological standards via the standardisation of the tests[50]. It also has an effect on participants’ motivation by giving them a more active, and interactive, role in the evaluation setting[59].

In summary, the features of this kind of hardware together with the advantages of the classic, current evaluation test, would achieve greater reliability and validity in the detection and intervention of neurodevelopmental disorders. Examples of this include the adaptation of classic tests, such as VR-Stroom[72], V-Tower (Virtual Tower of London) and V-West (Virtual Wisconsin Card Sorting Test)[73], along with the creation of new contexts such as the Virtual Store (V-Store)[73], the Virtual Action Plan-Supermarket (VAP-S)[74], the Virtual Supermarket (VMall)[37], and the well-known Virtual Classroom from Rizzo et al[55] and Armstrong et al[56].

Along with the complexity of evaluating within the RtI model, the evaluation of EF in the preschool stage is still difficult, especially considering the scarcity of EF assessment tools based on VR for this stage. One possibility for EF assessment in early years may be the adaptation of traditional measures. Examples include the Sorting Task and Animal Shifting (flexibility); the Animal Stroop, The Day and Night task, the Simon Task, and Local Global (inhibition); and Keep Track and Odd One Out (working memory)[16]. At the same time, in order to achieve more effective evaluation and intervention for SLD, VR tests must be adaptive in line with hypermedia systems.

THE NEXT STEP

The origins of educational hypermedia systems are linked to the appearance of the first intelligent tutors[75]. These tutors were student oriented and were based on the combination of study materials and the results of questionnaires, establishing an adaptive sequence for the course. As technology evolved, systems began to be developed known as Adaptive Educational Hypermedia Systems. The main objective of these systems is to adapt themselves to students based on certain characteristics such as what the student knows about the topic and the students’ learning styles or learning difficulties[76]. One of the things adaptive hypermedia systems bring to e-learning is the application of an adaptation model called the user model, which is the lens through which the e-learning system organises the visualisation of content, the order of presentation, the level of difficulty, and the type of feedback, based on various parameters (number of correct answers, errors, score, etc.)[77]. However the potential of the model applied to neurodevelopmental disorders is still to be developed, from some promising initial results[78,79]. For example, Cueli et al[90] found that students who worked with a hypermedia system in mathematics achieved better results in their knowledge of self-regulatory strategies. In addition, Taub and Azevedo [81] found that the use of hypermedia systems was more beneficial in the case of students with low prior knowledge.

Considering the potential of the RtI model for evaluation and intervention in EF and the benefits of VR and adaptive hypermedia systems, the future of evaluation and intervention in neurodevelopmental disorders should combine all of these main aspects. The challenge is to combine all of these new approaches (RtI, VR, and adaptive educational hypermedia systems) in order to improve EF in SLD students and incorporate this combination in schools. This will mean a new role for teachers who will have to support the process and the students at the evaluation points and in
the intervention levels that are currently part of RtI but with the new possibilities provided by VR and adaptive educational hypermedia systems.

CONCLUSION

The potential benefit of the RtI model is that, through its dynamic evaluation-intervention-evaluation processes, it is an effective system for measuring the progress and achievement of students who are lagging behind. It is also a safe method for monitoring low achievement in all students, whether due to lack of interest or motivation, boredom, or any other factor influencing insufficient learning[40]. The continued use of RtI in the classroom also aims to help children of average ability who present emotional and behavioural problems[41]. When a child is found to be exhibiting maladaptive behaviour or negative behaviours they are also evaluated and, depending on the results, evidence-based educational practices are added to reduce these behaviours and to promote positive attitudes towards the child’s social and academic life.

The scientific evidence indicates that students’ social and emotional competencies also play a key role in producing appropriate development and successful learning[78, 79,81]. Nevertheless, when children begin to exhibit difficulties in language and other academic areas, they run a high risk of their socio-emotional abilities deteriorating[24]. For example, children with SLD who are not dealt with early often begin to develop less involvement in school activities, by not paying sufficient attention to instructions and by interacting less with their classmates, which can lead to delays in learning and achievement and increased likelihood of developing behavioural problems. Because of that, RtI has become a model of pedagogical action that is able to meet these students’ needs so that they can achieve expected academic and socio-emotional results in preschool, avoiding school failures and behavioural problems at the beginning of primary education.

In this way, RtI is an ideal opportunity for inclusively improving education for all students, both with and without neurodevelopmental disorders, through the use of innovative approaches as VR and adaptive and motivating strategies.

REFERENCES


Fox L, Carta JJ, Strain PS, Dunlap G, Hemmeter ML. Response to intervention and pyramid model. Infants Young Child 2010; 23: 3-13 [DOI: 10.1097/YIC.0b013e3181c816e2]


Kuo N. Why is response to intervention (RTI) so important that we should incorporate it into teacher education programs and how can online learning help? J Online Learn Teach 2014; 10: 610-624


Kemp-Koo D, Claypool T. Response to intervention (RTI) in the province of saskatchewan. Communiqué 2011; 39: 4-6

Vaughn S, Bos CS, Schum RS. Teaching students who are exceptional, diverse, and at risk in the general education classroom. Boston: Pearson Allyn & Bacon, 2011


Rodas de Ruiz P. An initial approach to RTI model for students of a minority language group with needs in literacy. Interam J Psychol 2014; 48: 194-202


Rodríguez C et al. Neurodevelopmental disorders: RtI model

10.1007/s10055-006-0036-4


74 Klinger E, Chemin I, Lebreton S, Marié RM. A virtual supermarket to assess cognitive planning. Cyberpsychol Behav 2004; 7: 292-293


How non-rapid eye movement sleep and Alzheimer pathology are linked

Annelies Falter, Maarten J A Van Den Bossche

ORCID number: Annelies Falter 0000-0002-7728-1088; Maarten J A Van Den Bossche 0000-0002-2989-6600.

Author contributions: Falter A performed the literature study; Van Den Bossche MJA planned and coordinated the review and supervised the literature study; Falter A and Van Den Bossche MJA wrote the manuscript; all authors have read and approved the final manuscript.

Conflict-of-interest statement: The authors declare that there is no conflict of interest for this article.

Supported by the Funds Malou Malou, Perano, Georgette Paulus, JMJS Breugelmans and Gabrielle, François and Christian, Managed by the King Baudouin Foundation of Belgium, No. 2021-J1990130-222081.

Country/Territory of origin: Belgium

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report’s scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B

Annelies Falter, Maarten J A Van Den Bossche, Department of Geriatric Psychiatry, University Psychiatric Center KU Leuven, Leuven 3000, Belgium

Maarten J A Van Den Bossche, Center for Neuropsychiatry, Research Group Psychiatry, Department of Neurosciences, Leuven Brain Institute, KU Leuven, Leuven 3000, Belgium

Corresponding author: Maarten J A Van Den Bossche, MD, PhD, Assistant Professor, Staff Physician, Department of Geriatric Psychiatry, University Psychiatric Center KU Leuven, Herestraat 49, Leuven 3000, Belgium. maarten.vandenbossche@upckuleuven.be

Abstract
Alzheimer’s disease (AD) is a multifactorial neurodegenerative disorder characterized by the presence of senile plaques and neurofibrillary tangles. Research attempts to identify characteristic factors that are associated with the presence of the AD pathology on the one hand and increase the risk of developing AD on the other. Changes in non-rapid eye movement (NREM) sleep may meet both requirements for various reasons. First, NREM-sleep is important for optimal memory function. In addition, studies report that the presence of AD pathology is associated with NREM-sleep changes. Finally, more and more results appear to suggest that sleep problems are not only a symptom of AD but can also increase the risk of AD. Several of these studies suggest that it is primarily a lack of NREM-sleep that is responsible for this increased risk. However, the majority investigated sleep only through subjective reporting, as a result of which NREM-sleep could not be analyzed separately. The aim of this literature study is therefore to present the results of the studies that relate the AD pathology and NREM-sleep (registered by electroencephalography). Furthermore, we try to evaluate whether NREM-sleep analysis could be used to support the diagnosis of AD and whether NREM-sleep deficiency could be a causal factor in the development of AD.

Key Words: Alzheimer’s disease; Mild cognitive impairment; Sleep; Non-rapid eye movement sleep; Amyloid beta-peptides; Tau proteins; Electroencephalography

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.
INTRODUCTION

Alzheimer’s disease

Worldwide, approximately 50 million people suffer from dementia. The most common cause of dementia is Alzheimer’s disease (AD)[1]. The number of people affected by this neurodegenerative disease is increasing every year. The WHO considers AD as a priority because it is one of the major causes of morbidity and mortality in the aging population.

Pathogenesis

The two main pathological abnormalities found in AD are amyloid plaques and neurofibrillary tangles. Familial forms of AD exist in which hereditary mutations lead to the development of the disease. However, sporadic occurrence is much more common. Amyloid plaques accumulate extracellularly. They are largely composed of amyloid-beta-42 (Aβ42) and to a lesser extent of amyloid-beta 40 (Aβ40), both products of amyloid precursor protein (APP) metabolism. Aβ42 is less soluble than Aβ40. Generally, they are first detected in the neocortex. The subcortical structures are affected at a later stage[2,3]. The neurofibrillary tangles consist of paired phosphorylated tau-helices. These tangles accumulate in the neuronal cytoplasm and this damages the neurons. In contrast to plaques, the accumulation of neurofibrillary tangles is more predictable. Tau pathology typically begins in the medial temporal lobe allocortex (MTL; consisting of the hippocampal region, perirhinal, entorhinal, and parahippocampal cortices), where amyloid plaques accumulate less frequently, and then spread to the associative neocortex. Neuronal loss and synaptic damage usually develop after the tangles have formed[2].

Although now sometimes challenged, the most accepted hypothesis is still that the disease process is triggered by an increased production of the pathological Aβ peptides from APP by β- and γ-secretase, and by a reduced elimination of these pathological peptides (the amyloid hypothesis). According to this hypothesis, tangle formation and neuronal dysfunction follow downstream of the plaques. However, there is a strong association between the clinical symptoms, neurodegeneration and the tangles, and tau can also induce neurodegeneration independently of Aβ. It could therefore also be possible that tau and Aβ exert their effect via other pathways and amplify each other’s harmful effects[4]. Both seem essential in the pathogenesis of AD, but it is still unclear how they precisely affect each other. While the current hypothesis is that the accumulation of these proteins is causally related to synaptic and neuronal loss in the brain, AD is considered a complex and multifactorial disorder, in which both genetic and environmental factors affect the development and course of the disease[4].

Clinical manifestations

Progressive problems with episodic memory are usually the first complaint of AD patients. In an early stage, mainly short-term memory is involved. However, as AD progresses, other cognitive domains are also affected. Often topographic disori-
entation, problems with word-finding, sentence comprehension and behavioral problems will follow[2]. When there is an objective memory impairment without affecting the other cognitive domains and with a limited impact on daily functioning, this is called amnestic mild cognitive impairment (aMCI). This is seen as an intermediate stage, as individuals with aMCI have a strong increased risk of progressing to AD[5].

Although memory complaints are usually the first symptom that alerts both the environment and the patient to AD, sleep problems often precede these by several years. However, problems with sleep without the presence of memory complaints are often not immediately associated with AD. As the disease progresses, the severity of these sleep problems grows, and this is accompanied by a major impact on the quality of life of both the patient and the people around.

In addition, sleep problems are no longer only considered to be a major symptom of AD, but also a factor that may adversely affect the pathogenesis and symptomatology of this neurodegenerative disease. Episodic memory impairment is indeed the main characteristic of AD, but sleep, independent of AD, is essential for the optimal functioning of this cognitive domain.

**Non-rapid eye movement-sleep**

Sleep, based on electroencephalogram (EEG) oscillations, is divided into non-rapid eye movement (NREM1-, NREM2-, NREM3- and REM-sleep. During a sleep cycle, different phases occur. One night has multiple cycles. Two characteristics of REM-sleep are low amplitude desynchronized fast waves and muscle atony. NREM1 is a short transition phase between awake and sleep in which alpha waves (10-15 Hz) decrease while theta waves (4-8 Hz) become predominant. Sleep spindles (11-15 Hz) and K-complexes (KC; < 1 Hz) that occur on a background of theta waves are typical of NREM2. Spindles (11-15 Hz) are generated in the thalamus and synchronized by thalamo-cortical interactions[6,7]. KC are thought to protect sleep from external stimuli. During NREM3 (also called slow-wave sleep; SWS), delta waves, waves with low frequency (0.5-4 Hz) and high amplitude, are predominant along with slow oscillations (SO), slow waves (< 1.25 Hz) with high amplitude[8]. The activity of delta waves and SO together is sometimes called slow-wave activity (SWA). The SO arise from the superficial layers of the frontal cortex. Spindles are also present during NREM3. Sharp wave ripples (SWR) (80-140 Hz), which can be described as flares of synchronized neuronal activity in the hippocampus, are a third feature of NREM3.

For storing information in the long-term memory, NREM-sleep in particular seems to be important. To ensure a transfer of information, the neurons that represent a specific memory in the hippocampus are repeatedly activated. If the neocortex is activated simultaneously, the transfer of information takes place[9]. This co-activation gives the slow-learning cortex the opportunity to integrate the new information into the long-term memory. This happens mainly during NREM3. Especially SWR, spindles and SO appear to be essential for this process. A recent study[10] concluded that it is the exact timing and precise coupling of SWR, spindles and SO that ensure the transfer of information from the hippocampus to the cortex. The SWR precedes the spindle and activates the neuronal sequence associated with a recent memory. The spindles and the subsequent cortical SO help process this memory. The timing of these events determines the quality of the information transfer that takes place during the SWS[8].

**NREM-sleep, normal aging and AD**

Aging is associated with physiological changes in sleep patterns. In general, older people sleep less, go to bed earlier, get up earlier, sleep more during the day, wake up more at night and have difficulties in falling asleep again[11]. Age-related changes in the oscillations can also be observed on the EEG. Both the quantity and quality of NREM3 decreases while the length of NREM1 in particular and to a lesser extent NREM2 increases[12]. Older people have a decreased percentage of SWA and the amplitude is on average lower compared to younger people. This difference is most pronounced in the frontal brain regions where the SWA originates[13]. The density, duration and peak and mean amplitude of sleep spindles also decrease with aging[14, 15].

As early initiation of therapy usually yields better results, attempts have been made to identify symptoms that are specific to AD and that can be measured early in the course of the disease. The previous paragraph indicates that sleep quantity and quality decreases with aging even in the absence of neurodegenerative pathology, and are therefore not unique to AD. Nevertheless, interest in sleep in AD has increased rapidly in recent years. This is because, on the one hand, specific sleep problems such as
excessive daytime sleepiness and insomnia are more prevalent in AD. On the other hand, these sleep problems, similar to the formation of senile plaques and neurofibrillary tangles, seem to appear several years before the cognitive decline. As AD is considered a multifactorial disease, another focus of current research is identifying risk factors. Sleep deprivation (SD; no sleep for one night) and Alzheimer’s pathology such as senile plaques and neurofibrillary tangles have been shown to be associated. Thus, there may be a bi-directional relationship between sleep problems and AD.

As NREM-sleep is of great importance for proper functioning of episodic memory, which is also the cognitive domain that often deteriorates first in AD, several studies have concluded that NREM-sleep in particular is responsible for the association found between AD-pathology and sleep. However, the majority of the studies on this topic analyzed sleep through subjective reporting and did not differentiate between the different sleep stages. The purpose of this literature review is therefore to provide an overview of the literature investigating the relationship between NREM-sleep and AD in humans. In addition, studies will be discussed that specifically investigated the relationship between AD-pathology (in particular amyloid and tau) and NREM-sleep and which analyzed the possible causative role of NREM-sleep in the pathogenesis of AD. Subsequently, we briefly review the experiments that investigated whether NREM-sleep might be one of the pathways through which AD-pathology affects memory. Finally, based on the results found, it is discussed whether measuring NREM-sleep via EEG could in the future be used in the prevention, diagnosis and treatment of AD.

**SEARCH METHODS**

We performed a search of the literature using the PubMed, Embase and Cochrane Library databases. The search terms used at the start of the literature study were “Alzheimer’s disease” AND “NREM” OR “non-rapid eye movement” OR “nonREM” OR “SWS” OR “slow-wave”.

English-language studies in which human sleep was objectively measured by EEG and which included aMCI or AD as the study population were included. Studies analyzing the sleep of subjects without cognitive impairment were included only if the purpose of these studies was to gain more insight into the association between AD and NREM-sleep, and if amyloid-beta or tau measurements were compared with NREM-EEG data. The references of the above articles were also searched for additional articles that met the inclusion criteria.

**RESULTS**

One hundred and twenty-nine individual studies were found and screened, of which 31 met our inclusion criteria. An overview of the article selection is shown schematically in Figure 1. The results found are further described below.

**NREM-sleep changes in AD and aMCI cognitive impairment**

Prinz et al[16] and Vitiello et al[17] were among the first to detect that the percentage of NREM (duration NREM/duration total sleep) was significantly lower in AD patients than in elderly with intact cognition (EIC). More recent studies[6,18,19] also confirmed that people suffering from AD have a lower percentage of NREM3 compared to EIC. In several studies[6,18,20,21], a smaller percentage of NREM-sleep was associated with lower MMSE scores, not only in the total study group but also within the AD group. As diagnostic markers that can contribute to early identification of AD are sought after, Westerberg et al[22] studied NREM in aMCI patients. They found that the percentage of NREM3 was less in the aMCI group than in EIC. Within this aMCI group, patients with a lower percentage of NREM-sleep scored worse on a test in which they were asked to reproduce in the morning what they had studied the night before[22]. Although the mean percentage of NREM3-sleep in aMCI patients in another study[19] was lower than in the EIC group, this difference was not significant. This study by Gorgoni et al[19] also included AD patients and they found a significant difference in NREM3-sleep percentage between AD patients and EIC. The findings by Reda et al[20] were similar: The difference in percentage of NREM3 was not significant between aMCI and AD, but significant between AD and the EIC group. In contrast to Gorgoni et al[19], the difference between aMCI and the EIC group in this study did
Figure 1 Schematic overview of article selection.

However, many functions of NREM sleep and its microstructures are not yet known. What is known is that SWR, SO, spindles and KC, that mainly occur during NREM sleep, are important for our memory function. Therefore, these microstructures of NREM sleep were also investigated in patients with aMCI or AD. Rauchs et al[14] hypothesized that patients with AD would have fewer spindles compared to EIC, but this difference was not significant. In a more in-depth analysis whereby a distinction was made between fast (13-15 Hz) and slow (11-13 Hz) spindles[9], only the amount of fast spindles was significantly reduced in AD patients[14]. Westerberg et al[22] made a distinction based on location, and studied the spindles separately in a parietal and a frontal lead, where the fast and slow spindles, respectively, are most pronounced[7]. They noted a significant reduction in the aMCI group of the fast spindles in the frontal but not in the parietal lead. For the slow spindles, there was no difference in both leads[22]. The reduction of fast spindles was confirmed by Gorgoni et al[19]. They found a significant reduction in density (spindle number/length NREM sleep) of all spindles and of the fast spindles registered via the parietal lead in aMCI and AD, while the slow spindles did not differ. There was no difference between aMCI and AD[19]. A recent study[18] analyzed the total amount of spindles during NREM sleep. The density, amplitude and duration of these spindles were lower in AD than in aMCI and EIC. They were also lower in aMCI compared to EIC. Fast spindle density and spindle amplitude were both positively correlated with the MMSE scores[18]. Within the aMCI and AD subgroups, the occurrence of the spindles was positively correlated with the performance on a memory test in which the studied material was examined in the morning[19,22].

De Gennaro et al[6] also analyzed the KC. AD patients had a drastic decrease (40%) in KC-density during NREM sleep. This decrease was most pronounced in the frontal lead[6]. Reda et al[20] reported similar observations. Patients with AD had a lower KC-density. The density was significantly less not only compared with EIC, but also with the aMCI group. The difference was not significant between aMCI patients and EIC[20]. A recent study[18] measured the frontal KC-density in EIC, aMCI and AD. There was no significant difference in density between EIC and aMCI, while the density and amplitude were significantly lower in AD. Only the KC-amplitude differed between aMCI and EIC. No difference was found within the three groups in the duration of the KC[18]. The first prospective cohort study[24] on KC-complexes in aMCI confirms most of these results. In this study, the sleep of both aMCI and EIC was recorded after specific time intervals during a follow-up period of 2 years. Subjects who developed AD and subjects whose aMCI status remained stable during two years of follow-up,
Falter A et al. NREM sleep and Alzheimer pathology

were in the analysis divided in a progressive aMCI group (pMCI) and a stable aMCI group (sMCI), respectively. The KC-density at baseline and at 6 mo did not significantly differ among the three groups. In contrast, the KC-amplitude was significantly lower in sMCI and pMCI compared with EIC, while there was no difference between sMCI and pMCI subjects. At 12 and 24 mo, the KC-density and KC-amplitude was higher in EIC compared with sMCI and pMCI, but was also higher in sMCI than in pMCI. At 12 mo, when the pMCI subjects did not yet meet the criteria of AD, the KC-density and KC-amplitude was nevertheless already significantly lower in pMCI than in sMCI. The KC-density correlated positively with the MMSE scores in all studies\[6,18,24\].

**The link between NREM-sleep and Aβ and tau**

Aβ42 and tau are pivotal in the pathogenesis of AD. Therefore, it was also investigated whether specific NREM EEG-abnormalities could reflect the presymptomatic presence of Aβ and tau in the cerebrospinal fluid (CSF) or on imaging.

**Aβ- and tau-specific positron emission tomography scans and the NREM-sleep EEG**

Mander et al\[25\] detected a relationship between Aβ deposition in the medial prefrontal cortex and a reduction in NREM3 in EIC. A detailed analysis showed that this association was due to an underlying association between Aβ and less activity in the 0.6-1 Hz range. This inverse relationship was strongest between Aβ in the medial prefrontal cortex and sleep recorded via the frontal EEG lead. There was no significant association with the total 1-4 Hz frequency spectrum\[25\].

Lucey et al\[26\] analyzed not only Aβ but also tau presence on the brain scans of older adults who were predominantly cognitively intact (some subjects with aMCI). The percentage of 1-4 Hz-SWA decreased when more Aβ and tau were seen on imaging. For Aβ, there was an inverse association between Aβ deposition in the frontal, temporal and parietal cortices and both the 1-2 Hz and 1-4 Hz activity. The total amount of tau, and tau at the level of the entorhinal, parahippocampal and inferior temporal areas, was in these regions most strongly associated with the 1-2 Hz activity. The associations found with tau were much stronger than those with Aβ, and the associations with Aβ disappeared after statistical correction for multiple testing\[26\]. In both studies\[25,26\], the associations found remained significant when corrected for cortical thickness. Winer et al\[10\] tried to determine whether there are EEG abnormalities that are only associated with Aβ and not with tau and vice versa. The MTL and the cortex were analyzed because these regions are particularly susceptible to tau- and Aβ-damage in AD, respectively\[2\]. They observed that a weaker coupling between SO and spindles was associated with more accumulation of tau in the MTL [10]. There was no association between the strength of the coupling and cortical Aβ. However, for Aβ, they found, in line with the results of Mander et al\[25\], a significant inverse correlation between cortical Aβ and activity in the 0-6-1 Hz range\[10\]. A reduction in 0.6-1 Hz-SWA predicted a higher amount of cortical Aβ. MTL-tau had no significant association with this frequency interval or with any other subgroup frequency of the 0.6-4 Hz range\[10\]. A new study of Winer et al\[26\] aimed to address the lack of longitudinal data on this topic. The sleep of EIC was registered during one night in the beginning of the study. Around the same time positron emission tomography (PET) scans were taken in order to analyze Aβ accumulation. During a follow-up period of on average 3.7 years, multiple follow-up PET scans were performed to measure the change in Aβ. They found a negative correlation between the proportion of 0.6-1 Hz SWA at baseline and the change in Aβ during the follow-up. This negative correlation was frequency range specific, as they did not find an association between total SWA (0.8-4.6 Hz) and Aβ change. In addition, the strength of SO-spindle coupling was not associated with the rate of Aβ change\[27\].

**Aβ- and tau-CSF concentrations and the NREM-sleep-EEG**

Lucey et al\[26\] observed no correlation between CSF-Aβ42 and total NREM-SWA while using a similar study design where imaging was replaced by CSF analysis. In this study\[26\], there was however a significant inverse relationship between the 1-4.5 Hz-SWA and the ratio between CSF-tau and CSF-Aβ42 (CSF-tau/CSF-Aβ42). The ratio between phosphorylated tau and CSF-Aβ42 (CSF-p-tau/CSF-Aβ42) was also analyzed separately and this relationship was also significant. The inverse relationship between CSF-tau/CSF-Aβ42 and the 1-2 Hz SWA was the strongest. Kam et al\[28\] explored associations with the microstructural features of NREM-sleep. A decrease in spindle density during NREM2 was correlated with an increase in CSF-tau and CSF-tau/Aβ42 in EIC. This relationship was strongest between specifically CSF-tau and fast spindle
density, while there was no relationship with the slow spindles. Frontally recorded 0.5-4 Hz-SWA was also inversely related with CSF-Aβ42 but not with CSF-tau[28]. In another study[29], CSF-Aβ42 was inversely related with the duration and percentage of NREM3, while there was no significant relationship between CSF-Aβ42, total sleep length and the length of the other sleep stages. Additionally, no association was found between the percentage of NREM3 and CSF-tau[29]. In mild and more advanced AD patients, there was only a correlation between CSF-tau and a decreased percentage of NREM3; this association did not apply to CSF-Aβ42 and the percentage of NREM3[21].

The impact of a NREM-sleep deficit on CSF-Aβ and CSF-tau
Several studies[30,31] have reported that CSF-Aβ fluctuates over the course of a 24-h period, with the highest concentration being measured in the evening and the lowest in the morning after sleep. Following this finding, and with the aim of gaining more insight into the relationship between NREM sleep and AD pathology, the effect of sleep deprivation (SD) on the CSF-Aβ and CSF-tau concentrations of adults with intact cognition were examined. Lumbar catheters were used in two studies[32,33] to frequently monitor CSF levels during both normal sleep and SD. In the study of Ooms et al[32] CSF-Aβ42 decreased significantly (6%) after sleep, while the concentration did not decrease after SD. Neither sleep nor SD significantly affected CSF-tau. Luc ey et al [33] found a similar but greater increase (30%) in CSF-Aβ after SD and, in contrast to Ooms et al[32], also noted an increase in CSF-tau (50%). Shokri et al[34] conducted their study assuming that radio tracers also identify soluble amyloid and therefore can show a direct increase in Aβ42 on imaging. They compared two scans of the same EIC that were taken after one night of normal sleep and after one night without sleep. SD resulted in a significant increase in Aβ42 in the right hippocampus, parahippocampus and thalamus on the second scan[34]. In an attempt to determine whether NREM3 specifically contributes to this decrease in CSF-Aβ following sleep, Ju et al[35] disrupted sleep only during NREM3. This intervention logically resulted in a reduced percentage of SWA and a reduced total sleep duration. The percentage of NREM1 increased, while the other sleep variables remained unchanged. After NREM sleep disturbance, CSF-Aβ40 increased. There was no significant association between CSF-Aβ40 and CSF-Aβ42 and NREM, REM, or total sleep length. No correlation was found between NREM3 disturbance and CSF-tau concentrations[35]. Olsson et al[36] wanted to address the lack of knowledge on the chronic effects of SD on AD markers by allowing subjects with intact cognition to sleep only partially (max. 4 h) in their sleep laboratory for five consecutive nights. In the same subjects, physiological sleep was also monitored during five other nights. There was an interval of one month between both circumstances. After the completion of each period, a lumbar puncture was performed. Their hypothesis was that five nights of partial sleep deprivation (PSD) would lead to increased CSF-Aβ or CSF-tau. However, they found no significant difference in CSF-Aβ40, CSF-Aβ42 or CSF-tau concentrations between the two conditions. EEG analysis showed that despite the shorter total sleep time (207 min) in PSD, there was no significant difference in NREM3 (4 min) length. The lengths of NREM1, NREM2 and REM were shorter with PSD (19 min, 131 min and 52 min, respectively)[36].

The link between Aβ and tau, memory and NREM-sleep
Episodic memory impairment is characteristic of AD. As NREM-sleep is important for this cognitive domain, research groups have attempted to investigate whether NREM-sleep is still important for memory function, when AD pathology is already present in the brain.

In the previously mentioned studies[14,19,22], there was in the total study population (EIC, aMCI and AD) a proportional relationship between the results on memory tests or MMSE scores and, among other things, the percentage of NREM3-sleep, the spindle density and the KC-density. Furthermore, this association persisted when aMCI and AD were considered separately. Liu et al[18] detected in their study the strongest associations between NREM3 percentage and MMSE, and between KC-density and MMSE. These results are based on experiments in which the physiological sleep of aMCI, AD and EIC was measured via EEG during one night. A recent study [37] did more than just record physiological sleep. It used an intervention to investigate the link between NREM sleep, memory and AD. aMCI patients in this study had to study information before they went to sleep. The following night their SWA was stimulated via acoustic stimulation. The same experiment was repeated but without stimulation during sleep. The first observation was that acoustic stimulation significantly increased SWA. The second one was that on average aMCI patients remembered more after a night of stimulation, although the difference did not reach
Falter A et al. NREM sleep and Alzheimer pathology

DISCUSSION

This review shows that there is an association between AD and NREM-sleep. The number of EEG studies examining this association is rather limited up to now, but current results suggest that NREM-sleep analysis could in the future be useful in the diagnosis, prevention and therapy of AD.

Some sleep problems, such as excessive daytime sleepiness and insomnia, are characteristic of AD. The studies discussed suggest that certain specific changes in NREM-sleep could help in the diagnosis of aMCI and AD. Based on the current literature, a decrease in the NREM3 percentage should be able to differentiate persons with AD or aMCI from EIC. Although in each study this percentage was lower in AD compared to aMCI, the difference was not always statistically significant. Nevertheless, this suggests that the percentage might decrease progressively during the course of the disease. Some microstructural properties of NREM-sleep were also found to be significantly different in AD. A reduction in fast spindle density, amplitude and duration could help to identify aMCI and AD patients. While the spindle density could not distinguish the aMCI group from the AD group in the study by Gorgoni et al[19], Liu et al[18] showed a difference in both density and amplitude between both groups. Measuring spindles and NREM3 percentage with EEG could therefore assist in diagnosis at an early stage. In contrast, KC-density was consistently lower in AD compared to the aMCI group. This parameter might therefore be especially useful at a later stage, as in the majority of experiments the KC-density was not different between aMCI and EIC. A reduction in KC-amplitude may occur earlier and might be useful in distinguishing aMCI from EIC[24].

An association between sleep and AD pathology has been demonstrated before. Aβ and tau were associated with less qualitative sleep in humans[38]. However, the majority of these studies evaluated sleep subjectively and therefore could not differentiate between the different sleep stages. The discussed results suggest that AD pathology may cause EEG abnormalities particularly in NREM-sleep. The in vivo measurement of Aβ and tau via brain imaging is used in the clinic to support the diagnosis of AD. Pathological abnormalities on imaging characteristic of AD co-occurred with specific EEG abnormalities during NREM-sleep in the reviewed studies. A reduced percentage of 0.6-1 Hz-SWA reflected increased cortical Aβ in both Winer et al[10] and Mander et al[25], while there was no association with tau. Due to technical limitations, Lucey et al[26] did not analyze below 1 Hz, so they could neither confirm nor refute these results. In contrast, less than 1-2 Hz SWA could indicate tau pathology in the MTL and the orbitofrontal, inferior parietal and inferior temporal cortex[26]. Winer et al[10] did not confirm this association between tau and SWA, but identified a microstructural feature specifically associated with tau. The SO-spindle-coupling, and in particular the progressive loss of coupling intensity, might identify individuals with high MTL-tau earlier. The on average older population in the study by Lucey et al[26] could possibly explain the fact that they did detect an association between tau and SWA. As mentioned in the introduction, tau indeed first accumulates in the MTL and the orbitofrontal, inferior parietal and inferior temporal cortex.

An increase in the ratio of CSF-tau/CSF-Aβ reflects an increased risk of AD. Although there is more debate about the predictive value of CSF-Aβ, the literature suggests that CSF-Aβ increases in a preclinical stage which promotes plaque formation and decreases at a later stage only when plaques are more numerous[39]. These CSF biomarkers were also associated with NREM abnormalities in EIC. Reduced fast spindle density during NREM2 and 1-2 Hz-SWA were associated with an increase in CSF-p-tau, CSF-tau and CSF-tau/CSF-Aβ[26,28]. CSF-tau and CSF-p-tau are markers of neuronal damage and neurofibrillary tangles, respectively[40]. These NREM changes could potentially
be an early indicator of tau pathology. The results for amyloid were less consistent. Kam et al[28] and Varga et al[29] reported an inverse association between the percentage of NREM3 and CSF-Aβ42. Lucey et al[26] could not confirm this association. Again, their study population could have some influence on the results, as there were also aMCI patients in this study, and the presence of plaques might reduce CSF-Aβ fluctuations. Hence, EEG measurements, like CSF-analysis and imaging, could be used to support the diagnosis of AD and to identify individuals at increased risk in a non-invasive way. More research is needed however, to both confirm the results discussed and to quantify these abnormalities into clinically useful parameters.

Measuring NREM-sleep might have even more clinical potential. It has been acknowledged for some time that chronic sleep problems entail an increased risk of developing AD. For example, the risk that adults with intact cognition but with untreated obstructive sleep apnea hypopnea syndrome or insomnia will develop AD later on is higher than in persons without a sleep disorder[41-43]. Over the last decade, evidence is mounting that a lack of NREM3-sleep may be an important cause of this increased risk, and that NREM3 may play a causal role in the pathogenesis of AD. However, the design of the above-mentioned studies usually did not allow determination of the causality in the association found between CSF concentrations or imaging and NREM-sleep abnormalities. Additional arguments that sleep problems could potentially lead to Alzheimer’s pathology are found in studies in which participants are exposed to acute SD. All SD studies[32,33,35] discussed a CSF-Aβ concentration that was higher after SD than after normal sleep. For tau, the effects were less consistent. In the majority of cases, no nocturnal change in CSF-tau was observed in either physiological sleep or SD. One study[36] found a significant increase in CSF tau after SD[35]. One possible way in which a lack of NREM3 sleep could lead to a rise in certain concentrations in the CSF is through the so-called “glymphatic system”[44]. This perivascular network, found in animal models, promotes the elimination of soluble proteins, including Aβ, from the central nervous system. The glymphatic system is mainly active during sleep and its function is said to be maximal during SWS[45]. The findings of Ju et al[35] that a selective disruption of SWA was inversely correlated with an increase in CSF amyloid, and that there was no association with total sleep or REM sleep, are in line with this. The selective interruption of NREM3-sleep had no effect on CSF tau. This could possibly be due to the longer half-life of tau, which means that one night without sleep has a lesser effect on the clearance of tau[46]. On the other hand, it is also possible that sleep has a different effect on the clearance of Aβ and tau. The fact that NREM3-sleep was ensured with PSD and that PSD did not induce an increase in CSF-Aβ42 or CSF-tau also supports the hypothesis that it is NREM3 that mainly affects CSF concentrations[36]. Although it seems likely that there is an increased clearance of CSF-Aβ42 during NREM3, an increased synthesis, as suggested by Lucey et al[26], could also be partly responsible for an increase in CSF-Aβ42 after SD. A combination of reduced clearance and increased synthesis of CSF-AB42 during SD is also possible.

Studies on humans, combining EEG recording, CSF measurements and imaging over a long period of time are needed to gain more insight into the effect of NREM-sleep on Aβ and tau and to determine whether the CSF fluctuations are due to increased clearance or to increased synthesis of these proteins. Despite the lack of clear understanding about the mechanism, ensuring NREM-sleep could become important in the prevention of AD. In addition, stimulation of SWS could be used to normalize CSF-Aβ and in this way reduce the risk of plaques. SWS stimulation also has potential as a treatment for the memory symptoms that characterize AD[37]. Problems with the transfer of information to long-term memory are namely one of the first complaints of AD. NREM-sleep and the interaction of SO, spindles and SWR play an important role in this information transfer[8]. Acoustic stimulation during SWA has previously been shown to improve memory tests in individuals with intact cognition[47]. SWS stimulation was also able to improve memory in aMCI patients[37].

The hypothesis of Mander et al[25] that a reduction in SWA is one of the mechanisms by which cortical Aβ impedes the storage of information in the long-term memory, could provide a possible explanation for the observation that, despite the presence of AD pathology in an aMCI stage, SWA stimulation still had a positive effect on memory. This may explain why the decline of hippocampus-dependent memory is usually one of the first symptoms of AD, while Aβ usually does not accumulate at the hippocampus until later in the disease process. There is moreover a similarity between the regions that generate NREM-SWA and the cortical areas where Aβ usually first accumulates[48]. A similar EEG study has not yet been performed in aMCI patients. However, two recent studies[49,50] in a heterogeneous group of elderly (EIC, aMCI and AD) also suggested that sleep at an early stage may partly influence the impact of
cortical Aβ on memory, as the inverse relation between cortical Aβ on imaging and memory was statistically dependent on sleep continuity as an intermediate factor. Interestingly, EIC with high cortical Aβ had a more continuous sleep in this study than the EIC with little Aβ, which might be indicative of a possible protective role of sleep on cognition in the presence of high Aβ accumulation\[49,50\]. These studies all focused on Aβ. Whether NREM-sleep disturbance is also one of the pathways through which tau affects memory is not yet known.

Qualitative NREM3-sleep seems to be essential for memory in both EIC and AD patients. Enhancement of SWA, such as by acoustic stimulation, might therefore be a potential therapy for memory problems in AD patients.

**CONCLUSION**

This review discussed the studies that examined the link between NREM-sleep and AD. Based on the presented results, examining NREM-sleep could be a new diagnostic tool for detecting aMCI and AD in a non-invasive way. Enhancement of NREM-sleep could be an exciting new option in the prevention and treatment of AD. Prospective and longitudinal studies that combine EEG measurements with imaging and CSF analyses are needed to gain more insight into the causality and the underlying mechanisms of this relationship.

**REFERENCES**


Falter A et al. NREM sleep and Alzheimer pathology


Empirically supported psychological treatments: The challenges of comorbid psychiatric and behavioral disorders in people with intellectual disability

Laura E Gómez, Patricia Navas, Miguel Ángel Verdugo, Marc J Tassé

ORCID number: Laura E Gómez 0000-0002-0776-1836; Patricia Navas 0000-0002-5411-4025; Miguel Ángel Verdugo 0000-0002-5802-8220; Marc J Tassé 0000-0001-6168-669X.

Author contributions: Gómez LE wrote the draft; Navas P, Verdugo MA and Tassé MJ reviewed the draft and contributed equally to this work.

Conflict-of-interest statement: Authors declare no conflict of interests for this article.


Country/Territory of origin: Spain

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report's scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

Open-Access: This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract
This paper reviews the current state of knowledge on psychological interventions with empirical evidence of efficacy in treating common psychiatric and behavioral disorders in people with intellectual disability (ID) at all stages of their life. We begin with a brief presentation of what is meant by psychiatric and behavioral disorders in this population, along with an explanation of some of the factors that contribute to the increased psychosocial vulnerability of this group to present with these problems. We then conduct a review of empirically supported psychological therapies used to treat psychiatric and behavioral disorders in people with ID. The review is structured around the three generations of therapies: Applied behavior analysis (e.g., positive behavior support), cognitive behavioral therapies (e.g., mindfulness-based cognitive therapy), and contextual therapies (e.g., dialectical behavior therapy). We conclude with some recommendations for professional practice in the fields of ID and psychiatry.

Key Words: Behavior disorder; Intellectual disability; Psychiatry; Applied behavior analysis; Cognitive behavioral therapies; Contextual therapies

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.
Core Tip: Psychiatric and behavioral disorders are more common in people with intellectual disability (ID) than in the population without ID. Although there is some consensus that empirically supported psychological treatments for psychiatric disorders in the general population can be adapted for people with ID, there is scant published evidence about their application in this population. We discuss the empirical evidence for the application of psychological treatments in people with ID. Our review is structured around the first-, second- and third-generation therapies.

INTRODUCTION

Despite the significant challenges in establishing an accurate figure, it is estimated that approximately 1% of the population has some form of intellectual disability (ID)[1]. The causes of ID can be extremely varied, ranging from genetic alterations (e.g., chromosomal or single gene alterations) that are often associated with more severe forms of ID, to environmental factors (e.g., nutritional deficiencies, extreme social deprivation, or child abuse) that are often linked to milder forms.

Whatever the cause, the recommendation when defining ID is to adopt a biopsychosocial approach in which disability is conceived as the interaction of the person with their context[2,3]. Environments can therefore be more or less facilitating, and as a result, attenuate or accentuate the disability depending on existing barriers and supports. ID is a neurodevelopmental disorder whose diagnostic criteria can be found in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)[4] and the 11th edition of the International Statistical Classification of Diseases and Related Health Problems (ICD-11)[5]. It is characterized by significant limitations in intellectual functioning and adaptive behavior that appear before the age of 22 years [6].

For most of the disorders in these manuals, the diagnosis of a mental disorder relies on clinical judgment. To diagnose ID, clinical judgment is also important in the selection, administration, and interpretation of standardized tests used to determine the “significance” of the limitations in the person’s intellectual and adaptive functioning, which is defined as two standard deviations or more below the population mean. This article does not focus on psychological treatments for the core symptoms of ID (for an interesting review on this topic see Thurm et al[7]); rather, it discusses the treatment of comorbid psychiatric or behavioral disorders in people with ID, the reason being that most interventions with this population are tertiary prevention methods aimed at treating associated conditions rather than its core features. These comorbidities are in fact one of the main causes of exclusion for people with ID and one of the most frequent reasons why they are referred to mental health services.

PSYCHIATRIC AND BEHAVIORAL PROBLEMS

There is a growing body of literature on the prevalence and treatment of psychiatric and behavioral problems in people with ID[8]. Most of the studies report that such problems are common across the lifespan for this population group. Data on prevalence rise up to 49%[9], and rates are higher in people with more severe disability[10].

The range of problem behaviors exhibited is vast, the most frequent being self-injury, aggression toward others, disruptive behaviors, inappropriate sexual behaviors, stereotyped behaviors, and pica[11]. Problem behaviors may have a biological, psychological, social, or environmental etiology; or they may be the result of a combination of these factors. They may be a response to a specific social situation (e.g., rejection), stress (e.g., a major transition, loss of a loved one), or may appear as part
of the symptomatology of a psychiatric disorder (e.g., elevated irritability or paranoid ideation associated with a psychotic disorder may lead to aggression).

Thus, problem behaviors are not defined as a diagnostic category, but are framed within a contextual, non-etiologic model where the behaviors are not problematic per se. Whether they are considered to be a problem behavior depends on a number of factors, such as their characteristics (i.e., frequency, duration, intensity), personal variables (e.g., age), or contextual variables (e.g., social norms that govern the environment in which the behavior occurs; the interpretation and meaning given to the behavior by observers). It is vitally important to understand that a problem behavior serves a function; the responsibility of professionals is first-and-foremost to find out what the person wants or needs (e.g., attention, tangible), trying to communicate (e.g., pain, discomfort), or trying to escape (e.g., demand or situation), which will then allow them to design and implement an intervention that responds to the function of said behavior so that the problem behavior is no longer needed[12].

Attitudes toward psychiatric disorders in people with ID have evolved over the years: first, there was outright denial (e.g., until the mid-twentieth century, the possibility that this population might experience psychiatric disorders was not even recognized); there has also been a phenomenon known as diagnostic overshadowing (where the symptoms arising from physical or psychiatric disorders are misattributed to ID, leading to denied diagnosis and treatment); and, at the other extreme, it is common today to have reports of prevalence rates that double or triple those observed in the population without ID. In childhood and adolescence, for example, the most commonly reported comorbid disorders in this group are attention-deficit/hyperactivity disorder (30%), problem behaviors, autism spectrum disorder (ASD), anxiety disorders, and depressive disorders[9].

A study by Hughes-McCormack et al[13] confirmed ID as a predictor of psychiatric disorders [odds ratio (OR) = 7.1; 95% confidence interval (CI): 6.8-7.3], as they found a substantially higher prevalence of psychiatric disorders at all ages compared with the population without ID. The scientific literature also noted that when a diagnosis of ID is added to a diagnosis of ASD, the likelihood of other co-occurring psychiatric disorders is much higher than in the general population (OR = 130.8; 95%CI: 117.1-146.1)[14].

It is now widely accepted that people with ID experience the full range of psychiatric disorders as people without ID. The disagreement lies in whether these disorders manifest themselves (i.e., signs and symptoms) in the same way or rather in an “atypical” way. As Tasse[15] pointed out, some clinicians have proposed using certain behaviors as equivalents to DSM symptoms, whereas others feel that there is lack of evidence to support behavioral equivalents. The result is a dichotomy opposing those who defend the use of conventional diagnostic systems (DSM, ICD) and those who prefer adaptations specifically developed for people with ID. Reference manuals for the latter would include, for example, the Diagnostic Manual-Intellectual Disability (DM-ID-2)[16] — an adaptation of the DSM-5 — and the Diagnostic Criteria for Psychiatric Disorders for use with Adults with Learning Disabilities/Mental Retardation (DC-LD)[17] — an adapted version of the ICD-10.

**INCREASED VULNERABILITY**

The prevalence rates cited above can be explained by the increased biopsychosocial vulnerability in people with ID to develop problem behaviors and psychiatric disorders. It is thus generally accepted that there is a greater biological vulnerability, which predisposes people with ID, for example, to show certain behavioral phenotypes in specific genetic syndromes, alongside the interaction of other psychological and contextual factors often unique to people with ID. We will now look at some of the factors that are commonly associated with problem behaviors and psychiatric disorders.

**Health problems and difficulties accessing health services**

People with ID experience more health problems than the general population. They are more frequently affected by cardiovascular diseases, obesity, diabetes, epilepsy, chronic constipation, kidney disease, osteoarticular disorders, and thyroid disorders[18]. They are also more likely to have poorer oral health[19], lower physical activity levels, higher sedentary behavior, and poorer sleep hygiene[20].

People with ID experience a number of barriers to accessing preventive health care services compared to the general population[21], to receiving routine medical care[22].
and accessing mental health services[23]. The scientific literature cites inadequate training of mental health professionals, organizational barriers, lack of services, and poor quality of services as the main obstacles to accessing appropriate mental health care[24]. Barriers are even greater for those with the most severe disabilities[25,26].

**Lack of training and negative attitudes among health professionals**

The attitudes of health professionals toward people with ID have a major impact on the supports and care they provide. Yet, as we have already mentioned, multiple studies have highlighted the lack of ID training among social and health care professionals[27,28], which has repercussions on the quality of emotional support they provide[29].

Among the difficulties in providing satisfactory care, professionals cite not only heavy workloads and resource constraints, but also their own feelings of insecurity, stress, frustration, embarrassment, and fear of dealing with this population[28,30,31]. Worse still, some studies have reported blatant examples of disrespect, such as making negative comments while the person is present, or completely ignoring the person with ID and addressing their companion instead[32]. It is not uncommon for health care providers to stereotype people with ID[31], to disregard this population’s preferences and decisions regarding their own medical care[33], or even to deny them health care services altogether[34].

A possible explanation for these negative attitudes is that a large proportion of health care professionals are unfamiliar with the essential components that make up the daily lives of people with ID[31]. In fact, many health care professionals have never had contact with this population[30], and yet contact and interaction are precisely what is required in order to educate health care professionals about life with a disability and promote more positive attitudes.

**Polypharmacy**

The issue of polypharmacy in people with ID is also frequent and worrisome. Treating multiple comorbidities and the frequent secondary health problems experienced by people with ID often results in complex polypharmacological interventions, with significant health implications and a high risk of adverse side effects and interactions between medications[28], many of which often go undetected[35].

Studies tend to converge in pointing out the high use of antibiotics, antihypertensive drugs, and lipid-lowering agents, which are administered together with anticonvulsants, psychotropics, antidepressants and antipsychotics. Bowring et al.[36] found that 71% of people with ID took some type of medication and, of these, 38% took psychotropics. In fact, it is estimated that between 22% to 72% of people with ID take antipsychotics[36,37], and up to 13% take them at doses higher than what is recommended in medication guides[35]. In many cases, the underlying reason for these high polypharmacy rates is unrelated to the presence of a diagnosed psychiatric disorder. According to Sheehan et al.[10], 71% of people with ID using antipsychotics had no record of a diagnosed mental illness. It would therefore seem that psychotropic medication is a common treatment for psychiatric and behavioral problems in people with ID, to the detriment of other non-pharmacological approaches such as behavioral therapies or positive behavior support (PBS)[38].

The situation is all the more concerning when we consider that the evidence supporting the safety of these medications in this population is lacking because people with ID are often excluded from randomized clinical trials[39]. Given the polypharmacy that a single individual may be taking at a given time, it is impossible to ignore the potential for dependence and substance use disorder.

**Changes and transitions**

People with ID may be exposed to a greater number of stressful situations, which are directly (e.g., health problems) or indirectly (e.g., rejection and exclusion) related to their disability and impaired coping strategies. These stressors often lead to a deterioration in their emotional state, triggering emotional difficulties[40,41].

Major life events and transitions, which increase as a person ages[42], require appropriate planning that is sensitive to the person with ID and that anticipates changes in order to avoid emotional stress[43]. Yet gaps in planning occur frequently throughout their lives, whether in transition from one service to another, when they are of school age, in employment or residential settings, or at end of life. These situations contribute to a higher prevalence of depression and anxiety.

End-of-life care deserves particular attention, given the many difficulties and barriers faced by people with ID in later life stages[44]. The loss of significant others
and the (often complicated or pathological) grieving process are extremely important issues because of the contextual and emotional changes that can result from a bereavement.

Rights violations
Based on the above discussions, it will come as no surprise that people with ID are more likely to encounter stressors: these include exclusion, interpersonal loss and rejection, overprotection and infantilization, lack of family and social support, stigmatization and low expectations, and rights violations and victimization[45]. The socioeconomic disadvantage that often accompanies ID increases the risk of negative life events[46]. In this context, there is an urgent need for effective measures to facilitate the full realization of the rights enshrined in the Convention on the Rights of Persons with Disabilities (CRPD)[47] and to ensure that people with ID do not become victims of fundamental rights violations that increases their vulnerability to emotional difficulties. The quality of life conceptual framework has proven to be a useful vehicle to facilitate this task of fulfilling the specific rights highlighted in the CRPD for this collective[48-51].

EMPIRICALLY SUPPORTED PSYCHOLOGICAL TREATMENTS
Psychological treatments with strong empirical support for efficacy and efficiency in people with ID are extremely scarce. One of the reasons to explain this paucity of empirical evidence is that most people with ID who present with problem behaviors and psychiatric disorders are not receiving appropriate psychological treatments. Although there is ample evidence showing that pharmacological interventions alone are less effective than behavioral interventions, treatment is regrettably often reduced to the use of physical (e.g., exclusion), mechanical (e.g., restraints), or chemical (e.g., psychotropic) interventions.

Further, considerable interindividual differences mean that evidence of the efficacy of an intervention in people with a lower level of support needs may not be generalizable to people with more significant and complex support needs, and vice versa. Most research into psychological treatments has focused on people with lower levels of support needs.

It is extremely difficult to conduct randomized clinical trials and large-scale studies, as people with ID may have such significant support needs making it ethically questionable to assign them to a waitlist control group. Moreover, in this area, small differences may be clinically significant, even when they are not statistically significant.

Finally, we must not forget that there is also evidence pointing to the effectiveness of intrusive and aversive interventions that in effect pose a threat to people’s rights (hence introducing an ethical quandary). Considerable effort has been devoted to demonstrating that certain interventions are shown to be ineffective, as has been the case with facilitated communication[52,53].

We concur with Campbell et al[54], who warned against the oversimplification of the concept of effectiveness into “what works/doesn’t work”, and instead they recommended looking at what interventions work for whom, when, where, and how, assessing their appropriateness and identifying the factors that influence the outcomes (e.g., the quality of the therapeutic relationship). The fact remains, however, that there is a paucity of data supporting the efficacy or effectiveness of psychological treatments aimed at improving the mental health of people with ID. Empirical evidence in this area is still substantially limited[55-57] and the number of available studies are insufficient to make conclusive recommendations, although some interventions show promise, they still require testing in larger, more rigorous trials[58].

An example illustrating this dearth of evidence is a February 2021 Web of Science search for publications with “empirically supported” in the title; the search yielded 422 results, with only three including people with ID. The three articles were about the treatment of problem behavior (functional communication as a well-established treatment), phobic avoidance, and pica (behavioral intervention as a well-established treatment).

To begin to address this gap, Gómez and Navas[59] have listed some of the psychological interventions with the most empirical support for the treatment of problem behaviors or psychiatric disorders in people with ID, establishing their degree of recommendation and current level of evidence. Given the breadth of psychological treatments that could potentially be applied to the full range of problem behaviors and
psychiatric disorders in people with ID, in this paper we will use the classification based on the three generations or waves of therapies, described by Pérez-Álvarez[60]. A summary is presented in Table 1.

**First-generation therapies**
The first generation of behavioral therapy emerged as a rebellion against the prevailing clinical psychology based on psychoanalysis[60], for which the empirical evidence in people with ID was reduced to descriptive or anecdotal data that often generated more pain than progress. In opposition to these psychoanalytic approaches, the paradigms of classical and operant conditioning were advanced[60]. It was this second paradigm that eventually gave rise to behavioral therapy and applied behavior analysis (ABA), which are among the most widely used interventions and for which there exists the most empirical support for its effectiveness in the field of ID[54].

As Hartley et al[57] pointed out, interventions based on the principles of functional behavior analysis are, to date, the only ones included in published lists of effective treatments for people with ID. According to Hartley and colleagues, there is also consensus that differential reinforcement of other behaviors and differential reinforcement of incompatible behaviors increase the effectiveness of the intervention when combined with other behavioral techniques. In a review by Campbell et al[54] on the evidence-base for effective psychological interventions treating problem behavior in people with ID, functional analysis is categorized at the highest level of evidence (i.e., “A” with at least one meta-analysis, systematic review, or randomized control trial of high quality and consistency) in two of the reviewed studies. For more information about this system of classification, see Campbell et al[54].

Functional analysis is a process aimed at understanding and empirically testing the function of a problem behavior and how it relates to events in the individual’s environment (antecedents and consequences). Before determining the existence of a behavioral problem, a multidimensional assessment is needed to exclude health or psychological problems, or medication side effects that could account for the observed behavior in the individual. When these possible causes have been ruled out or treated, a functional analysis is carried out to understand the relationship between the problem behavior and its antecedents and consequences. We must not forget that problem behaviors can be an attempt to communicate a need (e.g., pain relief) or desire (e.g., food) by people who may lack the ability to effectively express themselves in other ways.

In this context, the principles of PBS emanate from the field of disability. PBS is based on respect for the person’s values, and its effectiveness is evaluated in terms of improvements to quality of life. In fact, person-centered interventions such as PBS are recommended by the National Institute for Health and Care Excellence[61]. Lavigna and Willis[62] reviewed 12 studies that used PBS involving a total of 423 subjects. While the results demonstrated the efficacy of this intervention, most investigations were single case studies or used very small non-randomized sample sizes. In the review by Campbell et al[54], studies using PBS were graded at a “B” level of evidence (i.e., well-conducted clinical studies but no randomized clinical trials directly applicable to the target population, and demonstrating overall consistency of results) and “C” level (i.e., widely held expert opinion but no available or directly applicable studies of good quality). The exception was the intervention by Hassiotis et al[63], a randomized controlled trial in which the combination of ABA-PBS together with a standard intervention was shown to be more effective than the standard treatment alone in reducing problem behaviors of lethargy and hyperactivity. McGill et al[64] demonstrated using a cluster randomized controlled trial of setting-wide PBS intervention was effective in significantly reducing problem behaviors in social care settings.

**Second-generation therapies**
The second generation evolved from the convergence of cognitive and behavioral therapy. Cognitive behavioral therapy (CBT) deconstructed the contextual and idiographic character of behavioral therapy to understand cognitions as the cause of emotional and behavioral problems, associating such cognitive “defects” with specific disorders[60].

Despite a paucity of evidence of its efficacy in people with ID, CBT has generated much interest and is often considered potentially useful, taking into account the significant methodological limitations of published studies. According to the research reviewed in a meta-analysis by Vereenooge and Langdon[65], CBT could be a potentially effective treatment for anger and depression (with better outcomes in individual over group-based interventions). Roberts and Kwan[66] found that CBT...
Table 1 Psychological interventions with empirical support for people with intellectual disability

<table>
<thead>
<tr>
<th>Generation</th>
<th>Intervention</th>
<th>Goal</th>
<th>Degree of recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>First: Behavioral therapies</td>
<td>Applied behavior analysis</td>
<td>Problem behaviors</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>Positive behavior support</td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>Second: Cognitive-behavioral</td>
<td>Cognitive-behavioral therapy</td>
<td>Anger, aggression; Depression</td>
<td>C</td>
</tr>
<tr>
<td>therapies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third: Contextual therapies</td>
<td>Mindfulness-based therapy</td>
<td>Aggression</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Mindfulness-based positive behavior support</td>
<td>Problem behaviors</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Compass-focused therapy</td>
<td>Self-criticism unfavorable social comparison; Negative</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>social comparison</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dialectical behavior therapy</td>
<td>Wellbeing</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Acceptance and commitment therapy</td>
<td>Obsessive thoughts</td>
<td>D</td>
</tr>
</tbody>
</table>

significantly reduced anxiety in adults with mild to moderate ID as measured by self, proxy, and clinician reported outcome measures.

While preliminary research exists evaluating the use of CBT for adults with ID, the potential use of CBT among children and adolescents with ID is significantly under researched[67]. To the authors’ knowledge, the only specifically adapted CBT program for children with ID is Fearless Me[67], a multimodal CBT program combining both face-to-face sessions and an online component. It is designed for children and adolescents with mild to moderate ID with the aim of helping them learn to manage and overcome anxiety.

Since cognitive therapies rely on verbal skills and the person’s ability to self-report emotions and feelings, they need to be adapted for use in people with ID. The level of adaptation will depend on the person’s level of support needs. The duration of the intervention will be longer and will require a greater number of sessions than for people without ID. National Institute for Health and Care Excellence (NICE) guidelines[68] recommended adapted CBT for the treatment of depression and anxiety disorders in people with ID, while acknowledging that the quality of evidence on its effectiveness in this population is low.

Mindfulness-based cognitive therapy is a form of CBT that merits particular attention, since it requires less dialogue with the therapist for the client to become aware of and normalize their emotions or bodily sensations, without applying unsuccessful efforts to discard those experiences. Abstract notions are simplified and made more concrete for people with ID through a series of adaptations. For example, Idusohan-Moizer et al[69] described and implemented a 10-session program to treat depression and anxiety in adults with ID. The intervention used mindfulness in breathing and basic yoga stretches characteristic of this approach, which they combined with metaphors and analogies typical of the third-generation therapies we examine in the next section. Although the results indicate a reduction in symptomaticity, they should be interpreted with some caution because of issues in determining the specific contributory variables.

**Third-generation therapies**

Originating as a protest against cognitivism, the third generation of behavioral therapy evolved from developments in behavior analysis, radical behaviorism, and an interest in Skinner’s verbal behavior. As pointed out by Pérez-Alvarez[60], contextual therapies are a response to the dissatisfaction with statistical diagnostic systems (DSM/ICD), by considering transdiagnostic dimensions (e.g., hyperreflexivity) rather than nosological categories, and by advocating therapeutic principles (e.g., activation and acceptance) rather than specific techniques or treatments. Third-generation therapies explain problems (as is the case in the field of disability) in interactive, functional, and contextual terms, not as something that is defective in the brain or person, in society or culture, but in the co-evolutionary interaction of the individual and the context in which he or she develops. Further, the effectiveness of these therapies is measured by personal achievements in different life domains (and not by reducing a list of symptoms); hence they align well with the current understanding of
interventions for people with ID, where success is measured in terms of supports and quality of life[30,70].

Patterson et al[71] highlighted several important findings in their recent review of third-generation therapies for adults with ID. First, only one of the 20 studies that met the inclusion criteria was a randomized controlled trial; the others were single case studies or single group designs with pre- and post-test evaluation. In all, 45% of the included studies used mindfulness-based therapy (MBT), 30% dialectical behavior therapy (DBT), 15% compassion-focused therapy (CFT), and 10% acceptance and commitment therapy (ACT). Adaptations included modifying the language, concretizing abstract concepts, chunking information, using physical and visual prompts, providing additional time to process information, checking whether participants understood, using role play and experiential exercises, reducing the duration of sessions, and involving carers. On average, interventions were conducted over a 35-wk period and included eight sessions, with each session lasting between 30 min (individual therapy) to 120 min (group therapy). Third-wave therapies improved some symptoms of psychiatric disorders for some adults with ID, but were highly effective at reducing problem and offending behavior in the community, enabling participants to maintain “at-risk” placements.

Singh et al[72] tested the effectiveness of mindfulness-based PBS, a customized form of MBT that enables parents and other caregivers to reduce their reported psychological stress to normative levels through mindfulness procedures and to support children with ID to self-manage their problem behaviors through PBS, reporting significant reductions in aggression and disruptive behavior and increases in compliance behaviors. Meditation on the Soles of the Feet is one of the few yet most cited resources to have been specifically developed for the treatment of anger and aggression in people with ID[73]. It involves the person shifting their attention from an emotionally intense situation to an emotionally neutral part of the body, the soles of their feet. Other techniques used are “mindful observation of thoughts” (i.e., focusing on breathing, visualizing, and observing thoughts like clouds passing through the consciousness) and observation of breathing or focusing on noises and objects in the environment. While the duration and form of mindfulness training varies across studies, “soles of the feet” training generally includes intensive weekly or daily sessions of supervised role-playing, practice, and homework. Newer promising mindfulness-based programs in the self-management of aggression are Surfing the Urge[72] and Stop, Observe, Breathe, Expand, and Respond Breathing Space[74]. Despite the limited and weak evidence, existing studies suggest that mindfulness-based programs are effective when applied individually, in groups, or in the community, as well as when parents or support persons are trained in their use, especially for the treatment of aggressive behavior[75-77]. When working with people with ID, however, it is more advisable to apply the programs individually[78].

DBT, which is aimed primarily at reducing behaviors such as self-harm, also has applications and adaptations in the field of ID. The review conducted by McNair et al[79] described seven studies on DBT adaptations for use with individuals with ID, but the quality of evidence of its efficacy was rated as methodologically “weak”; the reasons cited included the small number of participants, the absence of a comparison group, and the lack of specificity regarding the goals of the therapy. Glicksman et al[80] proposed using this therapy in conjunction with two core models from the field of ID: the person-centered model and the rights-based model. Hewitt et al[81] observed that improvements in scores on measures of psychological well-being, psychological distress, and quality of life appeared to be maintained at 6-mo follow-up, but scores at 2-year follow-up suggested a return to baseline for three participants who attended a DBT group for people with ID. In another investigation, Pearson et al[82] interviewed 11 individuals with ID about their experiences with this therapy, and noted that the empowering and validating environment created within DBT appeared to facilitate them having a voice in sessions, as individuals described asserting their needs, exerting control, and showing agency around their learning. They also experienced a positive therapeutic relationship, where the relationship was characterized by trust, validation, attunement, availability, empowerment, and a nonjudgmental stance. Moreover, participants described using skillful behaviors and generalizing these skills to other settings.

CFT is another approach that is gaining traction in the field of ID. The adaptation of this therapy for use in people with ID, proposed by Clapton et al[83], consisted of presenting the psychoeducational material in a more concrete and visual manner, avoiding or minimizing the use of abstract language, producing a support manual with simple summaries (written and visual information) for each session’s specific purpose, and designing practices and exercises to be done at home (e.g., use of colored
sand and water in a bottle to demonstrate the purpose of calming mind and body). The program, consisting of six sessions of approximately 90 min each, was rated as safe, accessible, and useful by the individuals with ID. The main difficulty with this intervention reported by participants was similar to what tends to be reported by participants from the general population: the initial challenge of being able to receive and generate compassion for oneself. The evidence of the efficacy of CFT, however, is based only on one study with a pre-post evaluation of six people (in which a significant decrease in the level of self-criticism and unfavorable social comparison was observed); and another by Goad and Parker[84] involving two groups of five and six people (in which clinically — but not statistically — significant differences were found).

Fewer studies have been published on treating this population with ACT, a behavioral-contextual approach that combines behavioral change techniques, present moment awareness, and acceptance procedures. A recent review by Byrne and O’Mahony[85], that included eight studies that used ACT with individuals with ID and/or ASD, concluded that ACT may be a useful intervention in reducing psychological difficulties and improving adaptive functioning. However, the small number of individuals included in the sample, as well as the poor methodological rigor of these studies, limit conclusive generalization of these findings. Brown and Hooper[86] outlined how the ACT model was adapted to treat Sarah, 18 years old, who had moderate/severe support needs and was experiencing depressive and anxious thoughts. Similarly, Pankey and Hayes[87] applied a four-session ACT intervention with a young woman with ID (intermittent support needs) and psychosis, and obtained limited but encouraging results.

In sum, it should be reiterated that all of these studies on the use of third-generation therapies in people with ID are quite preliminary and have major limitations (small sample sizes, absence of a control or comparison groups, insufficient evidence of the validity of the assessment instruments used); at best, we can conclude that these treatments show promise, but more effectiveness research is needed.

CONCLUSION

The significant conceptual changes that are taking place in the field of ID have important implications for the work of health professionals and their role in diagnosis, classification, and supports planning. For appropriate and accurate decisions to be made, Schalock et al.[6,88] recommended: (1) Ensuring that the required criteria for a diagnosis of ID are met; (2) Understanding related constructs and using precise terminology; (3) Recognizing the different conceptions of disability and perspectives that explain ID; (4) Using evidence-based practices in the design and provision of systems of supports[89]; and (5) Approaching subgroup classification as a post-diagnostic strategy, with a specific purpose and based on an explicit classification framework (e.g., according to support needs).

It is also important to bear in mind that, despite the occurrence of diagnostic overshadowing, people with ID have not escaped the omnipresent power of the marketing of psychotropic medications. The medicalization of the lives of people with ID has been promoted at the expense of psychotherapies that, although requiring more effort, have proven efficacy in people without disabilities. We must, therefore, exercise extreme caution to avoid pathologizing people’s everyday lives.

In terms of empirically supported psychological treatments, there is an urgent need to improve preventive measures and facilitate access to appropriate, efficacious, and ethical psychological treatments for people with disabilities. To this end, additional and more robust research in this area should be considered a priority, especially in light of the remarkable paucity of reliable empirical evidence for this population. The failure to meet the needs of people with ID and problem behaviors or psychiatric disorders is a clear violation of their rights[19].

Without wishing to undermine the positive examples of second-generation CBTs being adapted for use with people with ID, we believe it would be contradictory to recommend their use given the need to move away from deficit-based models in order to understand ID. Because of their contextual approach, in contrast to the medical model, first- and third-generation therapies are, in our view, more suitable. For third-wave therapies in particular, there is an urgent need for research into outcome measures with evidence of validity in homogeneous and adequately sized study groups; as well as evidence of cost and efficacy. In this regard, the evaluation of personal outcomes related to quality of life is highly recommended as a way of testing...
the efficacy of psychological treatments[90].

In accordance with the NICE guidelines[68], psychological interventions should be tailored to the preferences of people with ID, their level of understanding, their preferred mode of communication, their strengths, needs, and other associated conditions. In addition, interventions should respect the individual’s privacy (although they should also seek the collaboration of family and primary support providers) and provide additional supports during and between sessions. To achieve this aim, it is not enough for health care professionals — particularly psychologists and psychiatrists — to know what is ID. They must also learn about the relevant ethical aspects, principles, and values, including supports, inclusion, rights, self-determination, and quality of life[50]; and be willing to coordinate with professionals from other social, education, and health services.

REFERENCES


26 Verdugo MA, Navas P. All means all. Rights and quality of life of people with intellectual disability and major support needs. Madrid: Real Patronato sobre Discapacidad, 2017

27 Navas P, Verdugo MÁ, Martínez S, Sainz F, Aza A. Derechos y calidad de vida en personas con discapacidad intelectual y mayores necesidades de apoyo [Rights and quality of life of individuals with intellectual disability and extensive support needs]. *Siglo Cero* 2017; 48: 7-66


Foster CA. Dej-a Vu: Documentary revisits facilitated communication pseudoscience. *Behav Intervent* 2019; 34: 577-586 [DOI: 10.1002/bin.1687]


Pérez-Álvarez M. Las terapias de tercera generación como terapias contextuales [Third-generation therapies as contextual therapies]. Madrid: Síntesis, 2014


Hronis A, Roberts R, Roberts L, Kneebone II. Fearless me!©: cognitive behaviour therapy for children with intellectual disability and anxiety. Sydney: University of Technology Sydney & University of Adelaide, 2018


Verdugo MA, Schalock RL, Gomez LE. El Modelo de Calidad de Vida y Apoyos: La unión tras veinticinco años de caminos paralelos [The Quality of Life Supports Model: Twenty-five years of parallel paths have come together]. Siglo Cero 2021; 52: 9-28 [DOI: 10.14201/sig.2021523928]


Calderón R, Caballo C. Revisión sistemática sobre la aplicación y efectividad de tratamientos basados en mindfulness en personas con discapacidad intelectual [Systematic review on the application and effectiveness of treatments based on mindfulness in people with intellectual disabilities]. Rev Española Discapac 2018; 6: 49-74 [DOI: 10.5569/2340-5104.06.01.03]


https://www.wjgnet.com
Gómez LE et al. Psychiatric and behavioral disorders in ID


87 Pankey J, Hayes SC. Acceptance and commitment therapy for psychosis. *Int J Psychol Therapy* 2003; **3**:311-328


Phantom bite syndrome: Revelation from clinically focused review

Trang Thi Huyen Tu, Motoko Watanabe, Gayatri Krishnakumar Nayanar, Yojiro Umezaki, Haruhiko Motomura, Yusuke Sato, Akira Toyofuku

ORCID number: Trang Thi Huyen Tu 0000-0002-6062-4603; Motoko Watanabe 0000-0002-0767-197X; Gayatri Krishnakumar Nayanar 0000-0001-7722-9741; Yojiro Umezaki 0000-0003-0249-1977; Haruhiko Motomura 0000-0001-7923-2477; Yusuke Sato 0000-0003-0856-9173; Akira Toyofuku 0000-0002-9498-3014.

Author contributions: Tu TTH, Watanabe M, Umezaki Y and Toyofuku A were involved in manuscript drafting; Nayanar GK, Motomura H and Sato Y critically revised the manuscript; All authors read and approved the manuscript.

Conflict-of-interest statement: The authors declare that they have no competing interests.

Supported by: JSPS KAKENHI Grant, No. 19K10328 (to Professor Toyofuku A).

Country/Territory of origin: Japan

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Peer-review report’s scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): 0
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

Trang Thi Huyen Tu, Department of Basic Dental Sciences, Faculty of Odonto-Stomatology, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh 72714, Viet Nam

Trang Thi Huyen Tu, Motoko Watanabe, Gayatri Krishnakumar Nayanar, Akira Toyofuku, Department of Psychosomatic Dentistry, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

Yojiro Umezaki, Department of General Dentistry, Fukuoka Dental College, Fukuoka 814-0193, Japan

Haruhiko Motomura, Department of Psychosomatic Dentistry, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

Yusuke Sato, Gerodontology and Oral Rehabilitation, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan

Corresponding author: Trang Thi Huyen Tu, DDS, PhD, Lecturer, Research Fellow, Department of Basic Dental Sciences, Faculty of Odonto-Stomatology, University of Medicine and Pharmacy at Ho Chi Minh City, 652 Nguyen Trai, District 5, Ho Chi Minh 72714, Viet Nam. tu.ompm@tmd.ac.jp

Abstract

Phantom bite syndrome (PBS), also called occlusal dysesthesia, is characterized by persistent non-verifiable occlusal discrepancies. Such erroneous and unshakable belief of a “wrong bite” might impel patients to visit multiple dental clinics to meet their requirements to their satisfaction. Subsequently, it takes a toll on their quality of life causing, career disruption, financial loss and suicidal thoughts. In general, patients with PBS are quite rare but distinguishable if ever encountered. Since Marbach reported the first two cases in 1976, there have been dozens of published cases regarding this phenomenon, but only a few original studies were conducted. Despite the lack of official classification and guidelines, many authors agreed on the existence of a PBS “consistent pattern” that clinicians should be made aware. Nevertheless, the treatment approach has been solely based on incomplete knowledge of etiology, in which none of the proposed theories are fully explained in all the available cases. In this review, we have discussed the critical role of enhancing dental professionals’ awareness of this phenomenon and suggested a comprehensive approach for PBS, provided by a multidisciplinary team of dentists, psychiatrists and exclusive psychotherapists.
INTRODUCTION

A general dentist might come across many patients with disturbing or uncomfortable bite sensations in their daily practice. Most of the time, these typical complaints are due to reasons such as new crown delivery, dental restorations, orthodontic treatment and temporal mandibular joint dysfunction. Subsequently, dentists will perform some occlusal adjustments or provide a specific therapy, if on examination they observed any abnormal intercuspation or contact patterns. However, apart from these ordinary cases, there are some cases where patients present with non-verifiable occlusal discrepancies with strict demands for bite correction.

This phenomenon, named “phantom bite syndrome” (PBS), was firstly described by Marbach in 1976 with a report of two female cases[1,2]. The term was originally inspired by phantom limb pain, because it was suggested that what occurred after dental treatment in patients with hypochondriacal or severe personality disorder resembles the “ego defense” or “denial of loss” mechanism in postamputation patients[2]. Earlier, Posselt had mentioned this unusual sensation as a hyperawareness named “positive occlusal sense” in his textbook from 1960, saying “after occlusal grinding or adjustment, some person with a nervous predisposition may become too conscious of their own occlusion”[3]. However, Marbach’s evocative illustrations are useful for clinicians to visualize better the phenomenon.

In 1997, Clark et al[4] proposed an alternative term: “Occlusal dysesthesia”. Nowadays, this is commonly used to define “a persistent (more than 6 mo) uncomfortable bite sensation, which does not correspond to any physical alteration related to occlusion, pulp, periodontium, muscle or temporomandibular joint and cause significant functional impairment”[5]. Since then, there have been dozens of published cases regarding this phenomenon, but few original studies have been conducted. From the compiled case reports, it was observed that PBS patients were never satisfied with occlusal treatment resulting in “dental shopping”, and they refused to see psychiatrists because of their belief in “ideal bite correction” (Figure 1). Besides, neither an official classification nor any widely accepted treatment protocol currently exists. Given the new shreds of evidence from recent research, we set out to discuss: (1) Demographic characteristics and the typical manifestations of PBS; (2) The debate on etiopathogenesis; and (3) Recommended management strategies.
Tu TTH et al. A review of phantom bite syndrome

Figure 1 A typical course of phantom bite syndrome. Those nomadic patients usually travel from one dentist to another, desperately seeking an “ideal bite correction”. They normally refuse to see psychiatrists if being referred.

DEMOGRAPHIC AND THE TYPICAL MANIFESTATIONS OF PBS

Epidemiology

Generally in dentistry, uncomfortable occlusal sensations are a common finding among patients, while PBS (which is distinguishable if ever encountered) is quite rare. According to an email survey sent to United States orthodontists, 75% of respondents recalled encountering at least one patient with typical symptoms of phantom bite during their career, even though almost half of them were unfamiliar with the term itself[6]. Gerstner et al[7] in their study found that 20.5% of 127 temporal mandibular disorder (TMD) clinic patients had their uncomfortable bite all the time. However, whether their “discomfort” met the PBS diagnosis criteria was not discussed. In a study by Watanabe et al[8], PBS only accounts for less than 10% of outpatients who visit a specialized clinic of oral psychosomatic disorders. The discrepancy of the ratio between cases encountered by United States orthodontists and the cases in specialized clinic might be attributed to their obsession for “ideal bite correction” as mentioned above. Yet, to our knowledge, there has been no study that estimates the incidence of this phenomenon in the general population.

Demographic characteristics

Back in 2012, there were no original PBS data published. Hara et al[9] were the first to combine the results from 37 case reports in their first systemic review. Later, in two retrospective studies done by two different Japanese teams, the data confirmed similar patterns of female predominance, mean age and symptom duration. In particular, the female ratio varies from 72% to 84%, while the mean age at the initial visit ranges from 51.7 years to 53.1 years[8,10]. In terms of symptom duration, 39.5% of patients suffered from an abnormal bite for more than 5 years[8]. Except for the 2 cases that started since adolescence (as both described by Marbach), the majority developed their onset symptom around the age of 45[2,9].

One frequent observation among many PBS cases is that the first mild discomfort is often associated with some certain dental treatment (e.g., restorations, orthodontic treatment), then it becomes worse after further occlusal adjustment or extensive dental interventions (e.g., replacing crowns, extraction)[2,8,11-17]. However, 26.2% of 130 PBS patients reported symptoms that developed spontaneously or with triggers other than dental therapy. Looking for an ideal bite correction, they then visited approximately 4.4 ± 3.4 dental clinics[6]. The highest record belongs to one male patient who attended at least 200 appointments with 20 different dentists in 6 years[11]. To be able to pursue such a frustrating, time and money-consuming journey, the PBS patients are normally
assumed to be of moderate to high socioeconomic status\cite{11,15,18}. Such impression comes from expert consensus, but in fact, the situation may vary depending on national medical systems and policy.

**Comorbid psychiatric disorders**

In the very first announced case, Marbach suspected that PBS patients were mentally ill with delusion and paranoia, saying “hope for these unfortunate patients lies in part in the ability to make psychiatric research available for dentists”\cite{2}. This argument was questioned later by Greene and Gelb, when 4 out of 5 patients in their report did not qualify for any diagnosis of mental disorders\cite{19}. From a Japanese prosthodontist’s report, 46.2% of the patients complaining of occlusal disharmony had neuroticism, and 53.8% had manifest anxiety\cite{20}. This aligned with recent reports whose prevalence of psychiatric comorbidity (depression, anxiety disorder, insomnia, somatic symptom disorders) was between 45.9% and 59.51%. Schizophrenia and severe personality disorders were rare\cite{8,10}. Since PBS patients mainly complain of occlusal discomfort and rarely present severe psychiatric comorbidities, they seldom are provided active psychiatric treatment.

**Typical manifestations of PBS**

Despite the lack of official classification and guidelines, many authors agreed on the existence of a PBS “consistent pattern” that clinicians could easily recognize and should be made aware\cite{9,18,21,22}. A list of frequently observed clinical manifestations is summarized in Table 1.

Firstly, the occlusion or bite would be the center of their complaints, even though they may be expressed in many different ways (see Table 2). This would make PBS distinguishable from other oral conditions characterized by abnormal sensations or idiopathic pain without evident causes such as burning mouth syndrome, oral cnenestopathy and atypical odontalgia. Besides, PBS could be observed together with TMD and is sometimes even categorized as TMD’s subgroup\cite{5,23-25}.

Another important clinical aspect rarely mentioned in the literature is patients’ emphasis that their occlusal problems lead to concomitant somatic symptoms in other body parts (e.g., idiopathic headache, musculoskeletal pain)\cite{8,11,12,16,22,26}. “Our teeth is not separated from the body, after all”, one patient even said\cite{2}. Therefore, they firmly believe that all of their somatic dysfunctions would be cured if and only if their bites are corrected (Figure 2).

There has been no official record of specific triggers other than dental interventions at symptom onset. However, some cases reported patients who experienced a traumatic accident or underwent a pressured period of life (e.g., divorce, change of jobs), with or without dental treatments\cite{8,12}. Lack of a dental trigger becomes a predictor for psychiatric comorbidity, which affects the treatment outcome\cite{8}.

When patients describe their symptoms, they tend to use dental professional terminologies, even when they only have superficial knowledge about their conditions\cite{2,11,13}. Not only equipping themselves with a lot of arguments and self-research information, they often bring to the appointment pieces of evidence to prove occlusal dysfunctions, including their collection of diagnostic casts, occlusal splints, teeth pictures, radiographs or even extremely detailed resume and records from previously failed treatments\cite{2,6,11,18,26,27}. They would, sometimes, describe clearly and confidently that prior incompetent dentists are responsible for their exacerbated symptoms\cite{2,6,11,13,15,28}.

From our clinical observation, some patients are more obsessed with the idea of getting their occlusal equilibrium done than the “wrong bite” itself. In many cases, PBS patients even rigorously direct the dentists on what to be done. If not granted desired treatments, these patients would reject any other suggested treatment and quickly drop out after the first or second visit. Once the dentists recognize this pattern, inform patients about their normal examination results and gently recommend another specialist/psychiatric assessment, this rational approach will be met with prolonged discussions and denial. Meanwhile, even if patients’ demands are met, the absence of any tangible result will reinforce the existing erroneous belief that the occlusion problem has not been properly addressed. Dental services often affect PBS patients iatrogenically for worse (Figure 3). Such a vicious cycle of dental shopping thus continues.

In terms of psychological impacts, a study by Tsukiyama et al\cite{29} showed significantly higher scores of somatic symptoms and depression subscales in PBS patients in comparison with those of control groups. Nevertheless, as the author self-declared, these differences “only indicate that the patients may have psychiatric problems, not possible to prove that they have mental disorders”. Even in PBS cases
Table 1 Clinical characteristics of phantom bite syndrome

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Preoccupation with their dental occlusion and an enormous belief that their dental occlusion was abnormal</td>
</tr>
<tr>
<td>2  A long history of repeated dental surgery treatment failures with persistent requests for the occlusal treatment that they are convinced they need</td>
</tr>
<tr>
<td>3  A relatively high intelligence and socioeconomic status enabled them to undergo endless costly and time consuming dental treatments</td>
</tr>
<tr>
<td>4  Despite repeated failures of dental surgery, persist in seeking bite correction from a succession of dentists</td>
</tr>
<tr>
<td>5  A strong resistance to referral to psychiatrists and stick to dental procedures</td>
</tr>
<tr>
<td>6  A favorable attitude to dentists at first, gradually blaming them for the exacerbated symptoms, finally dropping out with disappointment</td>
</tr>
<tr>
<td>7  A tendency to use dental jargon</td>
</tr>
<tr>
<td>8  Bringing to the appointment pieces of evidence to prove occlusal discrepancies (radiographs, study cast, temporary crowns, mouthpieces, etc.)</td>
</tr>
</tbody>
</table>

Table 2 Summary of frequent complaints observed in patients with phantom bite syndrome and proposed terminologies

<table>
<thead>
<tr>
<th>Terminologies</th>
<th>Frequent complaints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phantom bite syndrome</td>
<td>Abnormal/uncomfortable bite</td>
</tr>
<tr>
<td>Occlusal dysesthesia</td>
<td>My bite is off/too high</td>
</tr>
<tr>
<td>Occlusal hyperawareness</td>
<td>My jaws are not biting correctly</td>
</tr>
<tr>
<td>Occlusal hypervigilance</td>
<td>Jaw looseness and weak bite</td>
</tr>
<tr>
<td>Occlusal neurosis</td>
<td>Uneven dental bite</td>
</tr>
<tr>
<td>Positive occlusal sense</td>
<td>Feel uneasy with the bite</td>
</tr>
<tr>
<td>Persistent uncomfortable occlusion</td>
<td>I try maneuver to position the bite correctly</td>
</tr>
<tr>
<td></td>
<td>I don’t know where my teeth belong anymore</td>
</tr>
<tr>
<td></td>
<td>Lack of familiarity with my own bite</td>
</tr>
</tbody>
</table>

Without psychiatric comorbidity, psychological distress is remarkable. They might lead to serious consequences on patients’ life quality, relationship with family, financial loss, career disruption or even suicidal thoughts[2,11,12,15]. Corresponding dentists, if trapped in these unusual cases, will quickly find these patients oppose any treatment and become increasingly challenging to manage. The worst scenarios would be litigation problems between patients and dentists[2,11,13].

THE DEBATE ON ETIOPATHOGENESIS

Initially, PBS was viewed as a psychotic disorder that was “rarely brought to the attention of psychiatrists” before being classified into monosymptomatic hypochondriacal psychosis (MHP)[2,18]. This speculation arises from the similarity between “wrong bite” obsession in PBS patients and “an erroneous and unshakable belief in a distorted body image” in MHP phenomenon[18]. In other words, PBS was suggested to be a sub-phenotype of MHP present in dental clinics, comparable with parasitosis that is often seen by dermatologists, sharing the common features of equal gender...
Figure 2 A written self-report of patient's multiple comorbidities of medically unexplained symptoms (e.g., headache, dizziness, neck and back pain, ambulatory impairment, numbness of hands and legs) associated with occlusal symptoms.

Figure 3 Occlusal view of mandibular arch in a 60-yr-old female phantom bite syndrome patient. Clinical examination revealed unnatural restorations for dentition owing to the fullest effort of dentists.

distribution and early adulthood onset. Unfortunately, this observation seems no longer to concur with recent demographic reports[8-10].

Twenty years later, this "psychodynamically oriented view" is replaced by the same author[30]. At this time, Marbach[30] adapted Melzack’s theory of neuromatrix and re-discussed PBS’s pathophysiology in a shared context with phantom tooth pain/atypical odontalgia[30,31]. The key element of this theory is that there exist individual differences in self-knowledge of occlusion, namely occlusal neurosignature.
Tu TTH et al. A review of phantom bite syndrome

Figure 4 Intra-oral photography and regional cerebral blood flow increasing map before and after phantom bite syndrome psychopharmacotherapy. The color bar indicates the Z-score comparing normal controls. A and B: The before (A) and after (B) pharmacotherapy. As the phantom bite symptom decreased, the asymmetrical regional cerebral blood flow pattern attenuated, and dental treatment was finally completed. Citation: Umezaki Y, Watanabe M, Takenoshita M, Yoshikawa T, Sakuma T, Sako E, Katagiri A, Sato Y, Toyofuku A. A case of phantom bite syndrome ameliorated with the attenuation of the asymmetrical pattern of regional cerebral blood flow. Jpn J Psychosom Dent 2013; 28: 30-34. Copyright ©The Japanese Society of Psychosomatic Dentistry 2013. Published by The Japanese Society of Psychosomatic Dentistry.

Such a unique coherent unit in the brain was created and is influenced by lifetime intercuspation and other tooth contacts. Whenever a dental intervention or routine adjustment is made, it would send a new input to the central nervous system. In the case of PBS patients, it is difficult for their neuromatrix to adapt to even a minor change, and they thus soon become unable to recognize the original bite itself.

From 1993 to 2000, Toyofuku conducted a clinical study using psychosomatic approaches to treat 16 serious PBS cases during hospitalization. As a result, it was observed that 15 of 16 PBS patients responded to the combination therapy of tricyclic antidepressants and supportive psychotherapies. From the result of these clinical observations, the author hypothesized that PBS might be due to several biochemical disorders involving neurotransmitters in the brain, the wrong connection between occlusion and medically unexplained complaints due to cognitive processes in the higher centers of the brain. This working hypothesis, however, had not been recognized widely due to the language barrier of Japanese publications. Notably, in the follow-up study 5 years later, about one-third of these cases began to complain of request for needless dental treatments again. Besides, a review of 130 PBS patients suggested that PBS is seldom associated with psychotic disorders. Central neuromodulator (antidepressant or antipsychotic) therapy may be effective for PBS. Most of these medications were given at very low "non-psychiatric" doses. These findings support the working hypothesis, suggesting the role of biochemical disorders involving neurotransmitters in the brain of PBS patients.

In 2003, Clark and Simmon proposed the theory of altered oral kinesthetic ability as another possible mechanism of PBS. In their speculation, some dysfunctions of muscle spindles in the jaw closers muscles would be responsible for the impairment of an individuals' ability in mandibular position discrimination. They did not invalidate Marbach's theory of diagnosable psychiatric disorders but rather agreed with Green and Gelb, stating that although patients' symptom and behaviors have certain psychological impact, the main underlying cause would be the unknown alterations in proprioceptive input transmission. Contrary to their expectation, the next two experimental studies comparing sensory perceptive and interdental thickness discriminative capacities in PBS and the control group both revealed insignificant results. Nevertheless, the possibility that the sensory test was not sensitive and accurate enough to tell the threshold differences could not be excluded.

Given this unsettled controversy, "brain imaging techniques...can be utilized to evaluate whether cortical map representations in fact correspond to patient's subjective occlusal complaints...", Hara et al suggested in their review. In 2013,
Tu TTH et al. A review of phantom bite syndrome

Umezaki et al[16] conducted a single-photon emission computed tomography in a PBS patient, revealing asymmetrical cerebral blood flow (CBF) patterns in the frontal lobe region. Interestingly, 1 year after taking psychopharmacology, not only had the patient’s symptoms remarkably improved, but the asymmetry patterns also attenuated. This finding reinforced the “altered central processing” theory, suggesting the involvement of central nervous system dysfunction in PBS manifestation. However, in a later case-control study conducted by the same research group, regional CBF (rCBF) in 44 PBS patients and 12 control subjects had insignificant differences. The author admitted that the initial idea of comparing one whole group of PBS with normal controls was “inappropriate” and interpreted this negative results as a reflection of the heterogeneous nature in PBS. A secondary analysis of these data later revealed different rCBF patterns are in accordance with certain clinical patterns, such as laterality of the symptoms or behavior of blaming dentists. In particular, PBS patients with right-side symptoms have significant right-side predominant parietal asymmetry and left-side predominant thalamus asymmetry[28]. Disturbance in the parietal area, which includes the secondary sensory cortex, and thalamus that relay information between different subcortical areas and the cerebral cortex, might imply the complexity of PBS symptoms. In the study, the tendency of frontal lobe asymmetry is also reported as the same with experimentally reproduced occlusal discomfort.

Generally speaking, all of these above interpretations were substantially built upon personal judgment and limited clinical observations. So far, none of these proposed theories fully explain all the available cases. For example, there are PBS cases (where patients had neither psychiatric disorder nor abnormal psychological condition(s) that cannot be explained using the theory of psychopathological influences alone[8,10]. Besides, there are also PBS patients whose symptoms spontaneously developed, meaning there are no dental treatments related inputs to trigger peripheral alterations. As long as this controversy remains unsolved, neither specific diagnostic testing nor effective treatment can be sufficiently developed.

RECOMMENDED MANAGEMENT STRATEGIES

Based on these perspectives of etiology and pathogenesis, various authors proposed different strategies. However, they all agree that treatment should primarily focus on patients’ education and therapies that improve overall function and well-being[9,23,27,35]. In this review, apart from those treatments, we want to discuss further other underestimated perspectives; including professional education, psychopharmacotherapy, successful guidance and reliable therapeutic relation.

Professional education

First and foremost, dental treatment would not be helpful and should be avoided. As many authors stated, PBS patients are considered to be “refractory to any dental treatment”[2,6,9,11,14,18,22]. However, they will always look for help from dentists, whom they believed to be the only people with enough expertise to understand their complaints and then be able to provide a “full bite correction”. In fact, on oral examination, some occlusal discrepancies may be detected, but they were far from the root cause of patients’ suffering[5,22]. Besides, a normally good occlusion can always be enhanced to become an ideal one with dentists’ intervention[23]. Such conventional treatment might initially relieve symptoms, but sooner or later, the condition only becomes worst since the patients’ occlusion was more and more distorted from the original. Hence, to prevent inappropriate, time-consuming, irreversible, extensive treatments; enhancing dental professionals’ awareness of this phenomenon is critical. Clinicians should be aware that there is no strong evidence to support that theoretically ideal occlusion must be fulfilled for a successful outcome of prosthodontic treatment[36].

Interestingly, we observed that the majority of PBS original research, including retrospective and case-control studies, came from either Japanese or German research teams[8,10,16,21,29,34,37]. This suggests that there are licensed specialists who treat the syndrome at specialized clinics in these two countries. In particular, thanks to the inclusion of PBS and other phenomena of oral psychosomatic disorders in the undergraduate dental curriculum of some Japanese universities since the early 2000s, a general dentist will be able to notice an early case of PBS and refer them to relevant treatment centers.
Table 3 Summary of medications used in phantom bite syndrome's management

<table>
<thead>
<tr>
<th>Classification</th>
<th>Drug’s name</th>
<th>Period of follow-up</th>
<th>Side effects</th>
<th>Treatment outcome</th>
<th>Mechanism</th>
<th>Level of evidences</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2 blocker</td>
<td>Pimozide/Haloperidol</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td>Prescribed as a treatment for monosymptomatic hypochondriacal psychosis</td>
<td>Expert’s opinion</td>
<td>Marbach [2], 1978</td>
</tr>
<tr>
<td>D2 partial agonist</td>
<td>Aripiprazole</td>
<td>Average 59 d from initial administration to clinical improved day</td>
<td>Drowsiness, constipation, weight gain, nausea, diarrhea, staggering, dizziness, malaise, irritation, headache</td>
<td>37% improved; 40.7% no change, 22.3% discontinued</td>
<td>Unspecified</td>
<td>Retrospective study, n = 27</td>
<td>Watanabe et al[8], 2015</td>
</tr>
<tr>
<td>Anticonvulsant</td>
<td>Clonazepam</td>
<td>No report</td>
<td>No report</td>
<td>No report</td>
<td>Reduce anxiety and increase tolerance to the symptom</td>
<td>Expert’s opinion</td>
<td>Clark et al [23], 2005</td>
</tr>
<tr>
<td>Tricyclic antidepressant (TCA)</td>
<td>Dothiepin</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Generally recovered</td>
<td>Prescribed as a treatment for somatic symptom disorder</td>
<td>Single case report</td>
<td>Wong and Tsang[12], 1991</td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>390 d</td>
<td>No</td>
<td>Significant improvement</td>
<td>Unspecified</td>
<td>Single case report</td>
<td>Watanabe et al[4][8], 2015</td>
<td></td>
</tr>
<tr>
<td>Paroxetine</td>
<td>No report</td>
<td>Drowsiness</td>
<td>1/3 improved; 2/3 no change</td>
<td>Unspecified</td>
<td>Retrospective study, n = 3</td>
<td>Watanabe et al[8], 2015</td>
<td></td>
</tr>
<tr>
<td>Serotonin-norepinephrine reuptake inhibitor</td>
<td>Duloxetine</td>
<td>Average 28 d from initial administration to clinical improved day</td>
<td>Drowsiness, constipation, nausea, decreased appetite</td>
<td>3/7 improved; 4/7 no change</td>
<td>Unspecified</td>
<td>Retrospective study, n = 7</td>
<td>Watanabe et al[8], 2015</td>
</tr>
<tr>
<td>Escitalopram</td>
<td>Average 18 d from initial administration to clinical improved day</td>
<td>Drowsiness, staggering, dizziness, malaise</td>
<td>3/4 improved; 1/4 discontinued</td>
<td>Unspecified</td>
<td>Retrospective study, n = 4</td>
<td>Watanabe et al[8], 2015</td>
<td></td>
</tr>
<tr>
<td>Selective serotonin reuptake inhibitor</td>
<td>Sertraline</td>
<td>Average 79 d from initial administration to clinical improved day</td>
<td>Drowsiness, constipation, nausea, edema, dry mouth, decreased appetite</td>
<td>7/9 improved; 2/9 no change</td>
<td>Unspecified</td>
<td>Retrospective study, n = 7</td>
<td>Watanabe et al[8], 2015</td>
</tr>
<tr>
<td>Fluvoxamine</td>
<td>Average 24 d from initial administration to clinical improved day</td>
<td>Drowsiness</td>
<td>2/4 improved; 2/4 no change</td>
<td>Unspecified</td>
<td>Retrospective study, n = 4</td>
<td>Watanabe et al[8], 2015</td>
<td></td>
</tr>
<tr>
<td>Noradrenergic and specific serotonergic antidepressant</td>
<td>Mirtazapine</td>
<td>Average 59 d from initial administration to clinical improved day</td>
<td>Drowsiness, constipation, weight gain, nausea, staggering</td>
<td>42.9% improved; 47.6% no change, 9.5% discontinued</td>
<td>Unspecified</td>
<td>Retrospective study, n = 21</td>
<td>Watanabe et al[8], 2015</td>
</tr>
<tr>
<td>Combination of TCA and D2 partial agonist</td>
<td>Amitriptyline; Aripiprazole</td>
<td>41 mo</td>
<td>Staggering</td>
<td>Remarkable improve</td>
<td>Altered biochemical abnormalities related to neurotransmitter and higher brain</td>
<td>Single case report</td>
<td>Umezaki et al[16], 2013</td>
</tr>
</tbody>
</table>
Tu TTH et al. A review of phantom bite syndrome

**Psychopharmacotherapy**

In terms of pharmacotherapeutics, although it has never been considered the primary choice of treatment in PBS management, medications appear to be the most applied treatment among clinical studies\[8,10,13,18,38-40\]. As presented in Table 3, the most frequently prescribed medications are antidepressants and antipsychotics. Originally, Marbach\[2\] prescribed pimozide and haloperidol, and Wong and Tsang\[12\] prescribed dothiepin, an antidepressant, since they regard their PBS patient as psychiatric disorder. As the etiological discussion matured, considerations for psychopharmacological mechanisms have been deepened. Other authors also applied psychotropic drugs and suggested such effects were related to biochemical alteration in central nervous system (especially the dopaminergic system)\[8,16\]. Besides, Clark et al\[23\] recommended clonazepam, an anticonvulsant, mainly for mood stabilizing, anxiety control and patients’ distraction. Since results were limited to single case reports, case series and a retrospective analysis study, prospective follow-up or clinical control study will be needed for further verification. At the same time, elucidation of psychotherapeutic mechanism for PBS should be required to be applied at scale.

**Successful guidance and reliable therapeutic relation**

The assessment of psychological components and the use of appropriate consultants if needed have been recommended\[6,9,18,21\]. Because dental professionals are not trained to practice psychological evaluation (either distraction or cognitive behavioral therapy), they are advised to refer patients to psychiatric care. However, based on our clinical observation and literature review, this is not always a practical choice. Among 12 cases of PBS in a series collected by Kelleher\[11\], none was successfully referred to psychiatrists for psychological assessment. Responses included “immediate rejections”, “declined help” or “eventually accepted but be extremely bitterly about”\[11\]. Such reluctance was also observed by other clinicians, saying “many patients will never accept just a referral”\[8,13,18\]. Ideally, collaborated management provided by dentists, psychiatrists and exclusive psychotherapists would be the best approach. Unfortunately, this is hardly available in the current dental clinic setting.

In addition, the most common barrier preventing a clinician to apply psychopharmacotherapy is persuading patients to accept treatment. According to Watanabe et al\[8\], PBS patients had remarkably high ratios of refusal of pharmacological treatment, especially in those with dental triggers. PBS patients have such a strong belief that only dental treatment can relieve the symptom, resulting in them refusing any therapy other than dental interventions. Such belief like obsession or dominant idea grows stronger via repeated dental interventions and temporary relief. In order to shift that insufficient belief and to stop never-ending dental interventions, there needs a positive patient-doctor relationship built upon trust, empathy and efficient communication\[22\]. In our clinical observation, such evidence of neuromodulators helping to balance rCBF asymmetry patterns in successfully treated PBS cases would aid in patients’ understanding of medication necessity (Figure 4)\[16\].

Treatment for PBS is indeed difficult in the dental setting but not impossible as reported\[27\]. Prudent patient education with an etiological explanation based on neuroscience including brain images would help PBS patients to understand their situation and to be convinced for pharmacotherapy instead of repeated dental procedures. Moreover, even in cases of acceptance, side-effects and slow drug response would affect patient’s tendency to withdraw quickly. It seems to be more related to resistance to taking medication than to actual adverse effects. Careful contact with patients and delicate dosing are important during the follow-up period\[8\].

| Combination of TCA, benzodiazepine and D2 blocker | Amitriptyline; Lorazepam; Sulpiride | Average 99.8 d for hospitalization and 3.8 yr from discharge | Weight gain, Liver dysfunction, hyperprolactinaemia | 15/16 improved | Altered biochemical abnormalities related to neurotransmitter | Retrospective study of inpatients, n = 16 | Toyofuku\[32\], 2000 |
| Combination of D2 blocker and benzodiazepine | Sulpiride; Flunitrazepam | 10 mo | No report | Symptom improved | Unspecified | Single case report | Nakamura\[40\], 1996 |
CONCLUSION

PBS is a tremendously difficult and unusual dental phenomenon that is underreported and deserves more attention. Recent studies have revealed unexplained diversity patterns among PBS’s clinical manifestation and functional brain imaging that likely represent the available sub-phenotypes of this syndrome. Further research must be focused on elucidating pathophysiological mechanisms to pave the way for efficient treatment strategy.

REFERENCES

11. Kelleher MG, Rasaratnam L, Djemal S. The Paradoxes of Phantom Bite Syndrome or Occlusal Dysaesthesia (‘Dysesthesia’). *Dent Update* 2017; 44: 8-12, 15 [PMID: 29172307 DOI: 10.12968/duen.2017.44.1.8]
15. Kelleher MG, Rasaratnam L, Djemal S. The Paradoxes of Phantom Bite Syndrome or Occlusal Dysaesthesia (‘Dysesthesia’). *Dent Update* 2017; 44: 8-12, 15 [PMID: 29172307 DOI: 10.12968/duen.2017.44.1.8]
Tu TTH et al. A review of phantom bite syndrome


Breast cancer in schizophrenia could be interleukin-33-mediated

Milica M Borovcanin, Katarina Vesic

ORCID number: Milica M Borovcanin 0000-0002-2992-814X; Katarina Vesic 0000-0001-8861-4987.

Author contributions: Borovcanin MM and Vesic K contributed equally to this work; Borovcanin MM presented the idea, structured the manuscript, searched the literature, integrated all parts of the manuscript, and drew two figures; Vesic K searched the literature, provided some new insights, and performed the final revision to the manuscript; and both authors read, additionally discussed and approved the final version of the manuscript.

Conflict-of-interest statement: Borovcanin MM has received research funding from Ministry of Science and Technological Development of the Republic of Serbia (NO. 175069); Faculty of Medical Sciences, University of Kragujevac (NO. JP15-05). Vesic K declared no conflict of interest.

Supported by Ministry of Science and Technological Development of the Republic of Serbia (NO. 175069); Faculty of Medical Sciences, University of Kragujevac (NO. JP15-05).

Country/Territory of origin: Serbia

Specialty type: Psychiatry

Provenance and peer review: Invited article; Externally peer reviewed

Abstract

Recent epidemiological and genetic studies have revealed an interconnection between schizophrenia and breast cancer. The mutual underlying pathophysiological mechanisms may be immunologically driven. A new cluster of molecules called alarmins may be involved in sterile brain inflammation, and we have already reported the potential impact of interleukin-33 (IL-33) on positive symptoms onset and the role of its soluble trans-membranes full length receptor (sST2) on amelioration of negative symptoms in schizophrenia genesis. Furthermore, these molecules have already been shown to be involved in breast cancer etiopathogenesis. In this review article, we aim to describe the IL-33/suppressor of tumorigenicity 2 (ST2) axis as a crossroad in schizophrenia-breast cancer comorbidity. Considering that raloxifene could be tissue-specific and improve cognition and that tamoxifen resistance in breast carcinoma could be improved by strategies targeting IL-33, these selective estrogen receptor modulators could be useful in complementary treatment. These observations could guide further somatic, as well as psychiatric therapeutical protocols by incorporating what is known about immunity in schizophrenia.

Key Words: Interleukin-33; Schizophrenia; Breast cancer; Neurodegeneration

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: The frequent occurrence of breast carcinoma in female patients with schizophrenia could indicate a specific immunological perturbation in both diseases. A common denominator could be interleukin-33 (IL-33). In this review article, we aim to describe the IL-33/suppressor of tumorigenicity 2 axis as a crossroad in schizophrenia-breast cancer comorbidity. Considering that raloxifene could be tissue-specific and...
INTRODUCTION

The initial question is whether patients with schizophrenia are resistant or more susceptible to developing breast cancer. Recently, it has been reported that the prevalence rate of schizophrenia of 0.3%-0.7% in the world population remains stable [1]. There is a possibility that some mechanisms underlying the pathogenesis of schizophrenia also have beneficial or even protective properties for the onset of some somatic disorders, while others lead to somatic comorbidity and mortality.

Schizophrenia is a neurodegenerative disease with a complex pathogenesis and pathophysiology. This type of disorder is characterized by a chronic course, exacerbations, and progression leading to neurodegeneration over time. Genetic predisposition, individual and environmental factors, and specific immune responses have a significant impact on disease onset and clinical presentation. Moreover, genome-wide association studies in schizophrenia have revealed similar genetic backgrounds with some immunological properties and genetic overlap with breast carcinoma[2,3].

Scientific advances in the last decade have led to the recognition that interleukin-33 (IL-33) and neuroinflammation play a role in some aspects of neurodegenerative diseases. It is possible that IL-33 has various biological activities at different stages of the disease. It is necessary to consider disease progression with exacerbations and neurodegeneration, and to discuss the impact of IL-33 on somatic disturbances.

The alarmin molecule IL-33 as a marker of innate immunity was first studied in different phases of schizophrenia, but also in breast carcinoma patients[4]. In summary, we will try to further elucidate the possibility of the involvement of IL-33 in the onset of breast cancer in schizophrenia patients.

BREAST CANCER RISK AND SCHIZOPHRENIA

Disturbingly, approximately 2.1 million new cases of breast cancer were diagnosed worldwide in 2018, accounting for nearly one in four cancers in women and resulting in approximately 630000 deaths[5]. Previous epidemiological studies suggest that breast cancer is more common in patients with severe mental disorders than in the general population[6,7]. Among mental illnesses, a higher incidence of breast cancer has been observed in schizophrenia than in mood disorders[8]. Patients with schizophrenia have an increased risk of breast cancer, but not cancer overall[9,10]. The mortality cancer risk in schizophrenia is estimated to be 40% higher than in the general population and 51% higher than in individuals without schizophrenia[11].

The aspects of genetics, as well as epigenetic disturbances due to unhealthy lifestyle and harmful habits should be considered in this comorbidity. Recently, the genetic overlap of schizophrenia and breast cancer has been extensively reported[12]. A study by Lu et al[3] reported the phenotypic and genetic positive association of schizophrenia and breast cancer, estimated the percentage of genetic overlap to be 0.14 [95% cumulative incidence (CI): 0.09-0.19], and identified a shared locus at 19p13 (GATA-D2A) as a significant risk factor for the development of both diseases. A polymorphism of the breast cancer resistance protein transporter (ABCG2) was associated with a 4-fold higher probability of a relevant reduction in positive and negative schizophrenia syndrome[13]. This suggests a potential genetic overlap between the two diseases.
In recent years, epidemiological evidence of a controversial association between cancer and neurodegenerative diseases has been increasing. Common etiological factors may play opposing roles in the pathogenesis of neurodegeneration and breast cancer\textsuperscript{[14-15]}. Characteristic brain pathology leads to neuronal cell death and neurodegeneration over time, whereas cancer is dominated by the process of unlimited cell proliferation\textsuperscript{[16,17]}. Evidence suggests that common biological mechanisms such as oxidative stress, metabolic dysregulation and inflammation underlie both diseases, all of which could promote apoptosis and cell proliferation\textsuperscript{[18-20]}.

Some immune-metabolic specifiers observed in patients with schizophrenia, such as diabetes, hyperinsulinemia, insulin resistance, and hyperlipidemia, may also be involved in the development of breast cancer\textsuperscript{[21-23]}. Inflammatory changes that occur in patients with schizophrenia could also be marked as a cancer risk factor\textsuperscript{[24]} (Figure 1). IL-33 is significantly higher in serum, indicating tumor recurrence, which also indicates poor prognosis in patients with breast cancer\textsuperscript{[25-27]}. Similarly, higher serum levels of IL-33 are measured in schizophrenia during psychotic episodes\textsuperscript{[4]}.

So far, many contributing risk factors have been considered for the onset of breast carcinoma, and they seem to overlap with those considered lifestyle-related qualifiers in patients with schizophrenia. Behavioral habits such as smoking, alcohol consumption, obesity, physical inactivity and stress exposure are considered hazardous in the etiology of breast carcinoma\textsuperscript{[28-31]}. It is particularly important to acknowledge gender specificities in schizophrenia. Gender differences in breast carcinoma risk have been observed in people without mental illnesses, but also in patients with schizophrenia. Taiwanese women with schizophrenia were observed to have a 1.94-fold higher risk of breast cancer than non-schizophrenic controls\textsuperscript{[32]}. In addition, hormonal disturbances have been implicated as a risk factor for schizophrenia and breast cancer. According to previous findings, estrogen levels in women decrease during certain periods of life, such as during menstruation, after childbirth, and menopause, which can often lead to an exacerbation of schizophrenia symptoms and resistance to antipsychotic treatment\textsuperscript{[33,34]}. At the same time, estrogen intermediates carcinoma development\textsuperscript{[35]}.

**IL-33 ROLE IN NEUROINFLAMMATION AND NEURODEGENERATION**

IL-33 is a multifunctional cytokine, and an alarmin, a damage-associated molecular pattern (DAMP), an endogenous molecule released into the extracellular space upon cellular stress and damage. DAMPs can correct altered physiological states and regulate homeostasis at low concentrations, propagate inflammatory reactions at high concentrations, or even lead to trauma and activation of surrounding cells and recruitment of distant cells in excessive release\textsuperscript{[36]}. The role of DAMPs in neurodegenerative diseases has already been recognized in sterile inflammation. As recently summarized by Pandolfo et al\textsuperscript{[37]}, acute or chronic stress may trigger sterile inflammation associated with DAMPs and may be an etiopathogenetic mechanism for affective disorders.

In recent years, there is increasing evidence for an important immunomodulatory role of IL-33 in neurodegenerative diseases. IL-33 is a multifunctional cytokine that acts intracellularly as a nuclear factor and extracellularly as a cytokine\textsuperscript{[38]}. By binding to the trans-membranes full length receptor (sST2), IL-33 exerts its biological activity through the IL-33/suppressor of tumorigenicity 2 (ST2) signaling pathway\textsuperscript{[39]}. In contrast, the soluble form functions as a decoy receptor and limits the biological activity of IL-33\textsuperscript{[40]}. IL-33 is highly expressed in the brain and released from astrocytes and oligodendrocytes, while the ST2 receptor is expressed in glial cells\textsuperscript{[41]}. It has a dual function and may exert pro-inflammatory or anti-inflammatory effects in the central nervous system (CNS)\textsuperscript{[42,43]}. Up-regulated expression of IL-33 in peripheral cells contributes to blood-brain barrier disruption\textsuperscript{[44]}. By binding to the ST2 receptor, IL-33 promotes microglial activation and proliferation and enhances production of different cytokines, leading to an acute inflammatory response\textsuperscript{[45]}. IL-33 signaling is associated with regulatory T and B cell responses\textsuperscript{[46,47]}. By inducing microglial and macrophage polarization to an anti-inflammatory type 2 phenotype and phagocytosis, IL-33 exerts a neuroprotective and reparative role in the CNS\textsuperscript{[45,48]}. In Alzheimer’s disease (AD), soluble decoys trap IL-33 already produced at lower levels, and its concentrations in serum and cerebrospinal fluid are significantly reduced\textsuperscript{[49]}. IL-33 could exert a neuroprotective effect by inducing innate immunity to reduce soluble amyloid β levels and amyloid plaque deposition\textsuperscript{[50]}. Carlock et al\textsuperscript{[51]} in 2017 pointed out that deficiency of IL-33 caused tau abnormality and late-onset...
neurodegeneration in the cerebral cortex and hippocampus, accompanied by memory impairment. A recent study suggested that IL-33 gene mutations affect susceptibility to late-onset AD, which in turn confirms that IL-33 may exacerbate neuroinflammation and cognitive decline[52,53].

As a neuroinflammatory and neurodegenerative disease, multiple sclerosis (MS) is associated with increased expression of IL-33 in the periphery, white matter, and plaque areas of the MS brain[54-57]. However, the exact role of IL-33 has not been fully elucidated, and data in the current literature are inconsistent. Previous studies have supported the involvement of IL-33 as a proinflammatory cytokine in disease pathology[56]. However, several recent observations suggest a dominant neurorepairative role of IL-33 in MS. Studies on experimental autoimmune encephalomyelitis (EAE) reported a detrimental effect of IL-33 treatment on EAE severity. The identification of ST2 expression by oligodendrocytes indicates an important role in the myelination process during CNS development and the repair phase of MS[57,58].

Recently, the protective effect of IL-33 in some acute neurological states has been reported. For example, Miao et al[59] in 2021 highlighted that increased IL-33 levels might reduce brain damage in patients with intracerebral hemorrhage. The reviewed results support the evidence of elevated IL-33 serum levels in the acutisation of schizophrenia and depression, suggesting the involvement of this axis in the processes of relapse and recurrence of mental disorder[37].

**IL-33 ROLE IN CANCEROGENESIS**

IL-33 exerts its function by binding to the ST2 receptor, expressed on T helper (Th) 2 cells, but skewing toward Th1 cytokines has been found in MS[60,61]. IL-33 has been categorized as a promoter of Th2 immunity by inducing the production of IL-4, IL-5 and IL-13, M2 polarization of macrophages, and eosinophil recruitment[62]. The Th2-related cytokine can act as a pro-tumorigenic factor by limiting anti-tumor immunity and promoting extracellular matrix remodeling, but the localization of IL-33 is crucial for accurately distinguishing its effect in tumor biology[63].

IL-33 stimulates innate type 2 lymphoid cells (ILC2s), leading to the release of proinflammatory cytokines and type-2 inflammation[64]. An association with allergic disease has been observed in patients with autism spectrum disorder, and an immune pattern initiated by IL-33, ILC2 and mast cells has been confirmed[65]. To date, there are no data on the role of ILC2s in schizophrenia, and these findings in autism spectrum disorder as a neurodegenerative disease may serve as a basis for further investigation in psychosis. ILC2s are the largest subset in the lung and skin associated with pro-allergic and antiparasitic immunity. Dysregulation of their signaling circuitry may accelerate fibrotic responses and have predominantly carcinogenic activities[66]. In the breast cancer model, increased endogenous IL-33 was observed during cancer progression, further facilitating the intratumoral accumulation of immunosuppressive
IL-33-producing innate lymphoid cells and promoting the growth of breast cancer and metastases in the lungs[67]. Escalation of systemic IL-33 secretion could precipitate the carcinoma development in neurodegenerative diseases and schizophrenia as a representative.

**IL-33 IN THERAPEUTIC STRATEGIES FOR BREAST CANCER IN SCHIZOPHRENIA**

The first line of treatment for schizophrenia is antipsychotics, but nowadays it is surprising that these drugs could also have anticancer effects[68]. Previous studies have shown that some antipsychotics may cause higher prolactin levels, called “prolactin-raising”, and thus an increased risk of breast cancer by promoting carcinogenesis and transition to invasive carcinoma[69]. Breast cancer risk was higher in patients receiving first-generation antipsychotics and second-generation antipsychotics alone, as well as a combination of both, regardless of the mean exposure dose[32]. However, it must be emphasized that according to the World Federation of Societies of Biological Psychiatry, breast cancer is not listed as an antipsychotic-induced hyperprolactinemia adverse effect[70]. De Hert et al[71] reported that studies in patients with idiopathic hyperprolactinemia, prolactinomas, and Parkinson’s disease (PD) found no carcinogenic effects of prolactin. Hoehn et al[72] first described the possible inverse association between PD and neoplasms. This may suggest that excessive dopamine or prolactin release is not the exclusive mechanism leading to breast cancer development, but rather a milieu for more complex interactions. Our study exploring the impact of risperidone and paliperidone as prolactin-raising long-acting injections suggested a decrease in IL-33 serum levels in patients with schizophrenia in remission with possibly balancing and antitumorigenic properties[4].

The clinical presentation of schizophrenia differs in men and women and could be partly attributed to the neuroprotective properties of estrogens[73]. Modulation of dopamine 2 receptor occupancy of antipsychotics can also be associated with estrogens[74]. Estrogens protect women from infections and prevent mortality associated with inflammation by downregulating levels of pro-inflammatory cytokines IL-1β, IL-10, and tumor necrosis factor alpha[75].

Selective estrogen receptor modulators such as raloxifene have an antiestrogenic effect in the breast and uterus, but not in the brain and bone tissue[76]. Of particular interest to us was that raloxifene was not only effective in improving schizophrenia, but even more effective in improving cognitive symptoms in postmenopausal women with low estrogen levels[77,78] (Figure 2A). This tissue-specific dual action of the drug could be simultaneously used for targeted breast cancer therapy and mental state in schizophrenia patients and specifically explored in the context of IL-33 secretion.

Guidelines recommend tamoxifen as a standard of care for premenopausal women for 5-10 years[79]. In vitro studies could guide further investigations on the beneficial properties of the prolactin-elevating antipsychotics thioridazine and chlorpromazine in enhancing the effect of tamoxifen in tamoxifen-resistant human breast cancer cells[80]. IL-33 overexpression in breast cancer cells results in resistance to tamoxifen-induced tumor growth inhibition, while IL-33 knockdown corrects this problem[81]. This knockdown could be achieved through the action of antipsychotics (Figure 2B).

**CONCLUSION**

Individualized treatment is of great importance in modern medicine, as patients with schizophrenia must be treated equally. Diagnosis and treatment of somatic states in schizophrenia patients could influence behavioral changes and improve outcome. Prevention strategies for breast carcinoma onset in schizophrenia patients could be developed by understanding and recognizing the genetic background, lifestyle, and individual factors, altogether resulting in phase-specific immune dysregulation. IL-33 as a marker of innate immunity and an alarmin has been discussed in other neurodegenerative diseases. Further exploration of IL-33 as an alarmin in mental disorders should take into account gender, age at onset, duration of illness, frequency of disease acutisation, antipsychotic treatment, and a variety of comorbid somatic states including breast cancer. There appears to be an impact on the occurrence of positive symptoms and exacerbation of schizophrenia, but also on the progression of breast cancer, making IL-33 a candidate for centered therapy. We have shown that
antipsychotics with their anticarcinogenic properties could be beneficial, possibly through prolactin elevation, tissue-specific estrogen-sparing drugs, and additional IL-33 downregulation.

ACKNOWLEDGEMENTS

This review was enriched by the valuable collaboration with the Center for Molecular Medicine and Stem Cell Research, at the Faculty of Medical Sciences, University of Kragujevac, Kragujevac, Serbia. We would like to thank Bojana Mircetic for language editing.

REFERENCES

women in the United Kingdom.}


Borovcanin MM et al. IL-33 in schizophrenia breast cancer

10.1186/s13058-016-0733-1


33 Seeman MV, González-Rodríguez A. Use of psychotropic medication in women with psychotic disorders at menopause and beyond. Curr Opin Psychiatry 2018; 31: 183-192 [PMID: 29528895 DOI: 10.1097/YCO.0000000000000410]

34 Searles S, Makarewicz JA, Dumas JA. The role of estradiol in schizophrenia diagnosis and symptoms in postmenopausal women. Schizophrenia Res 2018; 196: 35-38 [PMID: 28587815 DOI: 10.1016/j.schres.2017.05.024]


39 Chackerian AA, Oldham ER, Murphy EE, Schmitz J, Planz S, Kastelein RA. IL-1 receptor accessory protein and ST2 comprise the IL-33 receptor complex. J Immunol 2007; 179: 2551-2555 [PMID: 17655517 DOI: 10.4049/jimmunol.179.4.2551]


53 Reverchon F, de Concini V, Larrigadie V, Benmerzoug S, Briault S, Togbé D, Ryfbel F, Quesniaux VJ, Menaut E. Hippocampal interleukin-33 mediates neuroinflammation-induced cognitive


73 Brand BA, de Boer JN, Sommer IEC. Estrogens in schizophrenia: progress, current challenges and opportunities. Curr Opin Psychiatry 2021; 34: 228-237 [PMID: 33560022 DOI: 10.1097/YCO.0000000000000699]


Borovcanin MM et al. IL-33 in schizophrenia breast cancer

10.1517/14740338.2015.1014799]
77 de Boer J, Priksen M, Lei WU, Begemann M, Sommer I. The effect of raloxifene augmentation in men and women with a schizophrenia spectrum disorder: a systematic review and meta-analysis. NPJ Schizophreniat 2018; 4: 1 [PMID: 29321530 DOI: 10.1038/s41537-017-0043-3]
Basic Study

CPEB1, a novel risk gene in recent-onset schizophrenia, contributes to mitochondrial complex I defect caused by a defective provirus ERVWE1

Ya-Ru Xia, Xiao-Cui Wei, Wen-Shi Li, Qiu-Jin Yan, Xiu-Lin Wu, Wei Yao, Xu-Hang Li, Fan Zhu

Abstract

BACKGROUND
Schizophrenia afflicts 1% of the world population. Clinical studies suggest that schizophrenia patients may have an imbalance of mitochondrial energy metabolism via inhibition of mitochondrial complex I activity. Moreover, recent studies have shown that ERVWE1 is also a risk factor for schizophrenia. Nevertheless, there is no available literature concerning the relationship between complex I deficits and ERVWE1 in schizophrenia. Identifying risk factors and blood-based biomarkers for schizophrenia may provide new guidelines for early interventions and prevention programs.

AIM
To address novel potential risk factors and the underlying mechanisms of mitochondrial complex I deficiency caused by ERVWE1 in schizophrenia.

METHODS
Quantitative polymerase chain reaction (qPCR) and enzyme-linked immunosorbent assay were used to detect differentially expressed risk factors in blood samples. Clinical statistical analyses were performed by median analyses and Mann-Whitney U analyses. Spearman’s rank correlation was applied to examine the correlation between different risk factors in blood samples. qPCR, western blot analysis, and luciferase assay were performed to confirm the relationship among ERVWE1, cytoplasmic polyadenylation element-binding protein 1 (CPEB1), NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2), and NDUFV2 pseudogene (NDUFV2P1). The complex I enzyme activity microplate assay was
involved humans. Informed consent was obtained for experimentation with human subjects and their privacy rights were always observed.

**Conflict-of-interest statement:** No potential conflicts of interest.

**Data sharing statement:** The data used and analyzed during the current study are available from the corresponding author upon request.

**Supported by** the National Natural Science Foundation of China, No. 81971943, No. 81772196, No. 31470264, No. 81271820, No. 30870789, and No. 30300117; the Stanley Foundation of United States, No. 06R-1366 (for Zhu F); and the Medical Science Advancement Program (Basic Medical Sciences) of Wuhan University, No. TFJC 2018002.

**Country/Territory of origin:** China

**Specialty type:** Psychiatry

**Provenance and peer review:** Invited article; Externally peer reviewed

**Peer-review report’s scientific quality classification**

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: [http://creativecommons.org/License s/by-nc/4.0/](http://creativecommons.org/Licenses/by-nc/4.0/)

**Received:** April 25, 2021

**Peer-review started:** April 25, 2021

Carried out to evaluate the complex I activity induced by ERVWE1.

**RESULTS**

Herein, we reported decreasing levels of CPEB1 and NDUFV2 in schizophrenia patients. Further studies showed that ERVWE1 was negatively correlated with CPEB1 and NDUFV2 in schizophrenia. Moreover, NDUFV2P1 was increased and demonstrated a significant positive correlation with ERVWE1 and a negative correlation with NDUFV2 in schizophrenia. *In vitro* experiments disclosed that ERVWE1 suppressed NDUFV2 expression and promoter activity by increasing NDUFV2P1 level. The luciferase assay revealed that ERVWE1 could enhance the promoter activity of NDUFV2P1. Additionally, ERVWE1 downregulated the expression of CPEB1 by suppressing the promoter activity, and the 400 base pair sequence at the 3′ terminus of the promoter was the minimum sequence required. Advanced studies showed that CPEB1 participated in regulating the NDUFV2P1/NDUFV2 axis mediated by ERVWE1. Finally, we found that ERVWE1 inhibited complex I activity in SH-SY5Y cells via the CPEB1/NDUFV2P1/NDUFV2 signaling pathway.

**CONCLUSION**

In conclusion, CPEB1 and NDUFV2 might be novel potential blood-based biomarkers and pathogenic factors in schizophrenia. Our findings also reveal a novel mechanism of ERVWE1 in the etiology of schizophrenia.

**Key Words:** ERVWE1; CPEB1; NADH dehydrogenase ubiquinone flavoprotein 2; complex I; Pseudogene

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** Schizophrenia is a devastating psychiatric disorder. Clinical studies suggest complex I deficits and abnormal expression of ERVWE1 in schizophrenia. We found that CPEB1 might be a novel blood-based biomarker for schizophrenia. Analyses indicated that ERVWE1 was negatively correlated with CPEB1 and NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) in schizophrenia patients. CPEB1 functioned as a mediator of the NDUFV2 pseudogene (NDUFV2P1)/NDUFV2 pathway induced by ERVWE1. Moreover, ERVWE1 inhibited complex I activity via the CPEB1/NDUFV2P1/NDUFV2 pathway. Thus, CPEB1 and NDUFV2 may be independent risk factors for schizophrenia. Our findings also reveal the role of ERVWE1 in modulating mitochondrial energy metabolism in schizophrenia.

**Citation:** Xia YR, Wei XC, Li WS, Yan QJ, Wu XL, Yao W, Li XH, Zhu F. CPEB1, a novel risk gene in recent-onset schizophrenia, contributes to mitochondrial complex I defect caused by a defective provirus ERVWE1. *World J Psychiatr* 2021; 11(11): 1075-1094


**DOI:** [https://dx.doi.org/10.5498/wjp.v11.i11.1075](https://dx.doi.org/10.5498/wjp.v11.i11.1075)

**INTRODUCTION**

Schizophrenia is a chronic brain disorder with a worldwide prevalence of approximately 1%[1]. It causes many symptoms including hallucinations, delusions, breaking from reality, and a lack of motivation[2]. Schizophrenia imposes a financial burden on society because of high indirect costs[3]. There are numerous hypotheses about schizophrenia. One of the currently prevailing hypothesis proposes that mitochondrial metabolic dysfunction[4–6] is strongly involved in the pathophysiology of schizophrenia. Mitochondrial complex I deficiency is the most frequent cause of mitochondrial metabolic dysfunction implicated in the development of schizophrenia[7–9]. The mitochondrial oxidative phosphorylation (OXPHOS) system comprises five multimeric protein complexes, and plays a central role in cellular energy metabolism. Complex I is the first entry point of electrons into the respiratory chain and is suggested as the rate-limiting step in OXPHOS. NADH dehydrogenase ubiquinone...
flavoprotein 2 (NDUFV2) assembles in the catalytic component of complex I. Ample evidence has revealed that NDUFV2 is a candidate gene for schizophrenia[10-12]. NDUFV2 pseudogene (NDUFV2P1), an inactive locus of the NDUFV2 subunit of the complex I[13], has increased expression in schizophrenia-derived cells[14]. Knockout of CPEB1, a sequence-specific RNA-binding protein[15], reduces NDUFV2 protein expression and inhibits complex I activity in mice[16]. Nonetheless, there is no definitive hypothesis on the pathogenesis of schizophrenia.

Human endogenous retroviruses (HERVs), which comprise about 8% of the human genome, have emerged as novel risk factors for schizophrenia[17-19]. As remnants of ancient retroviral infections, most HERVs have acquired inactivating mutations during evolution[20]. In some instances, some HERVs sequences remain dormant functional copies encoding retroviral proteins[21] and play an active role in human placentation development and the innate immune system. HERVs possess a similar genome structure as typical exogenous retroviruses and are composed of gag, pol, and env regions sandwiched between two long terminal repeats[21]. Conventionally, based on sequence similarity with their exogenous retrovirus, HERVs are divided into three classes: class I (gamma-retroviruses), class II (beta-retroviruses), and class III (spuma-retroviruses)[22].

The HERV W family (HERV-W), the oldest group of HERVs, belongs to class I transposable elements. ERVWE1, HERV-W-derived envelope proteins at chromosome 7, is also known as HERV-W env or Syncytin-1[23]. ERVWE1 is a cell-cell fusion glycoprotein that is usually expressed on the surface of cytotrophoblasts and syncytiotrophoblast. It mediates trophoblast fusion[24] and blocks maternal immune[25], which are essential for normal placental development and the innate immune system. HERVs possess a similar genome structure as typical exogenous retroviruses and are composed of gag, pol, and env regions sandwiched between two long terminal repeats[21]. Conventionally, based on sequence similarity with their exogenous retrovirus, HERVs are divided into three classes: class I (gamma-retroviruses), class II (beta-retroviruses), and class III (spuma-retroviruses)[22].

The HERV W family (HERV-W), the oldest group of HERVs, belongs to class I transposable elements. ERVWE1, HERV-W-derived envelope proteins at chromosome 7, is also known as HERV-W env or Syncytin-1[23]. ERVWE1 is a cell-cell fusion glycoprotein that is usually expressed on the surface of cytotrophoblasts and syncytiotrophoblast. It mediates trophoblast fusion[24] and blocks maternal immune[25], which are essential for normal placental development and the innate immune system. HERVs possess a similar genome structure as typical exogenous retroviruses and are composed of gag, pol, and env regions sandwiched between two long terminal repeats[21]. Conventionally, based on sequence similarity with their exogenous retrovirus, HERVs are divided into three classes: class I (gamma-retroviruses), class II (beta-retroviruses), and class III (spuma-retroviruses)[22].

In summary, CPEB1 and NDUFV2 might be novel schizophrenia risk factors. To the best of our knowledge, no similar observations have been published to date. Moreover, our results indicated that ERVWE1 abolished complex I activity by regulating parental- and pseudo-NDUFV2 genes via suppressing CPEB1 in schizophrenia. Our findings might provide new insights into how ERVWE1 impairs mitochondrial energy metabolism and contributes to the pathogenesis of schizophrenia.

MATERIALS AND METHODS

Blood samples
Blood samples from patients and normal individuals, including whole peripheral blood for RNA isolation and serum for ELISA analyses, were collected from the Renmin Hospital, Wuhan University (Wuhan, China). All patients with recent-onset schizophrenia met the Diagnostic and Statistical Manual of Mental Disorders, 4th edition criteria for schizophrenia.
None of the patients had a history of psychiatric illness. Informed consent was obtained from all subjects, and the study was approved by the Institutional Review Board of Wuhan University, School of Basic Medical Sciences. All samples were preserved at -80°C before utilization. There were no significant differences in median age, education, body mass index, smoking status, and sex between control and schizophrenia patients. Demographics are presented in Supplementary Tables 1 and 2.

**Plasmid construction**

The pCMV-ERVWE1 plasmid was constructed as previously described[18]. Human CPEB1 (NM_030594.5) and NDUFV2P1 (NC_000019.10:c53224879-53223909) genes were amplified separately according to sequences of these genes in National Center for Biotechnology Information (Bethesda, MD, United States). The promoters of human CPEB1 (-800 to 0, -600 to 0, -400 to 0, and -200 to 0), NDUFV2 (-1100 to +100), and NDUFV2P1 (-900 to 0) were cloned into the luciferase reporter vector pGL3-basic, which contains a firefly luciferase gene. All primers were designed using Primer premier 5.0 and described in Supplementary Table 3.

One short hairpin RNA (shRNA) targeting NDUFV2P1 (shNDUFV2P1, 5′-CCAAGGAGTGGACGCTTCT-3′) and the control shRNA (sh-Control, 5′-CCAAGTGCGGCACGTAGTCT-3′) were cloned into the pSilencer 2.1-U6 neo shRNA expression vector (AM5764; Ambion Inc., Austin, TX, United States). All of the constructs were confirmed by sequencing (Sangon Biotech, Shanghai, China).

**Cell culture and transfection**

SH-SY5Y (CRL-2266) human neuroblastoma cells were obtained from American Type Culture Collection. SH-SY5Y cells were cultured in a 1:1 mixture of minimum essential medium (2225320; Gibco, Gaithersburg, MD, United States) and F-12 nutrient mixture (2209586; Gibco). All media were supplemented with 10% fetal bovine serum (2001003, Biological Industries, Beit HaEmek, Israel), 1% penicillin-streptomycin (2211093; Gibco) and 1% sodium pyruvate (2185865; Gibco) under 5% CO₂ at 37 °C. Transfection was performed with Lipofectamine 2000 transfection reagent (11668-019; Invitrogen, Carlsbad, CA, United States) according to the manufacturer’s instructions. Cells were incubated for further analyses after 24, 36, and 48 h post-transfection.

**RNA extraction and reverse transcription**

Total blood RNA was extracted from whole blood according to the standard protocols for TRIzol LS reagent (10296028; Invitrogen). Total cellular RNA was isolated with TRIzol reagent (15596018; Invitrogen), followed by chloroform separation and isopropanol precipitation. Then, 1 µg total RNA samples were reverse-transcribed into cDNA using the ReverTra Ace qPCR RT master mix with gDNA remover (FSQ-301; Toyobo, Osaka, Japan) according to the manufacturer’s recommendations.

**Quantitative PCR**

The mRNA levels were measured by quantitative PCR (qPCR) utilizing the 2x Sybr Green qPCR mix (2992239AX; Aidlab Biotechnologies Co. Ltd., Beijing, China) using a mini opticon detector (Bio-Rad, Hercules, CA, United States). GAPDH was used as an internal control, and the relative mRNA expression was calculated using the 2^(-ΔΔCt) method. The primers were designed according to gene sequences in NCBI using primer premier 5.0, and are listed in Supplementary Table 4.

**Protein extraction and western blot analysis**

Cultured cells were homogenized and lysed in M-PER TM mammalian protein extraction reagent (UC282138; Thermo Fisher Scientific, Waltham, MA, United States) containing protease and phosphatase inhibitors (Roche Applied Science, Indianapolis, IN, United States). Protein concentrations were quantified using the Pierce TM BCA Protein Assay (UD281372; Thermo Fisher Scientific). Protein samples were loaded onto a 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and was then electrotransferred to a PVDF membrane (IPVH00010; Amersham Biosciences, Piscataway, NJ, United States). The following antibodies purchased from ABclonal Technology (Woburn, MA, United States) were used in this study: rabbit anti-ERVWE1 antibody (dilution 1:1000, A16522), rabbit anti-CPEB1 antibody (dilution 1:1000, A5913), rabbit anti-NDUFV2 antibody (dilution 1:1500, A7442), mouse anti-GAPDH antibody (dilution 1:50000, AC002), goat anti-mouse IgG-horseradish peroxidase (HRP) (dilution 1:5000, AS003), and goat anti-rabbit IgG-HRP (dilution 1:5000, AS014). The bands were visualized using the Millipore Immobilon Western [18].
Chemiluminescent HRP Substrate (WBKLS0500; Millipore, Burlington, MA, United States). Images were captured using the Tanon 5200 chemiluminescence imaging system (Tanon, Shanghai, China). Target protein expression levels were normalized to GAPDH.

**Luciferase assay**
The luciferase activity was carried out using the Dual-Luciferase Assay System (E1960; Promega, Madison, WI, United States) in accordance with the manufacturer’s protocol. Firefly luciferase activity was normalized to pRL-CMV Renilla luciferase activity. The promoter activity is expressed as the mean ± standard deviation (SD) of at least triplicate replicates.

**ELISA**
The expression levels of human CPEB1 (072020; Huaer Biotech Co. Ltd, Wuhan, China) and NDUFV2 (202103; Meiman Industrial Co. Ltd, Jiangsu, China) in serum were measured by an ELISA kit according to the manufacturer’s recommendations. Absorbance was measured at 450 nm using a spectrophotometer (Thermo Fisher Scientific). The concentrations of CPEB1 and NDUFV2 were determined by comparing the optical density of the samples to the standard curve.

**Analyses of mitochondrial OXPHOS complex I activity**
Mitochondrial OXPHOS complex I activity was detected by the Immunocapture ELISA Kit following the manufacturer’s instructions (AB109721; Abcam, Cambridge, United Kingdom). After cells were lysed with detergent solution, protein concentration was adjusted to 1 mg/mL in incubation solution. The diluted 200 μL sample and control were loaded onto the wells of a microplate pre-coated with complex I capture antibody for 3 h at room temperature. Complex I activity was determined by measuring the oxidation of NADH to NAD⁺ and simultaneous reduction of a dye, which leads to increased absorbance at 450 nm. The activity was expressed as the change in absorbance per min using this equation: complex I activity = [Abs 1 – Abs 2] / [Time 1 - Time 2].

**Statistical analyses**
SPSS and GraphPad Prism were used for the statistical analyses. For the blood sample results, the median analyses and Mann-Whitney U analyses were performed to compare the different expressions of ERVWE1, CPEB1, NDUFV2P1, and NDUFV2 between the schizophrenia and control groups. Correlation analyses were performed using the Spearman’s rank correlation. All data were obtained from at least three independent experiments and presented as the mean ± SD. Statistical and data analyses were performed by one-way analysis of variance and Student’s t-tests. Significance was considered at \( P < 0.05 \).

**RESULTS**

**Degradation of CPEB1 and NDUFV2 proteins and the correlation among ERVWE1, NDUFV2, and NDUFV2P1 in schizophrenia**

Many biomarkers in serum have been studied to determine whether they can help diagnose schizophrenia. A genome-wide association study (GWAS) identified CPEB1 as a new susceptibility locus for schizophrenia[27]. Nevertheless, there has been no report on the expression of CPEB1 in patients with schizophrenia. Here, we showed that CPEB1 was significantly downregulated in schizophrenia compared to healthy people in blood samples using ELISA \( (P < 0.05; \) Figure 1A and Table 1). We further found decreased mRNA and protein levels of NDUFV2 in schizophrenia patients \( (P < 0.05, \) Figure 1B; \( P < 0.01, \) Figure 1C and Table 2). NDUFV2P1 (NDUFV2 pseudogene) was significantly upregulated in schizophrenia patients \( (P < 0.01; \) Figure 1D). We also found increased ERVWE1 mRNA expression in schizophrenia patients \( (P < 0.01; \) Figure 1E). Correlation analyses through linear regression revealed that ERVWE1 mRNA was negatively correlated with CPEB1 \( (P < 0.001; \) Figure 1F) and NDUFV2 protein \( (P < 0.001; \) Figure 1G), and positively correlated with NDUFV2P1 transcripts \( (P < 0.001; \) Figure 1H) in schizophrenia. Moreover, high levels of NDUFV2P1 mRNA were correlated with decreased NDUFV2 protein in schizophrenia \( (P < 0.001; \) Figure 1I). Univariate and multivariate analyses identified CPEB1 and NDUFV2 as two possible independent risk factors for schizophrenia (Supplementary Table 5).
In summary, CPEB1 and NDUFV2 might be novel potential blood-based biomarkers and possible pathogenic factors for schizophrenia, and are negatively correlated with ERVWE1 in schizophrenia.

**ERVWE1 reduces the expression of NDUFV2 by upregulating NDUFV2P1.**

The clinical data showed a significant correlation among ERVWE1, NDUFV2P1, and NDUFV2 gene expression in schizophrenia. Thus, to evaluate the causal relationship among ERVWE1, NDUFV2P1, and NDUFV2, experiments were performed in human neuroblastoma cells. ERVWE1 expression in transfected cells was detected by qPCR and western blotting (Supplementary Figure 1A and B). The mRNA level of NDUFV2P1 was significantly increased in SH-SY5Y cells after overexpression of ERVWE1 ($P < 0.05$; Figure 2A). Luciferase assays showed that overexpression of ERVWE1 greatly enhanced NDUFV2P1 promoter activity in SH-SY5Y cells ($P < 0.01$; Figure 2B). NDUFV2 is the parental gene of NDUFV2P1[13]. We found that ERVWE1 decreased both the mRNA ($P < 0.05$; Figure 2C) and protein levels of NDUFV2 in SH-SY5Y cells ($P < 0.05$; Figure 2D). Moreover, ERVWE1 reduced the expression of NDUFV2 by suppressing promoter activity in SH-SY5Y cells ($P < 0.01$; Figure 2E). These findings suggest that ERVWE1 downregulates the expression of NDUFV2 while upregulating its pseudogene, NDUFV2P1, in SH-SY5Y cells.

Pseudogene transcripts usually regulate the expression of their parental genes[28]. As shown in Figure 3A, the mRNA levels of NDUFV2 were significantly decreased in the NDUFV2P1 (+) group compared with the control ($P < 0.05$). Consistently, western blot analysis also confirmed that NDUFV2P1 inhibited the expression levels of NDUFV2 in SH-SY5Y cells ($P < 0.05$; Figure 3B). NDUFV2P1 expression in transfected cells was detected by qPCR (Supplementary Figure 2).

### Table 1 The concentration of CPEB1 in the blood of control and schizophrenia patients

<table>
<thead>
<tr>
<th>Control</th>
<th>n</th>
<th>Schizophrenia</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Mean</td>
<td>25.76</td>
<td>Mean</td>
<td>14.63</td>
</tr>
<tr>
<td>Median</td>
<td>22.47</td>
<td>Median</td>
<td>16.19</td>
</tr>
<tr>
<td>SD</td>
<td>11.70</td>
<td>SD</td>
<td>5.39</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.89</td>
<td>Skewness</td>
<td>-0.93</td>
</tr>
<tr>
<td>SE of Skewness</td>
<td>0.38</td>
<td>SE of Skewness</td>
<td>0.37</td>
</tr>
<tr>
<td>Range</td>
<td>53.26</td>
<td>Range</td>
<td>18.63</td>
</tr>
<tr>
<td>Minimum</td>
<td>16.46</td>
<td>Minimum</td>
<td>2.14</td>
</tr>
<tr>
<td>Maximum</td>
<td>69.70</td>
<td>Maximum</td>
<td>23.67</td>
</tr>
</tbody>
</table>

SD: Standard deviation; SE: Standard error.

### Table 2 The concentration of NADH dehydrogenase ubiquinone flavoprotein 2 in the blood of control and schizophrenia patients

<table>
<thead>
<tr>
<th>Control</th>
<th>n</th>
<th>Schizophrenia</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Mean</td>
<td>28.16</td>
<td>Mean</td>
<td>20.14</td>
</tr>
<tr>
<td>Median</td>
<td>26.87</td>
<td>Median</td>
<td>20.51</td>
</tr>
<tr>
<td>SD</td>
<td>11.02</td>
<td>SD</td>
<td>5.47</td>
</tr>
<tr>
<td>Skewness</td>
<td>2.691</td>
<td>Skewness</td>
<td>-0.67</td>
</tr>
<tr>
<td>SE of Skewness</td>
<td>0.38</td>
<td>SE of Skewness</td>
<td>0.37</td>
</tr>
<tr>
<td>Range</td>
<td>60.09</td>
<td>Range</td>
<td>23.23</td>
</tr>
<tr>
<td>Minimum</td>
<td>15.76</td>
<td>Minimum</td>
<td>4.91</td>
</tr>
<tr>
<td>Maximum</td>
<td>75.85</td>
<td>Maximum</td>
<td>28.14</td>
</tr>
</tbody>
</table>

SD: Standard deviation; SE: Standard error.
Figure 1 Cytoplasmic polyadenylation element-binding protein 1 (CPEB1) and NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2)
deficiency and the correlation among ERVWE1, CPEB1, NDUFV2, and NDUFV2 pseudogene (NDUFV2P1) in schizophrenia patients. A and C: Respectively represent the protein expression of cytoplasmic polyadenylation element-binding protein 1 (CPEB1) and NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) in the schizophrenia patients (n = 41) and in the control groups (n = 38) using enzyme-linked immunosorbent; B, D, and E: Respectively represent the mRNA levels of NDUFV2, NDUFV2 pseudogene (NDUFV2P1), and ERVWE1 in schizophrenia patients (n = 19) and the control groups (n = 38) using quantitative polymerase chain reaction; F: Correlation between ERVWE1 mRNA levels and CPEB1 protein levels in patients with schizophrenia and control groups (n = 57), where Y is the protein expression for CPEB1 and X is ERVWE1 mRNA value for each sample; G: Correlation between ERVWE1 mRNA levels and NDUFV2 protein levels in patients with schizophrenia and control groups (n = 57), where Y is the protein expression for NDUFV2 and X is ERVWE1 mRNA value for each sample; H: Correlation between ERVWE1 mRNA levels and CPEB1 protein levels in patients with schizophrenia and control groups (n = 57), where Y is the mRNA expression for NDUFV2P1 and X is ERVWE1 mRNA value for each sample; I: Correlation between NDUFV2P1 mRNA levels and NDUFV2 protein levels in patients with schizophrenia and control groups (n = 57), where Y is the protein expression for NDUFV2 and X is NDUFV2P1 mRNA value for each sample. *P < 0.05; **P < 0.01.

To further explore whether NDUFV2P1 mediated the regulatory effect of ERVWE1 on the expression of NDUFV2, the shNDUFV2P1 plasmid was constructed. Knockdown of NDUFV2P1 substantially elevated the ERVWE1-induced mRNA expressions of NDUFV2 in SH-SY5Y cells (P < 0.05; Figure 3C). Moreover, western blotting showed similar results of increased NDUFV2 protein expression after NDUFV2P1 knockdown (P < 0.05; Figure 3D). Hence, these data demonstrated that ERVWE1 could mediate the downregulation of NDUFV2 by upregulating NDUFV2P1 in SH-SY5Y cells.

We further examined whether NDUFV2P1 had an effect on NDUFV2 promoter activity. Luciferase assays indicated that NDUFV2P1 inhibited NDUFV2 promoter activity (P < 0.05; Figure 3E) in SH-SY5Y cells. To explore the role of NDUFV2P1 in the ERVWE1-induced promoter activity of NDUFV2, shNDUFV2P1 was used. The reduced promoter activity was significantly recovered after knockdown of NDUFV2P1 (P < 0.01; Figure 3F), suggesting that ERVWE1 suppressed NDUFV2 promoter by elevating NDUFV2P1 transcript.

**ERVWE1 mediates NDUFV2P1/NDUFV2 signaling pathway through the regulation of CPEB1**

Our clinical data revealed CPEB1 was downregulated, and demonstrated a negative correlation between CPEB1 and ERVWE1 in schizophrenia patients. It would be intriguing to assess the effect of ERVWE1 on CPEB1. We found that both mRNA (P < 0.01; Figure 4A) and protein (P < 0.05; Figure 4B) expression of CPEB1 was significantly lower after overexpression of ERVWE1 in SH-SY5Y cells than in control cells. Moreover, the luciferase assay demonstrated that ERVWE1 suppressed the promoter activity of CPEB1 (P < 0.01; Figure 4C). In order to define the minimum sequence essential for the transcriptional control of CPEB1 gene, four sequentially truncated promoters were examined in transfection assay. The luciferase reporter assay demonstrated that the promoter ranged from -400 to 0 was the minimum sequence required for biological activity (P < 0.05; Figure 4D). Furthermore, we found that CPEB1 could downregulate NDUFV2P1 mRNA levels in SH-SY5Y cells (P < 0.05; Figure 4E). qPCR and western blot analysis confirmed CPEB1 transfection (Supplementary Figure 3A and B). Additionally, overexpression of CPEB1 substantially impaired the ERVWE1-induced mRNA expression of NDUFV2P1 in SH-SY5Y cells, suggesting that ERVWE1-induced NDUFV2P1 was mediated by CPEB1 (P < 0.05; Figure 4F).

Several studies have reported that CPEB1 regulates NDUFV2 expression by a post-transcriptional mechanism[16]. We also found no change in the mRNA expression of NDUFV2 after transfection with CPEB1 (Figure 5A). Western blotting showed that CPEB1 elevated the protein levels of NDUFV2 (P < 0.01; Figure 5B). We further assessed the possible role of CPEB1 in ERVWE1-mediated NDUFV2 expression. Our results revealed that NDUFV2 mRNA was not altered, regardless of whether CPEB1 was overexpressed compared with control cells (Figure 5C). Regardless of whether CPEB1 was highly expressed, western blot analyses revealed that CPEB1 could relieve the inhibitory effect of ERVWE1 on NDUFV2 (P < 0.05; Figure 5D). Together, these results suggest that ERVWE1 regulates the NDUFV2P1/NDUFV2 signaling pathway through the suppression of CPEB1.

**ERVWE1 impairs electron transport chain complex I activity via the CPEB1/NDUFV2P1/NDUFV2 signaling pathway**

NDUFV2 is a subunit of mitochondrial complex I[11]. Our above results demonstrated that ERVWE1 significantly reduced the mRNA and protein expressions of NDUFV2. Thus, we further analyzed the impact of ERVWE1 on complex I activity. Successful
Overexpression of ERVWE1 elevated NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) pseudogene (NDUFV2P1) and reduced NDUFV2 levels. A and C: Respectively represent the mRNA levels of NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) pseudogene (NDUFV2P1) and NDUFV2 in ERVWE1-transfected SH-SY5Y cells using quantitative polymerase chain reaction; B: Luciferase assays of pGL3-NDUFV2P1 promoter co-transfected with pCMV-ERVWE1 or control vector in SH-SY5Y cells; D: Western blots of NDUFV2 and GAPDH in SH-SY5Y cells; E: Luciferase assays of pGL3-NDUFV2 promoter co-transfected with pCMV-ERVWE1 or control vector in SH-SY5Y cells. Each bar represents the mean ± SD of three independent experiments. *P < 0.05; **P < 0.01.

transfection was verified by western blot analyses in the transfected cells (Supplementary Figure 4). The complex I activity was decreased in ERVWE1-overexpressing cells compared to that in control cells (P < 0.05; Figure 6A).

We measured complex I activity after the knockdown of NDUFV2P1 in SH-SY5Y cells. Downregulating NDUFV2P1 could mitigate ERVWE1-induced inhibition of mitochondrial complex I in SH-SY5Y cells (P < 0.05; Figure 6B), indicating that ERVWE1 inhibited complex I activity by increasing NDUFV2P1.

Another experiment showed that CPEB1 could reverse the complex I deficiency induced by ERVWE1 in SH-SY5Y cells (P < 0.05; Figure 6C), suggesting that ERVWE1 could regulate complex I activity via suppressing CPEB1 expression.
Figure 3 Role of NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) pseudogene (NDUFV2P1) in ERVWE1-induced downregulation of NDUFV2. A: The mRNA levels of NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) in NDUFV2 pseudogene (NDUFV2P1)-overexpressing SH-SY5Y cells using quantitative polymerase chain reaction (qPCR); B: Western blots of NDUFV2 and GAPDH in SH-SY5Y cells; C: The mRNA levels of NDUFV2 in SH-SY5Y cells after co-transfection with ERVWE1 and shNDUFV2P1 using qPCR; D: Western blots of ERVWE1, NDUFV2, and GAPDH in SH-SY5Y cells; E: Luciferase assays of pGL3-NDUFV2 promoter co-transfected with pCMV-NDUFV2P1 or control vector in SH-SY5Y cells; F: Luciferase assays of...
In summary, our data indicate that ERVWE1 modulates mitochondrial energy metabolism, primarily of complex I, via the CPEB1/NDUFV2P1/NDUFV2 signaling pathway in schizophrenia.

**DISCUSSION**

Schizophrenia is a multifactorial mental disorder with abnormal cognitive functions and behavior. It is characterized by continuous or relapsing episodes of psychosis. Potential blood-based biomarkers have the potential to increase the precision of the diagnostic and prognostic process for schizophrenia. The field of blood-biomarker discovery for schizophrenia is coming to fruition as patient blood is an easily accessible biological sample[29]. An increasing number of molecular detection techniques with high specificity and sensitivity are currently applied in the search for serum markers or genetic markers in peripheral blood. Previous studies have addressed potential blood-based biomarkers for schizophrenia, such as microRNA[30,31], insulin-related peptides[32,33], and inflammatory cytokines[34]. However, no biomarkers have been successfully used in the clinical diagnosis of schizophrenia. GWAS has implicated CPEB1, which possesses a highly conserved RNA-binding domain and controls the cytoplasmic polyadenylation of mRNA by combining with the specific sequences, as a novel schizophrenia susceptibility gene[27,35]. However, there is no experimental evidence of an association between CPEB1 and schizophrenia. Clinical data of this study do not follow a normal distribution, so Spearman’s rank correlation was used for correlation analyses. In this paper, clinical data showed that CPEB1 was decreased and negatively correlated with ERVWE1 in schizophrenia. Univariate and multivariate analyses suggested that CPEB1 might be a novel potential blood-based biomarker and risk factor for schizophrenia.

The mouse model has shown that knockout of CPEB1 downregulates NDUFV2 protein expression[16]. The abnormal expression of NDUFV2 has been found in schizophrenia[7,36]. Our data displayed decreased NDUFV2 transcription in schizophrenia, in contrast to the data reported by Haghighatfard et al[36]. The reason for this might be the primer used by Haghighatfard’s group can detect not only NDUFV2 but also NDUFV2P1. Herein, our primers could successfully detect the transcript levels of NDUFV2 and NDUFV2P1. The NDUFV2P1 transcript was upregulated in schizophrenia and was negatively regulated NDUFV2. Additionally, our previous study showed false positives using nonspecific primers to detect both NDUFV2 and NDUFV2P1 (Supplementary Figure 5). We also found decreased NDUFV2 protein expression, consistent with the report by Bergman et al[14]. However, Bergman et al[14] showed no changes in NDUFV2 mRNA. The primary factor driving the difference between our and Bergman’s findings is that we focused on recent-onset schizophrenia patients. In our sample, all subjects with recent-onset schizophrenia were free from antipsychotics, whereas in the schizophrenia samples of Bergman’s group, all but one participant were using antipsychotic medications. A previous study demonstrated that exposure to antipsychotic medication can increase NDUFV2 mRNA expression in Sprague-Dawley rats[37]. Hence, antipsychotic medication treatment might reverse the downward trend of NDUFV2 and result in no change at the mRNA level. NDUFV2P1 is a non-transcribed pseudogene of NDUFV2. There are few studies on NDUFV2P1. Only one study has reported that schizophrenia-derived cells have increased expression of NDUFV2P1[14]. Similarly, we found that NDUFV2P1 was significantly increased and negatively correlated with NDUFV2 in schizophrenia patients. ERVWE1 is associated with schizophrenia[18]. Further analyses indicated that ERVWE1 was positively correlated with NDUFV2P1 in schizophrenia patients.

The etiology of schizophrenia remains unclear. Scientists believe that both environmental and genetic factors contribute to the risk of schizophrenia. Several studies have indicated that ERVWE1, a new pathogenic factor in schizophrenia, acts as a bridge and trigger between environmental and genetic factors. A series of environmental factors may activate the ERVWE1, a protein displaying inflammation and neurotoxicity[38]. Several studies have shown that ERVWE1 mediates neuroinflammation via the Toll-like receptor pathway in schizophrenia[19,38,39] or multiple sclerosis (MS)[40].
ERVWE1 elevated NADH dehydrogenase ubiquinone flavoprotein 2 pseudogene (NDUFB2P1) expression by suppressing cytoplasmic polyadenylation element-binding protein 1 (CPEB1). A: The mRNA levels of cytoplasmic polyadenylation element-binding protein 1 (CPEB1) after overexpression of ERVWE1 in SH-SY5Y cells using quantitative polymerase chain reaction (qPCR); B: Western blots of CPEB1 and GAPDH in SH-SY5Y cells; C: Luciferase assays of pGL3-CPEB1 promoter co-transfected with pCMV-ERVWE1 or control vector in SH-SY5Y cells; D: Luciferase assays of different
Moreover, ERVWE1 can activate inducible nitric oxide synthase activity and induce cytotoxic T lymphocyte responses[38]. Further studies on the delineation of a role for retroviruses in schizophrenia will bring new diagnosis and treatment of the devastating schizophrenia.

NDUFV2P1 is a potential predictive marker in schizophrenia by negatively regulating its parent gene, NDUFV2[14]. Based on the findings that ERVWE1 can regulate some schizophrenia-related genes[18,38], we conducted additional studies to examine the potential relationship among ERVWE1, NDUFV2P1, and NDUFV2 in schizophrenia. Previous research suggests that ERVWE1 is overexpressed in brain tissues[41]. To study the possible role of ERVWE1, we selected the SH-SY5Y cell line, a human-derived neuroblastoma cell with neuronal function and differentiation and widely used as in vitro model for neuronal studies. Our results indicated that ERVWE1 induced upregulation of NDUFV2P1 and downregulation of NDUFV2. Pseudogenes exhibit essential roles in gene expression and gene regulation. They are capable of regulating parental genes or act as a competitive endogenous RNA (ceRNA)[28] or small interfering RNA[42]. Our results showed that NDUFV2P1 acted as a ceRNA in NDUFV2 expression induced by ERVWE1.

The NDUFV2 promoter plays essential roles in its expression[11]. ERVWE1 has been proposed to regulate the activation of promoters and control the expression of some schizophrenia-related genes[18,19]. We found that NDUFV2P1 participated in the process of ERVWE1 suppressing the NDUFV2 promoter. It has been suggested that NDUFV2P1 possibly competes with NDUFV2 by interfering with its mRNA[43]. Retrocopies or processed pseudogenes, the majority of the human pseudogenes, are gene copies originating from mRNA retrotransposition. Our observation might reveal another way of regulating parental genes by their retrocopies. From these, we could conclude that ERVWE1 inhibited the NDUFV2 promoter by raising the NDUFV2P1 transcript.

Schizophrenia is linked to mitochondrial energy metabolism dysfunction. CPEB1 controls polyadenylation-induced translation of NDUFV2 mRNA and results in impaired brain-specific mitochondrial respiration[16]. In this work, we found that ERVWE1 dramatically lowered CPEB1 levels via inhibiting its promoter and that the range of -400 to 0 was necessary. We found a new regulatory mechanism that is different from the conserved mechanism of CPEB1. Interestingly, overexpression of CPEB1 led to increased protein levels of NDUFV2 with no change at the transcript level. The data are in line with another study[16]. Furthermore, CPEB1 reduced the NDUFV2P1 transcript. To date, there have been no published data on the mechanisms that regulate NDUFV2P1. Our finding might be the first report of NDUFV2P1 regulation in schizophrenia. Further studies are needed to determine precisely how CPEB1 affects pseudogene transcripts. Moreover, overexpression of CPEB1 decreased the NDUFV2P1 levels induced by ERVWE1. Simultaneously, overexpression of CPEB1 substantially elevated the expression of NDUFV2 inhibited by ERVWE1. Our results suggested that ERVWE1 regulated NDUFV2P1/NDUFV2 signaling pathway through the suppression of CPEB1.

Ample evidence indicates that impaired mitochondrial energy metabolism may be compromised in schizophrenia patients[44,45]. Complex I deficiency is among the most encountered defects of mitochondrial energy metabolism[46]. Complex I, also names NADH:ubiquinone oxidoreductase, is the largest and the most elaborate component of the mitochondrial respiratory chain. It catalyzes the oxidation of NADH, together with the transfers of two electrons to the ubiquinone. Complex I deficiency is associated with abnormalities in calcium signaling[46,47]. The possible role of calcium signaling in schizophrenia is initially presented by Jimerson et al[48]. Our previous report demonstrated that ERVWE1 can induce calcium influx in human neuroblastoma cells[49]. Therefore, we speculated that ERVWE1 might cause the decrease of complex I activity in schizophrenia. Herein, we found overexpression of ERVWE1 suppressed complex I activity. Knockdown of NDUFV2P1 significantly restored the complex I activity suppressed by ERVWE1. This observation suggested that NDUFV2P1 was required in the ERVWE1-induced complex I activity deficiency. CPEB1 is essential to promote mitochondrial energy production in neurons[15]. CPEB1 knockout cells have reduced complex I activity and mitochondria number[16,50].
Our results showed that CPEB1 overexpression could rescue complex I activity, which was inhibited by ERVWE1. As a core subunit of this complex, NDUFV2 is a crucial regulator of complex I activity, depletion and phosphorylation of this protein causes a decrease in complex I activity[51,52].

Figure 5 ERVWE1 reduced NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) expression by suppressing cytoplasmic polyadenylation element-binding protein 1 (CPEB1). A: The mRNA levels of NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) after overexpression of cytoplasmic polyadenylation element-binding protein 1 (CPEB1) in SH-SY5Y cells using quantitative polymerase chain reaction (qPCR); B: Western blots of NDUFV2 and GAPDH in SH-SY5Y cells; C: Represent the mRNA levels of NDUFV2 in SH-SY5Y cells after co-transfection with ERVWE1 and CPEB1 using qPCR; D: Western blots of ERVWE1, CPEB1, NDUFV2, and GAPDH in SH-SY5Y cells. Each bar represents the mean ± SD of three independent experiments. *P < 0.05; **P < 0.01; NS: Not significant.
ERVWE1 contributes to complex I deficits through the cytoplasmic polyadenylation element-binding protein 1 (CPEB1)/NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) pseudogene (NDUFV2P1)/NDUFV2 signaling pathway. A: Complex I activity in SH-SY5Y cells transfected with pCMV-ERVWE1 or control; B: Complex I activity in SH-SY5Y cells co-transfected with pCMV-ERVWE1 + shNDUFV2P1 or control; C: Complex I activity in SH-SY5Y cells co-transfected with pCMV-ERVWE1 + pcDNA3.1-CPEB1 or control. Each bar represents the mean ± SD of three independent experiments. *P < 0.05.

Schizophrenia mostly sets in after the age of 20. Mitochondrial energy metabolism dysfunction has been implicated in the etiology of schizophrenia in early adulthood from age 20-30[53]. Research from our and others suggests that environmental factors, including viral infection, drug stimulation, and genetic variation, can cause abnormal expression of ERVWE1. In this study, we demonstrated that increased ERVWE1 induces mitochondrial metabolism deficits, which ultimately leads to the pathogenesis of schizophrenia. In summary, ERVWE1 overexpression contributes to complex I deficits in schizophrenia via the CPEB1/NDUFV2P1/NDUFV2 signaling pathway.

CONCLUSION

In conclusion, the purpose of this study was to discover novel potential blood-based biomarkers and reveal the potential mechanisms underlying ERVWE1-mediated mitochondrial energy metabolism defects in schizophrenia. We reported that the expression of CPEB1 was lower and negatively correlated with ERVWE1 in schizophrenia blood samples. This study is to identify CPEB1 as a novel potential blood-based biomarker for schizophrenia. We also found that NDUFV2 exerted an inverse expression with NDUFV2P1 and ERVWE1 in schizophrenia. Further analyses showed a positive correlation and marked consistency between expressions of NDUFV2P1 and ERVWE1. In *vitro* experiments demonstrated that ERVWE1 mediated the downregulation of NDUFV2 by upregulating its pseudogene, NDUFV2P1. Moreover, CPEB1 was decreased after ERVWE1 overexpression and worked as a mediator of the NDUFV2P1/NDUFV2 signaling pathway induced by ERVWE1. Further studies
ERVWE1 causes CoI defect in schizophrenia

**ARTICLE HIGHLIGHTS**

**Research background**

Schizophrenia is a devastating psychiatric disorder that impairs mental and social functioning. There are several hypotheses regarding the pathogenesis of schizophrenia, including mitochondrial dysfunction. Both genetic and environmental factors contribute to the development of schizophrenia. Our previous studies showed that ERVWE1 acts as a bridge between genetic and environmental factors and plays an important role in the occurrence of schizophrenia.

**Research motivation**

Schizophrenia is a chronic disorder requiring long-term treatment and having a significant impact on patients and their families. Biomarkers and pathogenic mechanisms have become the most important undertaking for schizophrenia. In this manuscript, we focused on the relationship between mitochondrial energy metabolic deficits and ERVWE1 in schizophrenia.

**Research objectives**

This study aimed to identify novel potential blood-based biomarkers and risk factors for schizophrenia and to explore the underlying mechanisms ERVWE1-mediated mitochondrial deficits in schizophrenia.

---

**Figure 7** The proposed signaling pathways for ERVWE1 contributes to mitochondrial energy metabolism deficiency in schizophrenia. 1: ERVWE1 suppressed cytoplasmic polyadenylation element-binding protein 1 (CPEB1) mRNA and protein expression by regulating its promoter activity; 2: ERVWE1 upregulated the expression of NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2) pseudogene (NDUFV2P1) through enhancing its promoter activity. CPEB1 downregulated NDUFV2P1 expression and functioned as a mediator of NDUFV2P1 transcript induced by ERVWE1; 3: NDUFV2P1 reduced NDUFV2 promoter through increasing NDUFV2P1 transcript. Moreover, CPEB1 was involved in regulating NDUFV2P1/NDUFV2 signaling pathway triggered by ERVWE1; 4: ERVWE1 reduced complex I activity via regulating NDUFV2 and its retrocopies expression. Decreased CPEB1 was involved in the complex I activity deficiency induced by ERVWE1; 5: Deficient complex I activity might result in mitochondrial metabolic dysregulation and contributed to the development of schizophrenia.
Research methods
Quantitative polymerase chain reaction (qPCR) and enzyme-linked immunosorbent assay were used to detect differentially risk genes expression in blood samples. Statistical analyses of clinical data were performed by median analysis, Mann-Whitney U analysis, and Spearman’s rank correlation analysis. qPCR, western blotting, luciferase assay, and complex I enzyme activity microplate assay were performed to study the relationship between ERVWE1, CPEB1, NADH dehydrogenase ubiquinone flavoprotein 2 (NDUFV2), and NDUFV2 pseudogene (NDUFV2P1).

Research results
CPEB1 and NDUFV2 were deceased and negatively correlated with ERVWE1 in schizophrenia. Moreover, increased NDUFV2P1 demonstrated a significant positive correlation with ERVWE1 and negative correlation with NDUFV2 in schizophrenia. Cytological experiments demonstrated that ERVWE1 suppressed CPEB1 by suppressing its promoter activity and downregulated NDUFV2 expression through increasing NDUFV2P1 transcript. Further studies showed that CPEB1 participated in regulating the NDUFV2P1/NDUFV2 axis mediated by ERVWE1 in SH-SY5Y cells. Finally, we found that ERVWE1 regulated complex I activity through the CPEB1/NDUFV2P1/NDUFV2 signaling pathway in SH-SY5Y cells.

Research conclusions
CPEB1 and NDUFV2 were identified as novel risk factors in schizophrenia. We also proposed a novel mechanism that ERVWE1 mediated mitochondria metabolic deficits through the CPEB1/NDUFV2P1/NDUFV2 signaling pathway in schizophrenia.

Research perspectives
Further study is needed to verify if CPEB1 and NDUFV2 can be used as blood-based biomarker of schizophrenia. As a risk factor, further efforts need to be focused on better understand the role of ERVWE1, CPEB1 and NDUFV2 in the pathogenesis of schizophrenia. Although our study suggests that ERVWE1 may be a potential therapeutic target for schizophrenia, whether ERVWE1 can really benefit clinical treatment requires further research in animal experiments and clinical trials to confirm these data.

ACKNOWLEDGEMENTS
We thank Wang Y (Master of Science in Epidemiology and Biostatistics) for assistance with the biostatistical analyses.

REFERENCES
2 Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10:48-58 [PMID: 19050712 DOI: 10.1038/nrn2536]
7 Karry R, Klein E, Ben Shachar D. Mitochondrial complex I subunits expression is altered in...
Xia YR et al. ERVWE1 causes Col defect in schizophrenia


He K, Guo C, He L, Shi Y. MiRNAs of peripheral blood as the biomarker of schizophrenia. *Hereditas* 2018; 155: 9 [PMID: 28860957 DOI: 10.1186/s41065-017-0044-2]


Xia YR et al. ERVWE1 causes Col defect in schizophrenia

[PMID: 27412728 DOI: 10.1177/0706743716648290]


Retrospective Study

Psychiatric hospitalization during the two SARS-CoV-2 pandemic waves: New warnings for acute psychotic episodes and suicidal behaviors

Fabio Panariello, Sara Longobardi, Lorenzo Cellini, Diana De Ronchi, Anna Rita Atti

Abstract

BACKGROUND
The subsequent waves of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic have represented a dramatic health emergency characterized by significant consequences on mental health. Diachronic variations in the incidence rates of acute relapse of psychiatric disorders may represent significant “sentinel events” for assessing the mental health response to an unprecedented stressful event.

AIM
To investigate the variation in psychiatric hospitalization rates and differences in sociodemographic and clinical-psychopathological peculiarities at Bologna "Maggiore" General Hospital Psychiatric Ward (GHPW) between the first two waves SARS-CoV-2 pandemic and the same periods of the previous 3 years. The secondary purpose of the study was to suggest a diachronic response pathway to stress by reporting additional literature data on coping strategies.

METHODS
This observational and retrospective study collected information on admission to the GHPW at the "Maggiore" Hospital in Bologna in the index periods defined as follows: the first period between February 24, 2020 and April 30, 2020 (first epidemic wave) and the second period between October 8, 2020, and January 7, 2021 (second pandemic wave). Absolute numbers and proportion of admitted patients, their sociodemographic and clinical-psychopathological characteristics were compared with the same parameters recorded in the two same periods of the previous 3 years. No strict inclusion or exclusion criteria were provided in the data collection to collect information on all patients requiring acute psychiatric...
hospitalization.

RESULTS
During the first wave, there was a significant reduction in hospitalization rates, although there was a simultaneous increase in compulsory hospitalizations and the acute relapse of schizophrenia spectrum and other psychotic disorders. During the second wave, hospitalization rates reached those recorded during the same period of the previous 3 years, mainly due to the rise of bipolar and related disorders, depressive disorders, anxiety disorders, trauma- and stressor-related disorders and suicidal behaviors.

CONCLUSION
The coping strategies adopted during the first wave of the SARS-CoV-2 pandemic protected the vulnerable population from the general risk of clinical-psychopathological acute relapse, even if they increased the susceptibility to run into schizophrenia spectrum and other psychotic disorder relapses. In the medium-long term (as in the second pandemic wave), the same strategies do not play protective roles against the stress associated with the pandemic and social restriction measures. Indeed, during the second wave of the SARS-CoV-2 pandemic, an increase in total hospitalization rate, suicidal behaviors and the incidence rate of bipolar and related disorders, depressive disorders, anxiety disorders, trauma- and stressor-related disorders was observed.

Key Words: SARS-CoV-2; Schizophrenia spectrum and other psychotic disorders; Bipolar and related disorders; Depressive disorders; Anxiety disorders, trauma- and stressor-related disorders; Suicide behavior; Coping strategies

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: The severe acute respiratory syndrome coronavirus 2 emergency has impacted global health. This study describes psychiatric hospitalization rates and clinical characteristics during changes in the first two epidemic waves compared to the respective periods of the previous 3 years. Although the hospitalization rate decreased during the first pandemic wave, compulsory hospitalizations have increased. During the second pandemic wave, hospitalization rates returned to those recorded in the same period of the previous 3 years. Acute relapse in psychotic disorders increased during the first epidemic wave, but acute relapse in affective and anxiety disorders have increased in the second epidemic wave as well as suicidal behavior.


DOI: https://dx.doi.org/10.5498/wjp.v11.i11.1095

INTRODUCTION
In Italy, the first epidemic outbreak occurred at the end of February 2020 in Lombardy [1]. The Italian government adopted several social distancing measures and limitations of social recreational and work activities at national and regional levels to contain infection spread. These measures progressively increased until the lockdown imposed by the Prime Minister’s Decree of Ministers of March 9, 2020. Similarly, from the first decade of October 2020, in order to contain the spread of the second pandemic wave, the Italian Prime Minister issued another set of social distancing measures up to a level of limitation similar to the first lockdown. Details on social distancing measures during the first and second waves in Italy are reported in Table 1.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-related pandemic threatened the health of citizens and significantly contributed to a widespread sense of insecurity and a lessened quality of life for both individuals and...
Table 1 First and second pandemic waves and social distancing measures

<table>
<thead>
<tr>
<th>First wave (February 24, 2020 to April 30, 2020), 66 d</th>
<th>Second wave (October 8, 2020 to January 7, 2021), 92 d</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main governative measures</strong></td>
<td><strong>Main governative measures</strong></td>
</tr>
<tr>
<td>Smart working (more than 50%)</td>
<td>Smart working (8%-40%)</td>
</tr>
<tr>
<td>Distance learning (more than 75%)</td>
<td>Distance learning (75%-100%)</td>
</tr>
<tr>
<td>In and outdoor mask wearing (always)</td>
<td>In and outdoor mask wearing</td>
</tr>
<tr>
<td>Leisure activities limitations</td>
<td>Leisure activities limitations</td>
</tr>
<tr>
<td>Relatives and friends restricted visiting</td>
<td>Relatives and friends restricted visiting</td>
</tr>
<tr>
<td>Travelling limitations</td>
<td>Travelling limitations</td>
</tr>
<tr>
<td>Cinemas, theaters, museums, betting halls closure</td>
<td>Cinemas, theaters, museums, betting halls closure</td>
</tr>
<tr>
<td>Bars and restaurants closure from total closure to closure from 6.00 pm (prohibited on-site consumption)</td>
<td>Bars and restaurant closure from 6.00 pm (prohibited on-site consumption)</td>
</tr>
<tr>
<td>Night curfew (from 8:00 pm)</td>
<td>Night curfew (from 9:00 pm)</td>
</tr>
</tbody>
</table>

entire communities[2]. Such psychosocial effects on individuals (e.g., insecurity, fear, uncertainty, confusion, isolation, and stigma) and the whole community (e.g., school and workplace closures, the shutdown of social gathering places, suspension of economic activities, and uncertainty in health claims) may have contributed to threatening the mental health of individuals and the collective[3,4] and causing both new onsets of clinical pictures from adjustment disorders and relapses of psychiatric disorders from severe mental illness (SMI)[5-7]. These consequences have already been reported in the general population[8,9], vulnerable individuals[7,10], and patients with a history of psychiatric disorders[11]. Furthermore, the long-term consequences on mental health are likely to continue even after the end of the pandemic emergency, resulting, for example, from the economic burden due to the adopted social distancing measures. In epidemiological terms, it could be worse than the consequences on the mental health during the economic crisis in Italy and Greece in 2008, where a significant increase of suicidal behaviors and suicide occurred, as was well described in the literature[12,13].

The Italian mental health system is extensively community-based, and mental health care is freely available to all citizens. Besides outpatient settings, in Italy, there are nearly 318 psychiatric wards located within the general hospitals (General Hospital Psychiatric Wards [GHPWs]). Each ward contains, on average, 15 beds dedicated to treating psychiatric patients in both voluntary and compulsory ways[14]. The total national number of beds to take care of acute psychiatric clinical pictures is 3981, which means around 10 beds per 100,000 adult population[14]. Therefore, GHPWs are places of care dedicated to the hospitalization and treatment of acute psychopathological relapses that can be configured as psychiatric emergencies/urgencies. They represent a privileged observatory to survey the changes, in terms of incidence, of new-onset or acute relapse of clinical pictures related to SMIs such as those observed during the exposure to high-stress life events. For this reason, they play a crucial role during global crises that also involve mental health and represent a fundamental source of information to be collected in the early stages of a crisis, valuable both to better understand individual and collective coping strategies and to implement preventive actions and timely care[15-17].

Only a few studies[16,18] have examined and described qualitative and quantitative differences observed in the rates of urgent/emergent hospitalization within the psychiatric wards between the pre-pandemic and SARS-CoV-2 pandemic characterized by the consistent social restriction measures. Furthermore, to the best of our knowledge, no studies to date have investigated these differences during the first and second pandemic waves.

This study investigated the existence of any differences in terms of acute psychiatric relapse or new onset of psychiatric disorders observed at the GHPW of the "Maggiore" Hospital in Bologna between the first and second epidemic wave compared to the same periods of the previous 3 years by comparing the sociodemographic and clinical features.

To achieve this aim, we took into consideration: Voluntary hospitalization rates, compulsory hospitalization rates, diagnoses, and the severity of psychopathological-clinical pictures estimated by the mental health triage (MHTS)[19,20].
Furthermore, based on our results and the collection of scientific evidence that we have collected in the literature regarding the response of individuals and communities to stressful factors and their persistence over time, we propose an opinion review to depict a hypothetical model of the diachronic response of the community to unusually stressful and unprecedented events. We hypothesize that the coping strategies and their efficacy can change over time and, depending on the individual aptitude to answer to the stress, we have generated an "elastic" model that can give reason to the different psychopathological pictures observed over time, after and during a stressful event.

**MATERIALS AND METHODS**

**Participants**

This observational and retrospective study collected information on admissions to the GHPW at the "Maggiore" Hospital in Bologna in the index periods defined as follows: the first period between February 24, 2020 and April 30, 2020 (first epidemic wave) and the second period between October 8, 2020, and January 7, 2021 (second pandemic wave). Absolute numbers and the proportion of admitted patients, their sociodemographic and clinical-psychopathological characteristics were compared with the same parameters recorded in the two same periods of the previous 3 years. No strict inclusion or exclusion criteria were provided in the data collection to collect information on all patients requiring acute psychiatric hospitalization. This study followed the last version of the Declaration of Helsinki[21].

**Assessment**

For each patient hospitalized in the index periods, clinical data were collected, determining the diagnosis and the level of emergency/urgency at the time of admission based on the MHTS by the Victorian Emergency Department[19,20]. The MHTS is a triage scale aimed at identifying the severity level of psychopathological pictures to prioritize the intervention. It consists of a five-level scale: Level 1, the substantial danger of life to oneself and others; Level 2, violent, aggressive, probable danger of injury to oneself or others or with police escort; Level 3, very distressed or psychotic, liable to deteriorate, situational crisis, possible danger to oneself or others; Level 4, long-lasting semi-urgent mental health disorder, support agency present; and Level 5, long-standing non-acute mental health disorder, no support agency present. Patients were assessed using combined mental health and sociodemographic information.

Hospitalized patients were divided into emergent (Level 1 and 2) and urgent (Level 3, 4, and 5) cases. In the hospitalized population, the incidence rate of compulsory hospitalizations was calculated [according to the Italian law (Law number 180/1978), in case of acute psychopathological distress, when patients lack insight and refuse the treatments which they need, and there are no possibilities to provide outpatient treatments, psychiatric admissions against the patient’s will are possible - compulsory admissions][22]. We also collected other data such as sociodemographic data (sex, age) and the length of hospitalization. Age was used as a continuous variable. Diagnoses were assessed according to the DSM 5 through the administration of the Structured Clinical Interview for the DSM 5 Clinician Version (SCID-5-CV). The SCID-5-CV is characterized by excellent reliability and high specificity, and there is evidence for its easy use in daily clinical practice[23].

**Statistical analyses**

Continuous data for sociodemographic and clinical variables are expressed as the mean ± SD, while categorical variables are reported as percentages. The Kolmogorov-Smirnov test was performed to verify normal distributions of continuous variables. Multiple logistic regression analyses were performed by including variables that were significant for bivariate analysis. Pearson’s chi-squared test with Yates correction or the t-test for independent samples was used to compare categorical and continuous variables, respectively, between the identified subgroups. As the dependent variable, we used GHPW hospitalization during the pandemic. Odds ratio (OR) and 95% confidence interval (95%CI) were calculated to estimate the correlation between the frequencies of the hospitalization and the frequencies of independent variables. All statistical analyses were performed using the SPSS package (version 27.0, SPSS; SPSS Inc., Chicago, IL, United States) for Windows, and P < 0.05 (two-tailed) was considered statistically significant.
RESULTS

During the first index period, we observed 47 inpatients at GHPW "Maggiore" Hospital in Bologna (84 hospitalized patients during the second index period); 42.58% of the hospitalized population was female (vs 43.2% during the second lockdown). The mean age was 42 + 2 (years + SD) during the first index period (vs 43.9 + 1 during the second index period). The length of hospitalization was 8.8 + 3 d in the first index period and 8.5 + 2 d in the second index period. According to MHTS, 85.7% of the admitted patients were classified as urgency during the first index period (80.7% during the second index period). No differences were observed between the two pandemic waves and the same periods of the previous 3 years in length of hospitalization, mean age, and sex.

A reduction in total psychiatric hospitalization number during the first lockdown of 33.49% compared to the hospitalization rate in the same period of the previous 3 years was reported (47 in 2020 vs 81 in 2019, 68 in 2018, and 63 in 2017 for further details; Table 2). On the other hand, an increase in compulsory admissions was observed: 19.7% of total admissions that occurred during the first lockdown compared to 10.1% (P < 0.01) in the same period of 2019, 13.3% (P < 0.05) in 2018, and 7.2% in 2017 (P < 0.01). Consistently with these features, hospitalized patients presented a higher degree of psychopathological emergency at MHTS than in the same period of the previous 3 years (28.2% vs 13.6% [P < 0.01] in 2019, 25% [P < 0.05] in 2018, and 14.3% in 2017 [P < 0.01]) (Table 2).

During the first pandemic wave and related restriction measures, schizophrenia spectrum and other psychotic disorders diagnoses increased by 10% on average. As indicated in Table 2, 39.5% in 2020 vs 32.1% in 2019 (P < 0.05), 31.5% in 2018 (P < 0.05), 25.1% in 2017 (P < 0.01). Anxiety disorders and trauma- and stressor-related disorders (Group C in Table 2) showed an increasing trend during the first lockdown (8% in 2020 vs 5.2% in 2019, 5.9% in 2018, and 5.8% in 2017). Logistic regression analyses considering hospitalization at GHPW during the first pandemic wave and the estimated probability ratio (OR) was as follows: Male/female (OR = 1.754; P < 0.001), self-harm behavior (OR = 2.066; P < 0.001) as an independent variable (Table 3). During the second wave, hospitalization numbers increased reaching the same rate recorded during the previous 3 years (84 vs 89 in 2019, 85 in 2018, 88 in 2017) while the proportion of compulsory admissions and the level of emergency, according to the MHT, decreased to the percentages observed in the same period of the previous 3 years (Table 4).

Bipolar and related disorders, depressive disorders, anxiety disorders, and trauma- and stressor-related disorders according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM5) were those that were statistically increased significantly during the second pandemic wave compared to the same period of the three previous years. Bipolar and related disorders and depressive disorders were 32.9% of total hospitalizations vs 28.8% (P < 0.01) in 2019, 29.4% (P < 0.05) in 2018, 30.1% (P < 0.05) in 2017, in the same period of the previous 3 years. The disorders belonging to anxiety disorders and trauma- and stressor-related disorders according to the DSM5 during the second pandemic wave and related restrictive measures were 9% of total hospitalizations vs 4.2% (P < 0.01) in 2019, 3.7% (P < 0.01) in 2018 and 2.1% (P < 0.01) in 2017. We also observed a statistically significant increase in suicidal behavior during the second pandemic wave compared with the average recorded during the same period of the previous 3 years (20.22% vs 11.10%; P < 0.01). Disorders belonging to the schizophrenia spectrum were more prevalent in the second pandemic wave (10.4%) than in the same period of the previous 3 years, albeit not statistically significant. Logistic regression analyses considering hospitalization at GHPW during the second pandemic wave and the estimated probability ratio (OR) as an independent variable was as follows: male/female (OR = 1.938; P < 0.001) and suicidal behavior (OR = 2.410; P < 0.001) (Table 5).

DISCUSSION

Our results showed that, in patients with mild and moderate mental illness, during the first wave and the adoption of the first set of measures to contain the pandemic spread, some protective factors seemed to balance and outweigh the risky ones related to the pandemic and quarantine. The total number of hospitalizations at the GHPW decreased with a relative increase in the pictures listed as a psychiatric emergency according to the MHT (codes 1 + 2). On the other hand, the acute relapse of acute
Table 2 Absolute numbers of total admissions, proportion of emergencies, urgencies and compulsory admissions occurred during the first pandemic wave and in the same period of previous 3 years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>63</td>
<td>85.7%</td>
<td>14.3% (&lt; 0.01)</td>
<td>7.2% (&lt; 0.01)</td>
<td>25.1% (&lt; 0.01)</td>
<td>28.2% (&gt; 0.05)</td>
<td>5.8% (&gt; 0.05)</td>
</tr>
<tr>
<td>2018</td>
<td>68</td>
<td>75.0 %</td>
<td>25% (&lt; 0.05)</td>
<td>13.3% (&lt; 0.05)</td>
<td>31.5% (&lt; 0.05)</td>
<td>28.0% (&gt; 0.05)</td>
<td>5.9% (&gt; 0.05)</td>
</tr>
<tr>
<td>2019</td>
<td>81</td>
<td>86.4%</td>
<td>13.6% (&lt; 0.01)</td>
<td>10.1% (&lt; 0.05)</td>
<td>32.1% (&lt; 0.05)</td>
<td>27.9% (&gt; 0.05)</td>
<td>5.2% (&gt; 0.05)</td>
</tr>
<tr>
<td>2020</td>
<td>47</td>
<td>71.8%</td>
<td>28.2%</td>
<td>19.7%</td>
<td>39.5%</td>
<td>28.6%</td>
<td>8.0%</td>
</tr>
</tbody>
</table>

Emergencies defined for codes 1 + 2, Urgencies defined for codes 3 + 4 + 5 at Mental health Triage. Diagnosis DSM-5 295.9-295.40-295.70-297.1-297.3-298.8-298.9: Schizophrenia Spectrum and Other Psychotic Disorders. Diagnosis DSM-5 296.2-296.3-296.9-311-296.XX: Bipolar and Related Disorders and Depressive Disorders. Diagnosis DSM-5 309.81-308.3-309.0-309.24-309.28-309.3-309.4-309.9-309.89: Anxiety Disorders and Trauma- and Stressor-Related Disorders.

Table 3 Multivariate logistic regression analyses: odds ratio, and 95% confidence interval for hospitalization during the first pandemic wave

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>P value</th>
<th>OR</th>
<th>95%CI for EXP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/female</td>
<td>&lt; 0.001</td>
<td>1.754</td>
<td>1.331–2.312</td>
</tr>
<tr>
<td>Self-harm behaviour</td>
<td>&lt; 0.001</td>
<td>2.066</td>
<td>1.477–2.890</td>
</tr>
</tbody>
</table>

CI: Confidence interval; OR: Odds ratio.

Table 4 Absolute numbers of total admissions, proportion of urgencies, emergencies and compulsory admissions occurred during the second pandemic wave and in the same period of the previous 3 yr

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>88</td>
<td>80.7%</td>
<td>19.3% (&gt; 0.05)</td>
<td>9.1% (&gt; 0.05)</td>
<td>35.9% (&gt; 0.05)</td>
<td>30.1% (&gt; 0.05)</td>
<td>4.1% (&lt; 0.01)</td>
</tr>
<tr>
<td>2018</td>
<td>85</td>
<td>81.2 %</td>
<td>18.8% (&gt; 0.05)</td>
<td>10.7% (&gt; 0.05)</td>
<td>35.6% (&gt; 0.05)</td>
<td>29.4% (&lt; 0.05)</td>
<td>3.7% (&lt; 0.01)</td>
</tr>
<tr>
<td>2019</td>
<td>89</td>
<td>81.5%</td>
<td>18.5% (&gt; 0.05)</td>
<td>9.9% (&gt; 0.05)</td>
<td>33.7% (&gt; 0.05)</td>
<td>28.8% (&lt; 0.01)</td>
<td>4.2% (&lt; 0.01)</td>
</tr>
<tr>
<td>2020/21</td>
<td>84</td>
<td>80.7%</td>
<td>19.3%</td>
<td>10.4%</td>
<td>36.1%</td>
<td>32.9%</td>
<td>9.0%</td>
</tr>
</tbody>
</table>

Emergencies defined for codes 1 + 2, Urgencies defined for codes 3 + 4 + 5 at Mental health Triage. Diagnosis DSM-5 295.9-295.40-295.70-297.1-297.3-298.8-298.9: Schizophrenia Spectrum and Other Psychotic Disorders. Diagnosis DSM-5 296.2-296.3-296.9-311-296.XX: Bipolar and Related Disorders and Depressive Disorders. Diagnosis DSM-5 309.81-308.3-309.0-309.24-309.28-309.3-309.4-309.9-309.89: Anxiety Disorders and Trauma- and Stressor-Related Disorders.

Psychiatric clinical pictures belonging to schizophrenic spectrum disorders has undergone a statistically significant increase during the first pandemic wave and the first social restriction measures. In the literature, other authors have already described the increase in incidence rates of psychotic clinical pictures, and this phenomenon may probably have a generalizable value in all the countries most severely affected by the SARS-CoV-2-related pandemic and which have been subject to measures to contain the spread of the virus[24-26]. These preliminary data were consistent with results from other countries, such as China, where the pandemic would have originated and where the highest incidence of the first psychotic episode was reported in the months following the origin of the pandemic[27].
It can be argued that the highly stressful scenario deriving from the fear of contagion, the confinement and the reduction of freedom imposed by the restrictive measures, the fear of the future evolution of the pandemic situation, the fear of the potential economic crisis could have played a significant role in the new-onset and the exacerbation of pathologies belonging to the schizophrenia spectrum according to international literature.

Although the relationship between the pressure exerted by highly stressful events during adult life and the onset of psychosis has been poorly studied, the results of a large meta-analysis suggest that the risk of psychosis onset is three times higher in subjects exposed to highly stressful events[28].

The data relating to the increase in the onset of new psychoses recorded in our data collection during the first pandemic wave was consistent with Pollice et al[29] in 2012. They observed the same boost in first psychosis episodes in connection with the earthquake in L’Aquila in 2009, an event whose traumatic extent can be compared, in some ways, to the traumatic impact of the related SARS-CoV-2 pandemic and to the consequent deterioration of social life and which share the same cultural scenario (both the stress events and the related results happened in Italy)[29].

We hypothesized that the coping strategy could be called into question to understand better the phenomenon from a psychological point of view in the biopsychosocial perspective.

Coping is identified as a complex of strategies adopted to manage stress better and reduce the emotional impact of stress events. They have been described as emotion-focused coping strategies, problem-oriented coping strategies and avoidance-focused coping strategies. Problem-focused coping strategies are termed “positive coping” as they relate to better outcomes[30]. Patients with more severe symptoms of schizophrenia are characterized by a reduced use of positive coping strategies[30]. Furthermore, the severity of psychotic symptoms is inversely related to positive coping, which is related to better functional outcomes[31,32]. Overall, results from the scientific literature suggest that schizophrenic patients rely on maladaptive coping strategies or have the potential to exacerbate distress.

The observation, already reported, of a global decline in the number of hospitalizations at "Maggiore" Bologna Hospital's GHPW also may be analyzed considering the theory of coping strategies. If several social factors such as the shared experience at the national level of a sense of cohesion, trust and social compassion perceived by the whole nation and the increase of the importance of informal social networks, it can also be argued that most individuals have adopted positive coping strategies. It could mean that individuals with a higher vulnerability to psychotic disorders cannot adopt on their own problem-oriented coping strategies. From the biological point of view of the biopsychosocial model of psychiatric disorders, it can be speculated that severe psychiatric pictures, including schizophrenia, bipolar disorders, and depressive disorders for which a greater etiological weight is recognized to the biological factor in the biopsychosocial model, could have been triggered by the direct neuroinvasiveness of SARS-CoV-2 or by the consequences of consequences of peripheral infection or by the subsequent inflammatory state on the central nervous system (CNS). Our group proposed a generating hypothesis that SARS-CoV-2 would demonstrate a neuroinvasive and neurotropic capacity through different anatomical pathways[33] through the angiotensin-converting enzyme-2 receptor. Possible speculation regarding the observed data is that SARS-CoV-2, through the colonization of the CNS, may act as an etiological cofactor in the exacerbation or in the onset of psychopathological pictures that recognize a more significant causal load of the biological component in the biopsychosocial model.

Further evidence of the present study suggests that in the medium to long term, such as during the adoption of restrictive measures in the second pandemic wave, there was an overall increase in the incidence of acute psychopathological relapse.

### Table 5 Multivariate logistic regression analyses: odds ratio, and 95% confidence interval for hospitalization during the first pandemic wave

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>P value</th>
<th>OR</th>
<th>95%CI for EXP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/female</td>
<td>&lt; 0.001</td>
<td>1.938</td>
<td>1.357–2.766</td>
</tr>
<tr>
<td>Self-harm behaviour</td>
<td>&lt; 0.001</td>
<td>2.410</td>
<td>1.731–3.354</td>
</tr>
</tbody>
</table>

CI: Confidence interval; OR: Odds ratio.
Indeed, the hospitalization rate reverted to the pre-existing incidence recorded in the same periods of the previous 3 years. This feature was mainly due to the rise of Mood (bipolar and related disorders and depressive disorders) and anxiety disorders. This result suggests that positive coping strategies, in the case of the ongoing pandemic, have proven to be effective in the short-term prevention of acute mood and anxiety disorders but have lost this in the long term. Other notable data is the increase in the suicidal behavior rate. This finding can be, in part, explained considering the increase of depressive disorders incidence.

On the other hand, the impact of the stress on suicidal behavior can also be considered. Indeed, as highlighted by Monica Starkman, in a brief report on the suicide attempt rate after the terrorist attack in September 11, 2001, the stress related to that unprecedented attack contributed to the statistically significant increase in the number of suicidal behaviors during the 2 years following the September 11, 2001[34]. Furthermore, in our study, suicidal behaviors are also strong predictors of psychiatric hospitalization during the second pandemic wave at multiple regression analysis.

CONCLUSION

The findings in this study showed that the higher incidence of hospitalizations for mood disorders and suicidal behavior must also be evaluated considering the growing and looming economic crisis due to the restrictive policy measures influencing the indicators of social well-being. Deaths from suicidal behavior increase dramatically during economic crises. In 2016, as Merzagora et al[35], in Italy, suicides increased between 2011 and 2016, during and immediately after the severe economic crisis that began in 2008[35]. Our preliminary data highlight the requirement to adopt preventive social and health policies to contain a potential "psychopathological pandemic" following the SARS-CoV-2 pandemic and invite the scientific society to pay attention to the problem.

By combining our clinical observations with the epistemological theories present in the literature[36-38], we propose an opinion hypothesis based on applying, to the phenomenon of psychological stress and coping strategies, the same laws of physics as the law by Hooke. It states that, in a mechanical system, the elastic response to a force F applied to an object is proportional to the variation of variable length (ΔL) multiplied by a specific constant of the material of which the object is composed [F = ΔL K]. In biological systems, as reported by Kültz and Somero[38], organisms represent homeostatic systems in continuous dynamic equilibrium. Stress is constantly fluctuating around the norm throughout the life history of any organism. This fluctuation can push psychological and physiological set points out of homeostasis H.

The extent of the deviation (increase or decrease in value) of the most critical (limiting) variables (ΔHc) from the homeostatic set point (Hc) determines the behavior of the system. Therefore, for biological systems, Hooke’s law can be defined as [F = ΔHc k], where k is a constant describing the phenotype of the biological system at the moment of exposure to force F and Hc represents the homeostatic norm H of most critical (limiting) physiological variables c. Our opinion is that this construct is also valid for psychological systems where homeostasis is ensured by multiple factors, including coping and stress response strategies. The challenge is to identify which biopsychological variables might be the most critical and limiting ones. These variables will depend on stress type and severity and the individual's state at the time of the stressful event. Within this theoretical hypothesis, the time factor would also play a key role. If a perturbing event acts for a time interval Δt, some psychological homeostatic set points could alter at different time points, depending on the "elastic capacity" of the same set points and the coping strategies adopted. In our case, the set point of the affective dimension would have "broken" later than that of the psychotic dimension.

Limits and perspectives of the study

The current study had several limitations. Meanwhile, it is a monocentric study and describes what happened in one of the Italian GHPWs. As such, it does not claim to draw ecumenical conclusions. Our data are consistent with what was described by other authors during the first pandemic wave. However, our study represents the first that follows diachronically, during the first and second pandemic waves, the trend of hospitalizations for acute psychiatric relapse in response to the Trauma of the pandemic and the consequent restrictive measures. The last "take-home messages" that deserve to be taken into consideration and on which to calibrate the subsequent
research trajectories are the following: The severe mental illnesses have not suffered, contrary to the moderate-mild mental disorders, a contraction in terms of acute relapse, during the first SARS-CoV-2 pandemic wave. The "advantage" of moderate/mild disorders is lost during the second pandemic because the protective factors probably no longer balance the precipitating factors. The findings from the first pandemic wave suggest that mild/moderate psychopathological clinical pictures can be successfully treated in settings other than GHPWs. This key point indicates that: It is possible to reduce hospitalization rates by enhancing and customizing the offers of the community mental health care settings; it is confirmed that traumatic experiences are important causal factors of psychotic relapse in the short-term period and mood disorders relapse in the long-term period; and the suicide risk could, with every reasonable probability, become the real post-pandemic emergency in light of the emerging and consequent economic-financial crisis.

ARTICLE HIGHLIGHTS

Research background
Significant consequences on mental health characterize unprecedented events like the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Therefore, variations in psychiatric hospitalization rates represent significant ‘sentinel events’ for assessing the mental health response to stress. Responses to the stress may lead to new-onset or relapses of severe mental illness and suicidal behavior rates.

Research motivation
To reduce the further consequences on the mental health of the SARS-CoV-2 Pandemic by describing the repercussions of the first two waves of the pandemic on the psychiatric new-onset or relapse pictures rates.

Research objectives
The present study investigates any differences in terms of acute psychiatric relapse or new onset of psychiatric disorders observed at the General Hospital Psychiatric Ward (GHPW) at "Maggiore" Hospital in Bologna between the first and second epidemic.

Research methods
We took into consideration: Voluntary hospitalization rates, compulsory hospitalization rates, diagnoses, and the severity of psychopathological-clinical pictures estimated by the mental health triage and Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition interview. Multiple logistic regression analyses were performed to estimate the correlation between the frequencies of the hospitalization and the frequencies of the independent variables.

Research results
The total admissions rate at GHPW decreased during the first SARS-CoV-2 pandemic wave compared to the same period throughout the 3 previous years. The compulsory admission rate and acute psychotic relapse rate increased during the first SARS-CoV-2 pandemic wave in comparison with the 3 previous years. During the second SARS-CoV-2 pandemic wave the total psychiatric admission rate reached the same ratio recorded during the same period of the previous 3 years. Suicidal behaviors, depressive disorders, bipolar disorders, anxiety spectrum disorders and trauma- and stressor-related disorders relapse arose among hospitalized patients during the second SARS-CoV-2 pandemic wave compared with the 3 previous years.

Research conclusions
We hypothesize that problem-focused coping strategies may be protective for the risk of acute psychopathological relapse in the short term but not in the medium-long term.

Research perspectives
The suicidal rate could arise in the future, after the first two pandemic waves therefore may be important to survey it and to adopt preventive strategies.


Panariello F et al. Acute psychiatric hospitalization during two pandemic waves
Observational Study

Self-compassion and resilience mediate the relationship between childhood exposure to domestic violence and posttraumatic growth/stress disorder during COVID-19 pandemic

Xin-Li Chi, Qiao-Min Huang, Xiao-Feng Liu, Liu-Yue Huang, Meng-Jian Hu, Zhi-Jing Chen, Can Jiao, Brendon Stubbs, M Mahbub Hossain, Li-Ye Zou

ORCID number: Xin-Li Chi 0000-0002-3901-8107; Qiao-Min Huang 0000-0003-0870-3125; Xiao-Feng Liu 0000-0002-8971-6624; Liu-Yue Huang 0000-0001-8293-2220; Meng-Jian Hu 0000-0003-0429-845X; Zhi-Jing Chen 0000-0003-2438-120X; Can Jiao 0000-0002-8278-0834; Brendon Stubbs 0000-0001-7387-3791; M Mahbub Hossain 0000-0002-3223-0949; Li-Ye Zou 0000-0001-6411-5710.

Author contributions: Chi XL drafted the manuscript, completed the literature searches, and revised the final version of the manuscript; Huang QM drafted the manuscript and performed the statistical analyses; Liu XF, Huang LY, Jiao C, Stubbs B, and Hossain MM drafted the manuscript; Hu MJ and Chen ZJ revised the manuscript and performed the statistical analyses; Zou LY contributed to the study design.

Institutional review board statement: The study was reviewed and approved by the Medical Ethics Committee of Medical Department of Shenzhen University (Approval No. 20200005).

Informed consent statement: Xin-Li Chi, Liu-Yue Huang, Meng-Jian Hu, Zhi-Jing Chen, Can Jiao, Institute of KEEP Collaborative Innovation, School of Psychology, Shenzhen University, Shenzhen 518060, Guangdong Province, China
Qiao-Min Huang, Guangdong Polytenech, Foshan 528041, Guangdong Province, China
Xiao-Feng Liu, Longhua No. 2 Experimental School Affiliated Hongchuang Kindergarten, Shenzhen 518100, Guangdong Province, China
Brendon Stubbs, Physiotherapy Department, South London & Maudsley NHS Fdn Trust, London WC2R 2LS, United Kingdom
Brendon Stubbs, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London WC2R 2LS, United Kingdom
M Mahbub Hossain, School of Public Health, Texas A&M University, Texas, TX 77843, United States
Li-Ye Zou, Exercise Psychophysiology Laboratory, Institute of KEEP Collaborative Innovation, School of Psychology, Shenzhen University, Shenzhen 518060, Guangdong Province, China

Corresponding author: Li-Ye Zou, PhD, Professor, Exercise Psychophysiology Laboratory, Institute of KEEP Collaborative Innovation, School of Psychology, Shenzhen University, No. 3688 Nanhai Avenue, Nanshan District, Shenzhen 518060, Guangdong Province, China. liyezou123@gmail.com

Abstract

BACKGROUND

Studies have indicated that childhood exposure to domestic violence is a common factor in posttraumatic growth (PTG) and posttraumatic stress disorder (PTSD), but it is unclear whether PTG and PTSD share a common/different underlying mechanism.

AIM

To explore the common/different underlying mechanism of PTG and PTSD.
INTRODUCTION

Domestic violence has always been a serious social problem with a high incidence rate. Data from the World Health Organization[1] indicate that one-third of women worldwide have experienced physical violence, with many of these incidents happening in front of children and adolescents, therefore making it even more disturbing. Previous studies have consistently demonstrated that children who were exposed to domestic violence are significantly associated with an increased risk of psychological problems such as posttraumatic stress disorder (PTSD)[2-4].

Although PTSD symptoms in posttraumatic (PT) adolescent groups have been reported, many researchers found that PT individuals may also show positive psychological changes or posttraumatic growth (PTG)[5-7], which is defined as the positive psychological changes experienced as a result of the struggle with a traumatic event. PTSD and PTG following trauma are two completely different psychological constructs but may coexist in traumatized individuals, especially individuals who were exposed to domestic violence[8,9]. However, few studies have simultaneously
examined the mechanisms by which domestic violence affects PTG and PTSD. Therefore, it is unclear whether a common mechanism underlies the association between domestic violence and PTG/PTSD. To better understand the process of reactions to stressful events among young adults and to develop prevention or intervention programs, this study aimed to identify factors that mediate the association between physical domestic violence and PTG and PTSD. This study further examined and compared the mechanisms underlying these associations. It may inform future mental health interventions for vulnerable individuals. In addition, this study has critical significance as it was conducted amidst the coronavirus disease 2019 (COVID-19) outbreak, thus it may provide insights into the development of PTG and PTSD in the context of a global pandemic. Such insights emphasize unique psychosocial dynamics, especially among those who experienced trauma in the past and those currently experiencing global psychosocial stressors attributable to the pandemic.

**Potential mediating role of self-compassion and resilience**

The Kumpfer's resilience framework believes that individuals who encountered the traumatic event would gather their own resources to cope with the stress based on their cultural context. The more risk factors (such as childhood exposure to domestic violence) or the fewer protective factors in the cultural context there are, the higher probability individuals perform maladaptively, and *vice versa*. Hence, the stress would be alleviated or strengthened by the cultural context. Then, interactions between individuals and the environment would happen to handle stress, which may lead to the changes of the internal resources and traits associated with resilience. Consequently, the process of resilience influenced by such traits would bring about the adaption or maladaptation which may reduce the possibility of developing PTG and PTSD[10].

Self-compassion is the ability to treat oneself with the same kindness and compassion as one would treat others in the same situation[11], which was regarded as a trait associated with resilience. Existing literature indicated that self-compassion is an important predictor of resilience and promoting self-compassion may facilitate individual resilience[12-14]. This suggests that individuals with high self-compassion can exhibit a higher level of resilience and better adapt and recover when facing stress or difficulties. In contrast, individuals with a low level of self-compassion are inclined to criticize or reject themselves, which often form negative self-beliefs and low self-esteem, and feel unworthy of love[15]. These eventually lead to a low level of resilience.

**Present study**

Based on the Kumpfer’s resilience framework and empirical evidence, the present study hypothesized that the effect of childhood exposure to domestic violence on PTG and PTSD is significantly mediated by self-compassion (hypothesis 1) and resilience (hypothesis 2), respectively (1-step indirect path). It was also hypothesized that childhood exposure to domestic violence may affect PTG and PTSD via self-compassion to resilience (hypothesis 3) (2-step indirect path). Specifically, suffering from the trauma of COVID-19, children and adolescents who were exposed to domestic violence regarded as a risk factor may show a lower level of self-compassion as one would treat others in the same situation, which was regarded as a trait associated with resilience. Existing literature indicated that self-compassion is an important predictor of resilience and promoting self-compassion may facilitate individual resilience[12-14]. This suggests that individuals with high self-compassion can exhibit a higher level of resilience and better adapt and recover when facing stress or difficulties. In contrast, individuals with a low level of self-compassion are inclined to criticize or reject themselves, which often form negative self-beliefs and low self-esteem, and feel unworthy of love[15]. These eventually lead to a low level of resilience.

**MATERIALS AND METHODS**

**Study participants**

Between February 12 and 17, 2020, college students who have been or have not been exposed to domestic violence from more than 180 universities in China were recruited to participate in this study. Prior to filling out several self-reported questionnaires, all volunteers signed an online consent form which was opaque about the aim of the study to control the hawthorn effect. Overall, 2500 students from 29 provinces and cities of China were invited on the basis of the following inclusion criteria: Being at least 18 years of age and fluent in Chinese; however, only 2126 students completed the questionnaire (85.04% response rate). Of these 2126 students, 88 were excluded due to missing or incomplete responses. Finally, data of the remaining 2038 students (755 males and 1283 females, mean age: 20.56 ± 1.90) was analyzed. To ensure the adequate power, we calculated the sample size using the sampling formula \[ N = \left\lceil \frac{Z^2p(1-p)}{d^2} \right\rceil \]
[16], which determined that at $a = 0.05$, $P = 0.3$, and $d = 0.03$, the sample size needed was approximately 896 individuals. The sample size of 2038 in the present study should be sufficient.

**Procedure**

In an attempt to control the pandemic, the government had imposed a nationwide lockdown. Hence, this cross-sectional survey was conducted online, which was also safer and convenient. For 6 d, students were invited to participate in the survey via Tencent’s QQ, WeChat, Weibo, and college-related websites, such as university association websites and bulletin board system forums. Participants who clicked the survey link would be automatically directed to the questionnaire website page. It would not be proceeded to the next page unless participants completed all the items on the current page so that there was no missing data for those who completed the questionnaire. Participants received 10 RMB via online payment (equivalent to USD 1.5 at the time) on completion of the survey, which took approximately 20 min. Recruitment and data collection procedures were approved by the Human Research Ethics Committee (No: 2020005) of Shenzhen University.

**Measurement**

**Dependent variables**: The independent variables were PTSD and PTG. PTSD symptoms were assessed using the abbreviated PTSD Checklist-Civilian version (PCL-C)[17,18]. It consists of six items that correspond to six different symptoms. The respondents rated the frequencies of symptoms over the past 1 mo on a 5-point Likert scale that ranges from 1 (not at all) to 5 (extremely). To assess participants’ PTSD related to the COVID-19 pandemic, we revised “stressful experience” in this scale to “COVID-19 pandemic” (e.g., “Feeling very upset when something reminded you of COVID-19 pandemic from the past”). Higher total scores indicate more severe PTSD symptoms. The PCL-C possesses good psychometric with a sensitivity of 0.78 and specificity of 0.71[17]. In the present study, the results indicated a good internal reliability (Cronbach’s $\alpha = 0.81$) of this scale.

PTG was assessed using the 21-item Posttraumatic Growth Inventory (PTGI)[19]. It includes five dimensions: Relating to others (seven items), new possibilities (five items), personal strength (four items), spiritual change (two items), and appreciation of life (three items). Previous studies conducted by two different research groups indicated that few persons, including Chinese college students, expressed religious beliefs and spiritual change (two items); therefore, these items were deemed not applicable to the local culture and were finally excluded from this study[20,21]. All items were rated on a 6-point scale ranging from 0 (no change) to 5 (complete change). Sum scores were obtained, ranging from 0 to 95, with higher scores representing higher levels of PTG. The inventory in prior studies was found to have good psychometric properties in the Chinese context with an internal reliability of 0.88[22,24]. In this study, the internal reliability of the inventory was 0.95 and the internal reliabilities for the 4 subscales were 0.88, 0.86, 0.81, and 0.82, respectively.

**Independent variable**: The independent variable was childhood exposure to domestic violence, which was selected from the Chinese version of the revised Adverse Childhood Experiences Question. More specifically, this scale was used to measure domestic violence experienced in the first 18 years of life[25,26]. It contains 4 items and each item was rated on a 5-point scale ranging from 1 to 5 (1 = never, 2 = once or twice, 3 = sometimes, 4 = often, and 5 = very often). The total scores were used, with higher scores representing a higher frequency of exposure to domestic violence. The Chinese version of the revised Adverse Childhood Experiences Question possesses good psychometric properties with an internal reliability of 0.83[26]. In this study, the internal reliability of the scale was good (Cronbach’s $\alpha = 0.84$).

**Mediation variables**: The mediation variables were self-compassion and resilience. Self-compassion was assessed using the Self-compassion Scale, which has 26 items on 6 dimensions: Self-kindness (5 items), self-judgment (5 items), common humanity (4 items), isolation (4 items), mindfulness (4 items), and over-identification (4 items). All items were rated on a 5-point scale ranging from 1 (almost never) to 5 (almost always). Self-judgment, isolation, and over-identification should be scored in reverse, then a grand mean was computed based on the mean of all six subscales, with higher scores reflecting a greater level of self-compassion. This scale was adapted in Chinese, with good reliability (Cronbach’s $\alpha = 0.96$) and validity reported among Chinese adolescents and college students[27-29]. In this study, the internal reliability of the scale was good (Cronbach’s $\alpha = 0.87$), and the internal reliabilities for the 6 subscales
were 0.81, 0.65, 0.68, 0.75, 0.81, and 0.66, respectively, indicating an acceptable reliability.

Resilience was assessed using a short version of the Connor-Davidson Resilience Scale (CD-RISC). It reflects the ability to tolerate experiences, such as change, personal problems, illness, pressure, failure, and painful feeling[30]. Participants responded to 10 items on a 5-point Likert scale (0 = not true at all to 4 = true nearly all of the time), with total scores ranging from 0 to 40 (higher points indicate greater resilience capacity). This original scale in Chinese was adapted and reported with good reliability (Cronbach’s α = 0.88) and validity[31,32]. In this study, the results indicated an excellent internal consistency (Cronbach’s α = 0.92).

**Statistical analysis**

First, the mean ± SD or distribution was calculated for all variables. Second, partial correlations were performed to examine the association between domestic violence and PTG, PTSD, self-compassion, and resilience while controlling the age, gender, subjective social economic status, and family structure. Third, to examine the mediating role of self-compassion and resilience in the relationship between domestic violence and PTG and PTSD, a series of structural equation models were built. First, we established a direct model from domestic violence to PTG and PTSD and relational paths among PTG and PTSD were then added. Second, based on the direct model, an indirect model with the mediators (self-compassion and resilience) inserted between domestic violence and PTG and PTSD was built. In this indirect model, the predictive paths from self-compassion to resilience were added. The specific mediation pathways were presented as below: domestic violence → self-compassion → resilience → PTG/PTSD; domestic violence → resilience → PTG/PTSD; domestic violence → self-compassion → PTG/PTSD. Finally, a parsimonious model constraining non-significant paths to zero was built to further identify the indirect effect of domestic violence and PTG/PTSD via self-compassion and resilience. In the process of path analysis, all aforementioned variables were treated as observed variables in terms of their total scores, with controlling the cofounders as above. For all pathways, standardized direct, specific indirect, total indirect, and total effects were estimated. Standardized values were reported for all estimations. The goodness of fit was assessed using the following fit indices: Comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Thresholds used were as follows: For CFI and TLI, excellent fit > 0.95 and moderate fit > 0.90; for RMSEA and SRMR, excellent fit < 0.05 and moderate fit < 0.08.

**RESULTS**

**Descriptive statistics and correlations of variables**

The descriptions of cofounders are as follows: Gender, 755 males and 1286 females; age, 20.56 (SD = 1.90) years; subjective social economic status, 4.85 (SD = 1.38); family structure, 1859 intact families and 179 non-intact families. Table 1 shows the mean ± SD for all variables, as well as results from partial correlation analyses between every two variables (domestic violence, self-compassion, resilience, PTG, and PTSD). The mean variable values were as follows: PTG, 60.97 (SD = 16.95); PTSD, 11.79 (SD = 4.25); resilience, 35.45 (SD = 6.62); self-compassion, 84.75 (SD = 11.88); and domestic violence, 4.76 (SD = 1.72). Further, the results of partial correlations while controlling the age, gender, subjective social economic status, and family structure, showed that PTG was significantly positively correlated with self-compassion and resilience (P < 0.001), but significantly negatively correlated with domestic violence (P < 0.001). Moreover, PTSD had a positive association with domestic violence (P < 0.001), but negatively correlated with resilience and self-compassion (P < 0.001). Similarly, resilience was significantly positively correlated with self-compassion (P < 0.001), whereas resilience and self-compassion were negatively correlated with domestic violence (P < 0.001).

**Indirect effect analysis**

First, the results indicated that domestic violence had a direct effect on PTG and PTSD. This model fits the data better (χ² = 424.026, df = 49, P = 0.000, CFI = 0.966, TLI = 0.951; RMSEA = 0.061, 90%CI (0.056, 0.067), SRMR = 0.024), while controlling the age, gender, subjective social economic status, and family structure. Results from the path
Table 1 Descriptive statistics and correlation analyses among posttraumatic growth, posttraumatic stress disorder domestic violence, self-compassion, and resilience

<table>
<thead>
<tr>
<th>Variable</th>
<th>mean ± SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTG</td>
<td>60.97 ± 16.95</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD</td>
<td>11.79 ± 4.25</td>
<td>0.02</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resilience</td>
<td>35.45 ± 6.62</td>
<td>0.32</td>
<td>-0.28</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-compassion</td>
<td>84.75 ± 11.88</td>
<td>0.33</td>
<td>-0.34</td>
<td>0.56</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Domestic violence</td>
<td>4.77 ± 1.72</td>
<td>-0.10</td>
<td>0.09</td>
<td>-0.08</td>
<td>-0.56</td>
<td>1</td>
</tr>
</tbody>
</table>

*P < 0.05.
**P < 0.01.
***P < 0.001.

analysis showed that domestic violence was negatively associated with PTG (β = -0.100, 95%CI: 0.147 to -0.052, P < 0.001) and positively associated with PTSD (β = 0.083, 95%CI: 0.033 to 0.135, P < 0.005). This indicated that domestic violence was a common factor in PTG and PTSD.

Based on the results of the direct effect model, an indirect effect model was established controlling the cofounders as above. The results indicated a complete model with fit indices as follows: χ² = 891.522, df = 128, P < 0.001, CFI = 0.960, TLI = 0.943; RMSEA = 0.054, 90%CI (0.051-0.057), SRMR = 0.045. The path analysis showed that the paths between domestic violence and resilience were non-significant, and the remaining paths were statistically significant. Next, we removed the non-significant path above; constraining these paths to zero caused no degradation in fit. Using these procedures, a parsimonious model was built (Figure 1) controlling the cofounders as above. The result also indicated a good model with the following fit indices: χ² = 891.564, df = 129, P = 0.000, CFI = 0.960, TLI = 0.943; RMSEA = 0.054, 90%CI (0.051 to 0.057), SRMR = 0.045. The path analysis showed that domestic violence was significantly negatively related to PTG via a 1-step indirect path of self-compassion (β = -0.023, 95%CI: -0.044 to -0.007, P < 0.01) and by one 2-step indirect path of self-compassion to resilience (β = -0.008, 95%CI: -0.018 to -0.002, P < 0.05). Domestic violence was significantly positively related to PTSD via a 1-step indirect path of self-compassion (β = 0.008, 95%CI: 0.002 to 0.021, P < 0.07) and by one 2-step indirect path of self-compassion to resilience (β = 0.013, 95%CI: 0.004 to 0.024, P < 0.01). The detailed information is listed in Table 2. These findings suggest that self-compassion to resilience mediates the relationship between domestic violence and PTG and PTSD and that the mechanisms underlying the association between domestic violence and PTG and PTSD are similar.

DISCUSSION

This study examined the mechanisms underlying the relationship between childhood exposure to domestic violence and PTG and PTSD among Chinese college students during the COVID-19 pandemic. The findings of this study indicated that childhood exposure to domestic violence was significantly associated with PTG and PTSD[4,20]. Furthermore, these associations were significantly mediated by self-compassion (1-step indirect path) and from self-compassion to resilience (2-step indirect path). These results suggested that there were similar indirect paths from childhood exposure to domestic violence to PTSD and PTG. Possible explanations for these results are elaborated below.

First, childhood exposure to domestic violence was directly associated with PTG and PTSD, which further supports Kumpfer’s resilience framework. According to the theory, when individuals are under stress, those with more protective factors can easier get through the crisis, but those with more risk factors may become maladaptive.

Furthermore, this study found that childhood exposure to domestic violence had a positive association with PTSD and a negative association with PTG via a 1-step indirect path of self-compassion (supported hypothesis 1). Individuals with high self-compassion are more likely to pay attention to the positive aspects of the struggle with...
Table 2 Bootstrapping indirect effect and 95% confidence interval for multiple mediation model

<table>
<thead>
<tr>
<th>Indirect effect</th>
<th>Estimated effect</th>
<th>95% CI</th>
<th>Relative mediation effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent variable: PTG</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total indirect</td>
<td>-0.032&lt;sup&gt;c&lt;/sup&gt;</td>
<td>(-0.055, -0.010)</td>
<td>32.99%</td>
</tr>
<tr>
<td>Specific indirect 1</td>
<td>0.000</td>
<td>(-0.006, 0.004)</td>
<td>0%</td>
</tr>
<tr>
<td>Specific indirect 2</td>
<td>-0.023&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(-0.044, -0.007)</td>
<td>23.71%</td>
</tr>
<tr>
<td>Specific indirect 3</td>
<td>-0.008&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(-0.018, -0.002)</td>
<td>8.25%</td>
</tr>
<tr>
<td><strong>Dependent variable: PTSD</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total indirect</td>
<td>0.022&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(0.007, 0.039)</td>
<td>28.95%</td>
</tr>
<tr>
<td>Specific indirect 1</td>
<td>0.001</td>
<td>(-0.007, 0.008)</td>
<td>1.32%</td>
</tr>
<tr>
<td>Specific indirect 2</td>
<td>0.008&lt;sup&gt;a&lt;/sup&gt;</td>
<td>(0.002, 0.021)</td>
<td>10.53%</td>
</tr>
<tr>
<td>Specific indirect 3</td>
<td>0.013&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(0.004, 0.024)</td>
<td>17.11%</td>
</tr>
</tbody>
</table>

n = 2038. Bootstrap sample size = 5000.
<sup>a</sup>P < 0.05.
<sup>b</sup>P < 0.01.
<sup>c</sup>P < 0.001.

Total indirect: domestic violence → posttraumatic growth/posttraumatic stress disorder (PTG/PTSD);
Specific indirect 1: domestic violence → resilience → PTG/PTSD;
Specific indirect 2: domestic violence → self-compassion → PTG/PTSD;
Specific indirect 3: domestic violence → self-compassion → resilience → PTG/PTSD.

Figure 1 Multiple indirect effects model after controlling age, gender, subjective social economic status, and family structure. <sup>a</sup>P < 0.05, <sup>b</sup>P < 0.01, <sup>c</sup>P < 0.001. PTG: Posttraumatic growth; PTSD: Posttraumatic stress disorder.

Chi XL et al. Common mechanism of PTG and PTSD

trauma and positively reframe their understanding of the PT world[8]. These strategies can be used to help adolescents find meaning, recover from negative emotions, and realize PTG, and vice versa[21]. Additionally, we identified a 2-step indirect path from domestic violence to PTG and PTSD via self-compassion to resilience, which was consistent with hypothesis 3. When facing challenges, individuals with more exposure to domestic violence in their childhood are less likely to be self-compassionate and lead to lower resilience. By contrast, adolescents who are more self-compassionate tend to be resilient and “bounce back” from challenges[33]. This process may reduce PTSD symptoms and promote PTG[34,35]. During the COVID-19 outbreak, individuals with a lower level of self-compassion and resilience may find it more difficult to cope with the stressful event, compared to those with a higher level of self-compassion and resilience. This may eventually lead to PTSD symptoms and reduce the possibility of PTG[36].

Of note, childhood exposure to domestic violence had non-significant effects on PTG and PTSD using a 1-step indirect path via resilience. Taking a closer look, the relationship between childhood exposure to domestic violence and resilience were not significantly correlated, which could be attributed to the fact that resilience is a dynamic development process, accorded with the latest academically accepted definition that it refers to an individual’s efforts to adjust and actively adapt under
stress on the one hand instead of a steady trait[37]. Besides, it emphasizes an individual’s adaptive outcome on the other hand. Thereby, the profile of resilience at one measurement time cannot fully represent the dynamic process.

CONCLUSION
This study has several limitations. First, findings from the self-reported scales need further confirmation, using data from clinical interviews that can determine the presence of being diagnosed with PTSD. Second, the study employed a cross-sectional design, and the corresponding data before the COVID-19 outbreak was not available, thus limiting the causal interpretation of the findings. Thus, a latent profile transition analysis may be conducted within a longitudinal study to examine the developmental trajectory and longitudinal predicting mechanism of PTSD and PTG throughout the COVID-19 outbreak. Third, psychopathological impacts of self-compassion and variations of developing self-compassion among individuals were not evaluated, which may have provided further insights on how PTG and PTSD are developed among individuals. However, this warrants future research. Fourth, due to the cross-sectional design, the participants have to recall their childhood experiences instead of reporting existing circumstances, which leads to the potency of recall bias.

Despite these limitations, this is an exploratory study that examined the mechanisms underlying the association of domestic violence with PTG and PTSD among college students during the COVID-19 pandemic. Findings indicate that domestic violence is common to both PTSD and PTG, which may explain the coexistence of PTG and PTSD in young adults. In addition, it suggests that domestic violence affects PTG and PTSD through similar underlying mechanisms. Self-compassion and resilience are individual’s internal resources, which may increase the PTG and reduce the PTSD symptoms. From a clinical perspective, the unique roles of self-compassion and resilience in PTSD and PTG should be considered. Psychological services engaged to relieve PTSD and promote PTG should develop the competence of self-compassion and resilience of youth. The government should pay attention to domestic violence with timely and appropriate actions that should be taken: (1) Establish and improve the maternal and child protection system; (2) strengthen the support for community work to reduce the occurrence of domestic violence; (3) provide shelter for victims in the domestic violence during the epidemic period; and (4) form a public opinion environment against domestic violence.

ARTICLE HIGHLIGHTS

Research background
It was reported that domestic violence had a 30 percent incidence, which always happened in front of children or adolescents, causing an increased risk of psychological problems such as posttraumatic stress disorder (PTSD) symptoms. But some people who are exposed to domestic violence perform positive changes in psychology such as posttraumatic growth (PTG). It is unclear whether PTG and PTSD share a common underlying mechanism. The present study is exploratory to reveal it during the coronavirus disease 2019 (COVID-19).

Research motivation
Based on the shattered world assumption theory, the PTG model, and conservation of resources theory, this study hypothesized that the self-compassion and resilience are the common factors for PTG and PTSD. The present study addressed the mediator roles of self-compassion and resilience. It may inform future mental health interventions for certain individuals.

Research objectives
The present study aimed to explore the common mechanism of PTG and PTSD, revealing the mediating role of self-compassion and resilience between exposure to domestic violence and PTG/PTSD.

Research methods
A nationwide cross-sectional online survey was conducted in China during the
COVID-19 pandemic. The data was collected using the Chinese version of revised Adverse Childhood Experiences Question, Self-compassion Scale, Connor-Davidson Resilience Scale, PT Growth Inventory, and the Abbreviated PTSD Checklist-Civilian version. A structural equation model was conducted to analyze the data.

**Research results**

The path analysis indicated that exposure to domestic violence was significantly correlated with PTG and PTSD via a 1-step indirect path of self-compassion and via a 2-step indirect path from self-compassion to resilience. The 1-step indirect path of resilience did not reach significance.

**Research conclusions**

PTG and PTSD share a common mechanism for those who were exposed to domestic violence. Those who perform a lower level of self-compassion would tend to difficulty bounce back to normal, causing an increased risk of PTSD and decreased possibility of PTG, and vice versa.

**Research perspectives**

From a clinical perspective, the intervention study could be considered in investigating the roles of self-compassion and resilience for vulnerable individuals.

**REFERENCES**

6 Zhou X, Zhen R, Wu X. Shared and unique mechanisms underlying the association of trauma exposure with posttraumatic stress symptoms and growth among adolescents following the Jiuzhaigou earthquake. Psychol Trauma 2019 [PMID: 31697106 DOI: 10.1037/tra0000526]
Chi XL et al. Common mechanism of PTG and PTSD


26 Wang YR, Lin PZ, Cao FL. Validity and reliability of the Chinese version of the Revised Adverse Childhood Experience Questionnaire. Zhongguo Xinli Weisheng Za Zhi 2018; 13


Submit a Manuscript: https://www.f6publishing.com  
World J Psychiatry 2021 November 19; 11(11): 1116-1128  
DOI: 10.5498/wjp.v11.i11.1116  
ISSN 2220-3206 (online)

**ORIGINAL ARTICLE**

**Observational Study**

**Subgrouping time-dependent prescribing patterns of first-onset major depressive episodes by psychotropics dissection**

Hsi-Chung Chen, Hui-Hsuan Hsu, Mong-Liang Lu, Ming-Chyi Huang, Chun-Hsin Chen, Tzu-Hua Wu, Wei-Chung Mao, Chuhsing K Hsiao, Po-Hsiu Kuo

**ORCID number:** Hsi-Chung Chen 0000-0003-3191-0693; Hui-Hsuan Hsu 0000-0003-3503-3163; Mong-Liang Lu 0000-0003-3184-6554; Ming-Chyi Huang 0000-0001-9751-9834; Chun-Hsin Chen 0000-0002-1776-8556; Tzu-Hua Wu 0000-0002-3631-6248; Wei-Chung Mao 0000-0003-1620-1010; Chuhsing K Hsiao 0000-0003-4905-8484; Po-Hsiu Kuo 0000-0003-0365-3587.

**Author contributions:** Chen HC and Hsu HH drafted the manuscript; Kuo PH, Hsiao CK and Chen HC conceived and designed the study, and critically revised the manuscript; Kuo PH owns primary responsibility for the final content; Lu ML, Huang MC, and Chen CH assisted to refer the patients as participants; Lu ML, Huang MC, Chen CH, Wu TH, and Mao WC gave critical opinions on the study design and the manuscript; all authors have read and approve the final manuscript.

**Institutional review board statement:** This study was approved by the research ethics committees of National Taiwan University Hospital, Taipei Municipal Wanfang Hospital, and Taipei City Hospital, Songde Branch.

**Informed consent statement:** All participating patients provided informed consent.

**Abstract**

**BACKGROUND**

Subgrouping patients with major depressive disorder is a promising solution for the issue of heterogeneity. However, the link between available subtypes and distinct pathological mechanisms is weak and yields disappointing results in clinical application.

**AIM**

To develop a novel approach for classification of patients with time-dependent prescription patterns at first onset in real-world settings.

**METHODS**

Drug-naive patients experiencing their first major depressive episode ($n = 105$) participated in this study. Psychotropic agents prescribed in the first 24 mo
Major depressive disorder (MDD) is a common and challenging mental illness[1]. Even though research on the biological substrates of MDD is increasing, the heterogeneity of MDD substantially compromises the applicability of these study findings[2]. Thus, subgrouping MDD into homogeneous clusters, which could be used to investigate specific neurobiological and pathological mechanisms, is a promising solution for this.

A number of clinical features were previously used to subgroup MDD[3]. Most subtyping schemes (such as melancholic, psychotic, and atypical features) are based on differences in patterns of clinically observed characteristics. However, the link between these subtypes and distinct pathological mechanisms is weak and yields disappointing results in clinical application[3,4]. Even using data-driven latent class analyses, there is inadequate evidence to support the existence of symptom-based
Heterogeneous clustering for major depression

In this study, we developed a novel approach by clustering patients with similar biologically relevant potentials. Our approach was based on three conceptual aspects. First, parameters that were correlated with biological mechanisms were preferred to theoretical or symptom-based constructs. Second, non-descriptive parameters were considered superior to linguistic-defined symptoms. Third, creating parameters that enrich the available information helped sub-classify patients into data-driven clusters. While practicing physicians commonly initiate treatment according to clinical guidelines, the ensuing adjustments to the regimen are essentially based on the temporal change of clinical response. The medication class and dose of prescribed psychotropics evolve naturally and reflect the underlying disease entity, as well as the patient’s preference and treatment response at a given moment. The time-dependent change in regimen is dynamically determined by considering factors such as symptom profiles, clinical features, disease stage, treatment response, effectiveness and side effects of medications, and level of functional recovery. As a result, we adopted a time-dependent prescription pattern in the early disease stage as the preferred indicator for MDD subgrouping.

In summary, we aimed to develop a novel approach to tackle the heterogeneity in drug-naïve patients experiencing their first episode of MDD, using detailed prescription patterns of major psychotropic classes during the early disease course. Specifically, we developed new parameters to capture time-dependent information on prescription patterns in the first two years after the initial diagnosis. Using this information, we performed a K-means clustering analysis for data-driven subgrouping to yield empirical subtypes of MDD.

MATERIALS AND METHODS

Participants
This study used the cohort established by the Research Collaborating Group for New Insight, Strategy and Evaluation–Treatment-Resistant Depression Program (RECOGNISE-TRD program). Between October 2010 and April 2016, remitted individuals with a DSM-IV-TR diagnosis of MDD and between 18 and 65 years old were recruited to participate in this study. Participants were eligible if they had more than 2 years of complete medical records from their first episode to the date of recruitment. The duration of the MDD course beyond the 2 years after onset to the date of recruitment differed between participants. The medical records should document all of the participants’ psychiatric and medical history throughout the disease course, including prescriptions, use of outpatient and inpatient services, and information on their comorbidities. If participants discontinued their treatments within 2 years of disease onset because of full remission, they were contacted to ensure the absence of any depressive episode before recruitment. Participants were excluded if they had incomplete medical records, were comorbid with organic brain syndromes, dementia, substance abuse, psychotic disorders, and schizoaffective disorders or if their diagnosis had been revised to bipolar affective disorder. A total of 105 participants were eligible. This study was approved by the research ethics committees of National Taiwan University Hospital, Taipei Municipal Wanfang Hospital, and Taipei City Hospital, Songde Branch. Written informed consent was obtained from each participant. All methods were performed in accordance with the relevant guidelines and regulations.

Pharmacological dissection and the calculation of psychotropic loads
Three major classes of psychotropic agents are used to treat MDD, including antidepressants and adjunctive agents (augmentation agents and hypnosedatives). Following disease onset, the drug class, dosage, date of administration, and duration of prescription were recorded in 4-wk intervals. Since the natural course of a major depressive episode is usually between 6-13 mo, the present study evaluated the prescription pattern of psychotropics using monthly data collected over the first 2 years after disease onset. Observations collected over this period provided adequate time for patients to experience all treatment regimens suggested by the treatment

subtypes. The heterogeneity in MDD is derived from multiple aspects of the disease, including symptom presentation, clinical features, family history, comorbidities, phenotypic stability over time, and treatment response. With these complexities in mind, our goal was to identify the biologically relevant clusters of MDD. The fundamental scheme and analytical technique need to be enhanced to facilitate disease subtyping and improve the identification of the homogeneous clusters of MDD.

In this study, we developed a novel approach by clustering patients with similar biologically relevant potentials. Our approach was based on three conceptual aspects. First, parameters that were correlated with biological mechanisms were preferred to theoretical or symptom-based constructs. Second, non-descriptive parameters were considered superior to linguistic-defined symptoms. Third, creating parameters that enrich the available information helped sub-classify patients into data-driven clusters. While practicing physicians commonly initiate treatment according to clinical guidelines, the ensuing adjustments to the regimen are essentially based on the temporal change of clinical response. The medication class and dose of prescribed psychotropics evolve naturally and reflect the underlying disease entity, as well as the patient’s preference and treatment response at a given moment. The time-dependent change in regimen is dynamically determined by considering factors such as symptom profiles, clinical features, disease stage, treatment response, effectiveness and side effects of medications, and level of functional recovery. As a result, we adopted a time-dependent prescription pattern in the early disease stage as the preferred indicator for MDD subgrouping.

In summary, we aimed to develop a novel approach to tackle the heterogeneity in drug-naïve patients experiencing their first episode of MDD, using detailed prescription patterns of major psychotropic classes during the early disease course. Specifically, we developed new parameters to capture time-dependent information on prescription patterns in the first two years after the initial diagnosis. Using this information, we performed a K-means clustering analysis for data-driven subgrouping to yield empirical subtypes of MDD.
guidelines.

In this study, pharmacological dissection was used to categorize the psychotropic agents, including antidepressants, augmentation agents and hypnosedatives.[10] Antidepressants used in this study include selective serotonin reuptake inhibitors, serotonin and norepinephrine reuptake inhibitors, tricyclic antidepressants, mirtazapine, bupropion, moclobemide, and agomelatine. Augmentation agents included anticonvulsants, antipsychotics, lithium, thyroxin, and methylphenidate. All benzodiazepine receptor agonists (BzRA) were categorized as hypnosedatives. Cumulative doses of each psychotropic medication class were converted into equivalents and designated as antidepressant load (ADL), augmentation load (AUGL), and sedative load (SL). These three equivalent-based psychotropic loads have been used in the literature[10]. For each participant, the monthly cumulative equivalents for 3 classes of psychotropic loads were calculated in 4-wk intervals, starting with the initiation of treatment.

Doses of antidepressants and augmentation agents were transformed into equivalents using the Anatomical Therapeutic Chemical (ATC) system[11]. In analyzing hypnosedative use, equivalent doses of all BzRA were calculated, including benzodiazepines (BZDs) and the Z-drugs (zaleplon, zolpidem, zopiclone, and eszopiclone). As the defined daily dose in the ATC system for hypnosedatives do not adequately represent the clinical usage of BZDs as anxiolytics or hypnotics[12], the BZDs equivalents suggested by the Ashton manual were adopted in the present study[13]. Since conversion ratios for a number of BZD drugs commonly used in Taiwan are not provided in the original Ashton manual (including brotizolam, midazolam, oxazolam, and fludiazepam), study investigators defined the equivalents for these agents by consensus. The full list of BZD equivalents is provided in Supplementary Table 1.

**Mental and physical comorbidities**

In this study, anxiety disorders and physical comorbidities were incorporated as covariates in the multivariable models evaluating the differences in several clinical severity indicators between empirical clusters. The presence of an anxiety disorder was defined as more than three consecutive instances that matched the DSM-IV-TR diagnostic criteria during the first two years following MDD onset. Additionally, physical comorbidities were evaluated as the number of medical diseases identifiable using the Charlson comorbidity index[14].

**Clinical severity indicators**

Several domains of clinical severity were used to examine the validity of data-driven clusters. Firstly, the total number of prescribed antidepressants and the number of prescribed antidepressants with adequate dosage and adequate duration (ADAD) were calculated within the first two years after disease onset. The adequate dosage used was as per the dose suggested by the Ministry of Health and Welfare, Taiwan, for treating MDD[15]. An adequate duration was defined as at least two weeks of treatment[16,17]. Secondly, the cumulative frequency of patients’ utilization of psychiatric facilities, including outpatient services and psychiatric admissions, were documented for within 2 years of disease onset and from 2 years after disease onset to the date of recruitment, respectively. Additionally, the severity of depression was scored on the recruitment date using the Chinese version of the Beck Depression Inventory-II (BDI-II): Level 0 for scores 0–16 (euthymic); 1 for 17–22 (mild); 2 for 23–30 (moderate); and 3 for 31–63 (severe)[18].

**Statistical analyses**

Data-driven clustering analyses: A K-means clustering analysis was used for empirical subtyping. To optimize the data analyses, biological treatment history was longitudinally recorded for each participant in 4-wk intervals, yielding 24 sets of data points. Four parameters were used to capture the time-dependent prescription pattern for each class of psychotropic agents. In each participant, 24 data points for psychotropic load were calculated to determine the stability of prescribed dosage, frequency of high-dosage prescriptions, frequency of prescriptions, and the time trend of prescribed dosage. The stability of the prescribed dosage was a reflection of dosage fluctuation over the 2 year period. It was evaluated by estimating the coefficient of variance within the 24 points of data for each psychotropic load. Secondly, a count of high dosage prescriptions was obtained as an index of the need for high levels of specific interventions to achieve symptomatic control. The median of each class of psychotropic load was calculated for all participants using dichotomized cutoffs. The
24 data points for each class of psychotropic load in individual participants were compared with the cutoff. The total number of points in which the psychotropic loads were equal or greater than the median was recorded. Higher counts corresponded to a greater frequency of high-dosage prescription. Thirdly, the frequency of the prescription denoted the count of the presence of psychotropic load in the 24 data points. It represents the frequency of utilization of outpatient services and indicates the intensity of psychiatric care. Finally, the trend over time of the prescribed dosage evaluated the increments or decrements in psychotropic dosage. The correlation coefficient between each psychotropic load and time was determined for each participant to quantify the time trend of the prescribed dosage. A total of 12 parameters were generated for each participant. A K-means clustering analysis was performed using all collected parameters. The number of cluster K was an input parameter, and it was set at a priori between 3 and 5. The final class of subgroups was determined based on clinical relevance and implications. Additionally, we noted that if the scales of the input parameters are notably different, the contribution of these parameters may be uneven. Because the scales of ‘the frequency of prescription’ and ‘the level of dosage’ are greater than the other variables, we conducted sensitivity analysis to use these two parameters only, and perform the clustering analysis again. We then compared the differences of resultant clusters for evaluating the robustness of the current findings.

Between-cluster comparisons: After data-driven clustering analyses were completed using the K-means method, several clinical severity indicators were used to examine between-cluster differences. In these analyses, a Chi-square and analysis of variance (ANOVA) were used as univariate analyses. For multivariable regression analyses (excluding analyses of the severity of depression), all clinical severity data were taken into account, and Poisson regression analyses were conducted because of count data. General linear models were constructed in consideration of the between-cluster differences in depression severity at the end of the follow-up period. Data were analyzed using the SPSS version 17.0 (SPSS Inc., Chicago, IL, United States) statistical package. The statistical significance level was set at a P value of < 0.05.

RESULTS

A total of 105 individuals with first-onset MDD participated in the study. The average age of participants was 51.5 ± 13.2 years old, their age at MDD onset was 44.5 ± 13.6 years old, 75.5% were female, and 20.8% had concurrent anxiety disorders (Table 1).

According to the results of the K-means clustering analyses, 4 feature-based clusters were identified. Based on the differences between regimens, psychotropic load, and prescription duration, the clusters were designated as short-term antidepressant use (SA, n = 33), long-term antidepressant use (LA, n = 34), long-term antidepressants and sedatives use (LAS, n = 19), and long-term antidepressants, sedatives, and augmentation use (LASA, n = 19). The parameters used for the K-means clustering analyses are detailed in Supplementary Table 2. Additionally, SA and LA clusters were denoted as the simple pharmacy group, due to the relatively simple pattern of psychotropic agent use. In contrast, the LAS and the LASA clusters comprised the complex pharmacy group. Table 1 also summarizes the univariate comparisons of various demographic characteristics and clinical indicators between the 4 clusters.

Figure 1 depicts the change over time in the monthly equivalents in the 4 clusters based on three classes of psychotropic loads. The lowest average ADL for the first 2 years following disease onset in the SA reflects the lower initial dosage and shorter exposure period of these participants (Figure 1A). Figure 1A also demonstrates that ADL in the LA cluster is consistent over time and similar to the complex pharmacy group, while the dosage of antidepressants was lower in the LA cluster. This finding is consistent with a comparison of the averaged ADL values between clusters. Additionally, Figure 1B presents the time-course of SL in the 4 clusters. Complex pharmacy group clusters exhibited steady levels and the higher equivalent of hypnosedatives, which discriminates them from the simple pharmacy group. Figure 1C illustrates the unique properties of LASA in the prescription pattern of augmentation agents. The three clusters (excluding LASA) exhibited less monthly AUGL until 4 mo after MDD onset.

Table 2 summarizes the differences in various clinical severity indicators between clusters, after controlling for the confounding effects. In the first 2 years following MDD onset, no significant difference was observed between the 4 clusters in the numbers of antidepressants used. However, the number of ADAD in the SA was

Table 1 summarizes the differences in various clinical severity indicators between clusters, after controlling for the confounding effects. In the first 2 years following MDD onset, no significant difference was observed between the 4 clusters in the numbers of antidepressants used. However, the number of ADAD in the SA was
Table 1 Univariate comparisons of demographic and clinical features by empirically-derived clusters (n = 105), mean ± SD

<table>
<thead>
<tr>
<th>Demographic and general clinical data</th>
<th>Total</th>
<th>Simple pharmacy, short-term antidepressants use (SA, n = 19)</th>
<th>Simple pharmacy, long-term antidepressants use (LA, n = 33)</th>
<th>Complex pharmacy, long-term antidepressants and sedatives use (LAS, n = 34)</th>
<th>Complex pharmacy, long-term antidepressants, sedatives and augmentation use (LASA, n = 19)</th>
<th>P value for univariate omnibus test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>51.5 (13.2)</td>
<td>42.6 (14.4)</td>
<td>50.6 (15.2)</td>
<td>54.9 (9.1)</td>
<td>56.5 (10.3)</td>
<td>0.02</td>
</tr>
<tr>
<td>Age of onset (yr)</td>
<td>44.5 (13.6)</td>
<td>36.3 (14.4)</td>
<td>43.2 (15.3)</td>
<td>48.6 (10.0)</td>
<td>49.6 (11.0)</td>
<td>0.01</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>80 (75.5)</td>
<td>15 (78.9)</td>
<td>24 (72.7)</td>
<td>27 (79.4)</td>
<td>13 (68.4)</td>
<td>0.79</td>
</tr>
<tr>
<td>Comorbid anxiety disorder, n (%)</td>
<td>22 (20.8)</td>
<td>6 (31.6)</td>
<td>9 (27.3)</td>
<td>4 (11.8)</td>
<td>3 (15.8)</td>
<td>0.25</td>
</tr>
<tr>
<td>Number of comorbid physical diseases</td>
<td>0.7 (1.3)</td>
<td>0.5 (1.2)</td>
<td>0.7 (1.4)</td>
<td>1.1 (1.5)</td>
<td>0.6 (1.2)</td>
<td>0.39</td>
</tr>
<tr>
<td>Follow-up duration after 2-yr of disease onset (yr)</td>
<td>3.7 (3.5)</td>
<td>3.2 (3.2)</td>
<td>3.4 (3.1)</td>
<td>3.7 (3.2)</td>
<td>3.8 (3.9)</td>
<td>0.91</td>
</tr>
<tr>
<td>Psychotropic loads</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antidepressant load (defined daily dose/month)</td>
<td>34.5 (18.4)</td>
<td>13.3 (7.2)</td>
<td>32.1 (13.1)</td>
<td>42.8 (17.3)</td>
<td>44.8 (17.7)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sedative load (equivalent dose/month)</td>
<td>57.3 (56.8)</td>
<td>8.1 (10.8)</td>
<td>18.5 (20.9)</td>
<td>91.9 (45.4)</td>
<td>112.0 (56.7)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Augmentation load (defined daily dose/month)</td>
<td>1.2 (2.7)</td>
<td>0.1 (0.5)</td>
<td>0.3 (0.6)</td>
<td>0.3 (1.0)</td>
<td>5.3 (4.0)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Clinical feature within 2-yr of disease onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of antidepressants use</td>
<td>1.7 (0.9)</td>
<td>1.3 (0.6)</td>
<td>1.6 (0.8)</td>
<td>1.9 (1.1)</td>
<td>1.9 (0.9)</td>
<td>0.07</td>
</tr>
<tr>
<td>Number of antidepressants use with adequate dosage and duration</td>
<td>1.5 (0.9)</td>
<td>1.1 (0.5)</td>
<td>1.3 (0.8)</td>
<td>1.7 (1.2)</td>
<td>1.8 (0.8)</td>
<td>0.04</td>
</tr>
<tr>
<td>Visits of psychiatric outpatient service</td>
<td>17.3 (10.7)</td>
<td>10.6 (5.8)</td>
<td>15.5 (5.1)</td>
<td>20.2 (10.6)</td>
<td>21.8 (17.1)</td>
<td>0.002</td>
</tr>
<tr>
<td>Frequencies of psychiatric admission</td>
<td>0.2 (0.5)</td>
<td>0.1 (0.3)</td>
<td>0.2 (0.4)</td>
<td>0.1 (0.3)</td>
<td>0.7 (0.8)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Clinical feature beyond 2-yr of disease onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visits of psychiatric outpatient service</td>
<td>31.1 (66.7)</td>
<td>9.7 (10.3)</td>
<td>20.2 (24.7)</td>
<td>41.4 (91.6)</td>
<td>42.6 (78.4)</td>
<td>0.22</td>
</tr>
<tr>
<td>Frequencies of psychiatric admission</td>
<td>0.8 (0.5)</td>
<td>0.0 (0.0)</td>
<td>0.1 (0.4)</td>
<td>0.1 (0.7)</td>
<td>0.0 (0.0)</td>
<td>0.71</td>
</tr>
<tr>
<td>Severity of depression in the recruitment</td>
<td>1.5 (0.9)</td>
<td>0.6 (1.0)</td>
<td>0.9 (1.2)</td>
<td>1.2 (1.3)</td>
<td>1.7 (1.2)</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Severity of depression was scored according to the level of depression, which was evaluated by the Beck Depression Inventory.

Significantly lower than in LAS (P = 0.04) and LASA (P = 0.03). Although the average ADL of LA is significantly lower than the complex pharmacy group (Table 1), the number of used antidepressants and ADAD did not significantly differ between the LA and the complex pharmacy group clusters. Analysis of the use of outpatient services within the first 2 years of disease course, showed a significantly lower number...
Table 2 Multivariable analyses for the comparisons of clinical features by empirically-derived clusters (n = 105)

<table>
<thead>
<tr>
<th>Clinical feature within 2-yr of disease onset</th>
<th>Simple pharmacy, short-term antidepressants use (SA, n = 19)</th>
<th>Simple pharmacy, long-term antidepressants use (LA, n = 33)</th>
<th>Complex pharmacy, long-term antidepressants and sedatives use (LAS, n = 34)</th>
<th>Complex pharmacy, long-term antidepressants, sedatives and augmentation use (LASA, n = 19)</th>
<th>P value for pos-hoc comparison using multivariable regression test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of antidepressants use</td>
<td>0.57 (0.32-0.82)</td>
<td>0.74 (0.52-0.97)</td>
<td>0.94 (0.66-1.23)</td>
<td>0.96 (0.61-1.31)</td>
<td>SA &lt; LAS (P = 0.04); SA &lt; LASA (P = 0.03)</td>
</tr>
<tr>
<td>Number of antidepressants use with adequate dosage and duration</td>
<td>0.50 (0.26-0.73)</td>
<td>0.64 (0.44-0.85)</td>
<td>0.87 (0.59-1.15)</td>
<td>0.95 (0.60-1.31)</td>
<td></td>
</tr>
<tr>
<td>Visits of psychiatric outpatient service</td>
<td>4.76 (4.04-5.47)</td>
<td>7.46 (6.75-8.17)</td>
<td>10.34 (9.37-11.30)</td>
<td>11.37 (10.15-12.58)</td>
<td>SA &lt; LA (P &lt; 0.001); SA &lt; LAS (P &lt; 0.001); SA &lt; LASA (P &lt; 0.001); LA &lt; LAS (P &lt; 0.001); LA &lt; LASA (P &lt; 0.001)</td>
</tr>
<tr>
<td>Frequencies of psychiatric admission</td>
<td>0.03 (-0.02-0.08)</td>
<td>0.05 (-0.002-0.10)</td>
<td>0.03 (-0.01-0.08)</td>
<td>0.24 (0.04-0.44)</td>
<td>SA &lt; LASA (P = 0.01); LA &lt; LASA (P = 0.004); LASA &lt; LASA (P = 0.001)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical feature beyond 2-yr of disease onset</th>
<th>Simple pharmacy, short-term antidepressants use (SA, n = 19)</th>
<th>Simple pharmacy, long-term antidepressants use (LA, n = 33)</th>
<th>Complex pharmacy, long-term antidepressants and sedatives use (LAS, n = 34)</th>
<th>Complex pharmacy, long-term antidepressants, sedatives and augmentation use (LASA, n = 19)</th>
<th>P value for pos-hoc comparison using multivariable regression test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visits of psychiatric outpatient service</td>
<td>3.03 (2.58-3.49)</td>
<td>5.81 (5.34-6.30)</td>
<td>10.21 (9.49-10.93)</td>
<td>10.86 (9.98-11.74)</td>
<td>SA &lt; LA (P &lt; 0.001); SA &lt; LAS (P &lt; 0.001); SA &lt; LASA (P &lt; 0.001); LA &lt; LAS (P &lt; 0.001); LA &lt; LASA (P &lt; 0.001)</td>
</tr>
<tr>
<td>Frequencies of psychiatric admission</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Severity of depression in the recruitment</td>
<td>0.58 (0.23-0.93)</td>
<td>0.90 (0.54-1.25)</td>
<td>1.66 (1.10-2.22)</td>
<td>2.20 (1.37-3.30)</td>
<td>SA &lt; LAS (P = 0.002); SA &lt; LASA (P &lt; 0.001); LA &lt; LAS (P = 0.02); LA &lt; LASA (P = 0.001)</td>
</tr>
</tbody>
</table>

1Covariates included into the Poisson regression model were sex, age of onset, comorbid anxiety disorder, and number of medical comorbidities. OFFSET = logarithm (2).
2Covariates included into the multivariable Poisson regression model were sex, age of onset, comorbid anxiety disorder, and number of medical comorbidities. OFFSET = logarithm (duration of course beyond 2-yr of disease onset).
3Covariates included into the multivariable Poisson regression model were sex, age of onset, comorbid anxiety disorder, and number of medical comorbidities. OFFSET = logarithm (duration of course beyond 2-yr of disease onset). This model failed to converge because of zero count in some clusters.
4Covariates included into the general linear model were sex, age of onset, comorbid anxiety disorder, number of medical comorbidities, and duration of course beyond 2-yr of disease onset. Severity of depression was scored according to the level of depression, which was evaluated by the Beck Depression Inventory. IRR: Incident rate ratio; EMM: Estimated marginal mean. CI: Confidence interval.

The simple pharmacy group exhibited a lower frequency of visits to outpatient services after 2 years from disease onset than the complex pharmacy group. However, no significant difference was observed between LAS and LASA. The LASA cluster exhibited a higher frequency of psychiatric admission in the first 2 years after MDD onset, compared with the other 3 clusters. However, SA, LA, and LAS did not differ in terms of admission frequency. The simple pharmacy group exhibited a lower frequency of visits to outpatient services after 2 years from disease onset than the complex pharmacy group. However, no significant difference was observed between LAS and LASA. Finally, the clusters in the complex pharmacy group exhibited more severe depressive symptoms than the clusters of the simple pharmacy group. However, no within-group differences were found in terms of depression severity in both the simple and complex pharmacy groups (Table 2).
To evaluate the robustness of the clustering results, we re-analyzed the data with various parameter combinations. Our sensitivity analysis revealed similar clustering results. By adopting only 2 parameters, the frequency of high-dosage prescriptions, and the frequency of prescriptions, the K-means clustering analysis still yielded the same 4 clusters as were generated by the original 4 parameters.
DISCUSSION

The patients were empirically classified into 4 clusters according to the K-means clustering analysis that dissected the time-dependent prescription pattern over the first 2 years following the onset of the first episode of MDD. This analysis was facilitated by novel parameters that successfully capture the reciprocal interactions between physicians’ prescriptions and disease status in a real-world setting. The 4 identified clusters exhibited distinct patterns in terms of the dosage of psychotropic agents, duration of prescription, and the combinations of psychotropic agents used. Additionally, the 4 clusters also exhibited differences in severity of clinical indicators. To the best of our knowledge, this is the first attempt to classify MDD by pharmacological dissection considering time-dependent prescription patterns using a data-driven clustering analysis.

While no between-cluster differences were observed in the number of antidepressants used in this study, the SA exhibited lower number of ADAD used, compared to the clusters comprising the complex pharmacy group. In the clinical setting, the number of antidepressants used reflects the expense involved in the use of sequential trials to select appropriate antidepressants. In contrast, the number of ADAD is a reflection of inadequate treatment response after a series of interventions[20]. Therefore, data-driven clusters in the present study suggest a different level of treatment difficulty, beyond the number of trials used to identify a tolerable antidepressant regimen. Furthermore, the use of psychiatric facility (an index of clinical disease burden) showed differences between clusters. The complex pharmacy group exhibited higher number of visits to outpatient services compared to the simple pharmacy group within the first 2 years following MDD onset and beyond. The use of outpatient services reflects the need for re-evaluation or stabilization of symptoms, and serves as an indicator of long-term treatment stability. Taken together, our findings suggest that the patients in the complex pharmacy group require more intensive care to stabilize their symptoms. Similarly, the level of BDI-II-defined depression on the recruitment day corresponded to the pattern of residual symptoms over time in the complex pharmacy group. Finally, psychiatric admissions are often indicated in life-threatening conditions or situations where the treatment response remains inadequate after a series of guideline-based interventions. LASA, in comparison with other clusters with low AUGL, uniquely exhibited a highest frequency of psychiatric admissions. These findings support the clinical relevance of the prescription-based subtyping strategy.

According to the established pharmacological mechanisms of each class of psychotropic agent, the four prescription-based clusters correspond to distinct biological mechanisms. In the simple pharmacy group, SA and LA were associated with similar initial doses of antidepressants, but differed in the duration of prescription. Consistent with the clinical practice, the episodes in the SA cluster exhibited the expected response to antidepressants during the acute and maintenance phases. In contrast, LA cluster of patients needed a longer period of maintenance therapy due to vulnerability to relapse or recurrence, despite the apparent adequate antidepressant treatment. This finding suggests that the traditional monoaminergic antidepressants can correct the neurochemical deficits of MDD in clusters comprising the simple pharmacy group.

Compared with the relatively short-term and low-dose exposure to hypnosedatives in the simple pharmacy group, clusters in the complex pharmacy group were characterized by longer duration and high doses of hypnosedative use. Interestingly, the rate of comorbidity of anxiety disorders did not differ significantly between the 4 clusters. Although antidepressants were suggested as the primary therapy for anxiety[20], monoaminergic antidepressants alone seem insufficient for patients in the complex pharmacy group. Contrary to the recommendation of short-term adjunct use of hypnosedatives in the treatment guidelines[21], depending on study populations, 7.6 to 60% patients with MDD initiated antidepressant treatment with concurrent hypnosedatives, while 12% to 48% of patients received long-term combined treatment [22-28]. Indeed, the monoaminergic deficiency hypothesis was regards as insufficient to explain the heterogeneity of MDD [29]. Additionally, monoaminergic antidepressants and benzodiazepines may alleviate the anxiety symptoms through different mechanisms[29,30]. Therefore, clusters within the complex pharmacy group may have specific pathogenic mechanisms that cannot be resolved by monoaminergic antidepressants alone, such as deficits in the gamma-aminobutyric acid-related system[29]. This finding suggests that, instead of subgrouping by comorbidity of anxiety disorders, classifying individuals by hypnosedative use pattern over 2 years may be more informative and facilitate the identification of common neurochemical deficits among these patients. Moreover, in addition to long-term hypnosedative use, LASA is...
characterized by long-term concomitant use of augmentation agents. Because various medications with different pharmacological mechanisms were included as augmentation agents in this study, the heterogeneity of neurochemical deficits in the LASA cluster may be more complex than originally believed.

Clusters comprising the complex pharmacy group exhibited a higher initial SL that persisted over the first 2 year after MDD onset. This is consistent with the previously published findings showing that depressed patients who used a higher daily dosage of sedatives in the initial treatment phase tend to be long-term users[28]. Additionally, the association between sedative use and clinical severity was previously found to be independent of the use of antidepressants and augmentation agents[10]. The complex pharmacy group in this study exhibited high dosage and long duration of hypnotic and antidepressive agents, which correlated with worse indicators of disease severity. However, when AUGL is taken into account, participants with long-term use of hypnotic use could be further partitioned into LAS and LASA, which differ in aspects of clinical severity (Table 2). In this study, the time-dependent pattern of ADL separated the SA, LA, and LASA clusters comprising the complex pharmacy. When SL and AUGL are taken into consideration, a difference was detected between clusters, enhancing the clinical implications of these data-driven clusters. Therefore, simultaneous inclusion of the 3 classes of psychotropic loads has distinct and complementary function in the clustering analysis.

There are a number of limitations to this study. First, the strict inclusion criteria resulted in a sample size which is not large enough to ensure that all statistical models are sufficiently powered. However, under the current sample size, most of the analyses have successfully demonstrated between-cluster differences in terms of clinical severity indicators. This observation indirectly supports a significant extent of between-cluster differences between the resulting clusters. Additionally, people in Taiwan have high accessibility and availability to medical resources. Since eligible participants had to have at least 2 years of complete treatment history in a teaching hospital, they usually exhibited superb adherence, which may influence the selection of participants. Therefore, study findings may not be generalized to all patient groups. Second, medications compliance is unknown in our participants. Therefore, the presented psychotropic loads may not reflect the actual dose taken by the participants. Moreover, we did not know if the patients concurrently received pharmacological or non-pharmacological treatment from other hospitals, which could introduce misclassification bias. Third, baseline severity of depression was not evaluated in this study. However, findings of a previous study suggest that the initial severity of MDD is not correlated with treatment response[31]. Absence of this information is unlikely to change the clustering results. Fourth, the latency from the disease onset to the commencement of treatment is not recorded in this study. Since the delay in treatment makes it more difficult to treat the major depressive episode, this may introduce bias in inference of homogeneity in pathogenic mechanisms between clusters. Finally, the resultant clusters in this study may merely reflect therapists’ prescription preferences. According to the examination of clinical validity, we believe that these empirically-derived clusters, at least in part, reflect the underlying pathological mechanisms.

CONCLUSION

With the trend of increased data sharing of medical registries, prescription patterns are well established in most data banks. While claimed based datasets do not have detailed information on symptoms or dynamic disease status in a real-world setting, these kinds of data would have large sample size and large amounts of data regarding prescription patterns. By using such kinds of databases, researchers not only can investigate the heterogeneity of MDD but also explore the complexity feature with other comorbid health conditions, and can easily expand to study other mental disorders. Additionally, machine learning approaches, such as K-means clustering analyses can be used to optimize the subgrouping information. In the future, this technique can be enhanced by reducing the required number of input parameters and duration of prescription records. The biological validity, generalizability, and predictive value of this new clustering approach warrant further investigation.
ARTICLE HIGHLIGHTS

Research background
Major depressive disorder (MDD) is a common and challenging mental illness. Even though research on the biological substrates of MDD is increasing, the heterogeneity of MDD substantially compromises the applicability of these study findings. Thus, subgrouping MDD into homogeneous clusters, which could be used to investigate specific neurobiological and pathological mechanisms, is a promising solution.

Research motivation
The heterogeneity in MDD is derived from multiple aspects of the disease. There is inadequate evidence to support the existence of symptom-based subtypes. Thus, we intended to develop a novel approach by clustering patients with similar biologically relevant potentials.

Research objectives
We aimed to develop a novel approach to tackle the heterogeneity in drug-naive patients experiencing their first episode of MDD, using detailed prescription patterns of major psychotropic classes during the early disease course.

Research methods
Psychotropic agents prescribed in the first 24 mo following disease onset were recorded monthly and categorized as antidepressants, augmentation agents, and hypnosedatives. Four parameters were used to summarize the time-dependent prescription patterns for each psychotropic agent. A K-means cluster analysis was used to derive subgroups of participants based on these four parameters of psychotropic agents across 24 mo.

Research results
The patients were empirically classified into 4 clusters according to the K-means clustering analysis that dissected the time-dependent prescription pattern over the first 2 years following the onset of the first episode of MDD. The four identified clusters exhibited distinct patterns in terms of the dosage of psychotropic agents, duration of prescription, and the combinations of psychotropic agents used. Additionally, the four clusters also exhibited differences in severity of clinical indicators.

Research conclusions
Our novel parameters successfully captured the reciprocal interaction between physicians' prescriptions and disease status in a real-world setting. This study presents a novel clustering strategy that can be used to generate prescription-based subtypes.

Research perspectives
Machine learning approaches, such as K-means clustering analyses can be used to optimize the subgrouping information. The biological validity, generalizability, and predictive value of this new clustering approach warrant further investigation.

ACKNOWLEDGEMENTS
The authors would like to thank the Dr. Chin-Hao Chang for the assistance of statistical analysis; Dr. Chih-Kang Hsu, Dr. Hsien-hsueh Shih, Dr. Yen-Chin Wang, Dr. Hai-Ti Lin, Dr. Tzu-Yu Liu, Dr. Meng-Shiuan Shie and research assistants Miss I-Chen Huang and Miss Chieh-Chun Hu for their help with data collection.

REFERENCES
2 McIntosh AM, Sullivan PF, Lewis CM. Uncovering the Genetic Architecture of Major Depression.
WJP | https://www.wjgnet.com

Chen HC et al. Heterogeneous clustering for major depression


25 Pfeiffer PN, Ganoczy D, Zivin K, Valenstein M. Benzodiazepines and adequacy of initial


29 Luscher B, Shen Q, Sahir N. The GABAergic deficit hypothesis of major depressive disorder. Mol Psychiatry 2011; 16: 383-406 [PMID: 21079608 DOI: 10.1038/mp.2010.120]


New-onset depression after hip fracture surgery among older patients: Effects on associated clinical outcomes and what can we do?

Hao-Cheng Qin, Zhi-Wen Luo, Heng-Yi Chou, Yu-Lian Zhu

ORCID number: Hao-Cheng Qin 0000-0003-2723-4299; Zhi-Wen Luo 0000-0002-0524-9951; Heng-Yi Chou 0000-0001-6525-2899; Yu-Lian Zhu 0000-0001-5530-144X.

Author contributions: Qin HC and Luo ZW conducted the original search and wrote the first draft of the paper; Qin HC, Luo ZW, and Chou HY screened the selected articles and contributed to subsequent drafts of the manuscript; Zhu YL conceived the idea for the study and provided suggestions; Qin HC and Luo ZW made equal contributions to the work.

Conflict-of-interest statement: The authors admitted that there is no conflict of interest.

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the PRISMA 2009 Checklist.

Country/Territory of origin: China

Specialty type: Psychiatry

Provenance and peer review: Invited article;Externally peer reviewed

Peer-review report’s scientific

Abstract

BACKGROUND
Hip fracture in the elderly is a worldwide medical problem. New-onset depression after hip fracture has also received attention because of its increasing incidence and negative impact on recovery.

AIM
To provide a synthesis of the literature addressing two very important questions arising from postoperative hip fracture depression (PHFD) research: the risk factors and associated clinical outcomes of PHFD, and the optimal options for intervention in PHFD.

METHODS
We searched the PubMed, Web of Science, EMBASE, and PsycINFO databases for English papers published from 2000 to 2021.

RESULTS
Our results showed that PHFD may result in poor clinical outcomes, such as poor physical function and more medical support. In addition, the risk factors for PHFD were summarized, which made it possible to assess patients preoperatively. Moreover, our work preliminarily suggested that comprehensive care may be the optimal treatment option for PHFDs, while interdisciplinary intervention can also be clinically useful.
CONCLUSION
We suggest that clinicians should assess risk factors for PHFDs preoperatively, and future research should further validate current treatment methods in more countries and regions and explore more advanced solutions.

Key Words: Depression; Hip fracture; New onset; Risk factors; Treatments; Clinical outcomes

INTRODUCTION
Hip fractures are one of the most common physical injuries in the elderly, and they often occur after a sudden fall in older people with osteoporosis. In addition to osteoporosis, other risk factors include aging, osteoarthritis, malnutrition, muscle weakness, arthritis, and visual impairment[1-3]. At least 120 million people around the world are affected by hip fractures. Between 2012 and 2016, the cumulative incidence of hip fractures among those aged 55 years and over in urban China was 148.75 per 100,000. Moreover, the cost of hospitalization for hip fractures increased six-fold from 2012 to 2016, while per person, it increased by 1.59 times[4,5]. Given that the global population of the middle-aged and elderly is expanding, hip fractures will be sure to transform into a significant challenge and a critical health issue for all of humanity. The 1 mo mortality rate after hip fracture surgery is approximately 5% to 8%[6], while the cumulative mortality rate at 12 mo is 33%[7]. In addition to the high mortality rate, hip fractures threaten patients’ functional status in all aspects[8,9], especially ambulation ability in the elderly[10]. Even among those who lived in the community before their hip fracture, 11% ended up as bedridden, 20% did not regain normal walking ability, fewer than 30% returned to their previous level of activity, and 80% used walking aids for a year after their hip fracture[1,11].

Depressive symptoms are also prevalent among the old[12], and the onset of depression is obviously associated with a sudden injury or illness that can lead to hospitalization. This is also how postoperative hip fracture depression (PHFD) occurs. Wilson et al[13] also defined this as new-onset depression, a postoperative diagnosed disease following a depression-free preoperative period, which has been well described in certain cohorts. The prevalence of depressive symptoms is up to 50% among hospitalized older adults, and over one-third of the sample in the study by Ciro et al[14] reported high depressive symptoms during hospitalization. The incidence of depression is also high in patients with broken hips[15]. Information extracted from multiple studies[16-19] showed that the prevalence of depression after hip fracture ranged from 9% to 47%. In the systematic review and meta-analysis by Heidari et al[20], researchers concluded that the prevalence of depression among older hip fracture patients was 23% (95% confidence interval [CI]: 0.18-0.29). Many studies have proven that older adults with depression are associated with emotional suffering, rising health expenditures, higher risk of suicide, and diminishing quality of life (QOL)[21-23]. There is a clinical association between hip fracture and depression. Loss of walking

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: Although postoperative hip fracture depression (PHFD) has terrible effects on older patients, the risk factors, associated clinical outcomes, and treatments have not been systematically summarized. The comprehensive identification of risk factors is clinically useful in preoperative assessment and early prevention of PHFD. In addition, the summarized treatment options for PHFDs can suggest what circumstances treatment works best, providing a credible guide for effective management for clinicians.
ability, poor QOL, pain, and a sense of loss of future life caused by hip fracture are potential risk factors for depression.

This review explores the risk factors related to PHFD, its adverse clinical outcomes, and effective interventions by means of a literature search. It is expected to provide ideas for clinical workers to deal with the disease and provide direction for researchers.

MATERIALS AND METHODS
We referred to previously published literature search methods[24]. We used the PubMed, Web Of Science, EMBASE, and PsyclNFO databases to conduct a systematic review of the published literature and searched up until December 31, 2020. The search terms were hip fracture, femoral neck fracture or femoral intertrochanteric fracture and depression, and depressive disorder or low mood (search strategy, Figure 1). We researched reference lists of previously published reviews and studies in these reviews to ensure that no papers were missed. There were no restrictions on language or country of publication. After removing the same study from different databases, two reviewers (Qin HC, Luo ZW) assessed the titles and abstracts of initial papers independently and determined the final inclusion eligibility. Then either reviewer identified the full text of all possibly relevant records and included all subtypes of PHFD. Differences were settled by consensus (Figure 2).

RESULTS
How does PHFD develop (pathogenesis of PHFD)?
From the micro point of view regarding the formation of postoperative depression, postoperative peripheral inflammation can lead to an increase in proinflammatory cytokines in the brain in various ways, resulting in cognitive impairment[25]. In turn, these inflammatory factors continue to affect structures in the brain such as glia (microglia and astrocytes), the neurotrophic support of neurons, and changes in the cytotoxicity of neurons.

Looking at this from a macro perspective, the changes in all aspects of life brought about by a hip fracture operation have varying degrees of psychological impact. These include the fear of becoming too dependent on caregivers and placing too much burden on family members; fear that the body will not be able to function as before; loss of hope, confidence, or anxiety about future life; feeling lonely and isolated because their daily activities are restricted; and fear of falling again or even death. These psychological changes after surgery all contribute to the formation of depressive symptoms to a certain extent.

Risk of developing depression in hip fracture
In patients with PHFDs, there are many risk factors that contribute to the progression of the disease including age, sex, complications, types of fracture and preoperative status, and emotional-social support. With a thorough understanding of its risk factors, we can efficiently prevent and diagnose the disease in a timely manner.

Age and frailty: Epidemic data show the link between age and hip fractures; the number of hip fractures and the cost of treatment are increasing in China due to the aging of the population. In the study by Cheng et al[26], there were statistically significant differences in the incidence of hip fractures among different age groups. Meanwhile, as people grow older, loss of bone and muscle mass increases the risk of falling, leading to fractures[27]. Age is also an important factor in the development of depression. Both the prevalence and the incidence of major depression double after the ages of 70–85[28]. Combining these two factors, age as a risk factor for both hip fracture and depression, suggests that age is a potential risk factor for the development of PHFD. Williams et al[29] discovered that the odds ratios (ORs) for depression following fractures among younger women (≤ 65 years) were 0.62 (0.35 to 1.11; P = 0.12) and 3.33 (1.24 to 8.98; P = 0.02) for older women (> 65 years). However, new knowledge is emerging that frailty, an age-related cumulative decline in multiple physiological systems, is a better predictor of mortality and morbidity than chronological age[30]. At the same time, many of the items in the assessment of frailty are risk factors for fractures and depression[31]. The identification and assessment of frailty...
PHFD: Postoperative hip fracture depression. Risk factors:
1. Age and frailty
2. Gender
3. Complication
4. Types of hip fracture
5. Preoperative status and co-morbidity
6. Status of prognosis
7. Psychosocial factors

PHFD-associated clinical outcomes
1. Bad physical function
2. Higher pain severity
3. Poor treatment adherence
4. Extend medical cost
5. Increase the mortality
6. More postoperative infections

What can we do?
Treatment:
1. Nurse-led treatment?
2. Interdisciplinary intervention
3. Comprehensive care: Psychological and physical interventions

Figure 1 The full-text summary diagram. PHFD: Postoperative hip fracture depression.

may facilitate the identification of vulnerable preoperative patients so that management can be implemented in a timely manner. The association between the frailty index (as a risk factor) and the risk of PHFD is still unknown, and subsequent research is required to prove this association.

Sex: Among several factors that can influence the epidemiology of hip fractures, sex seems to be the most relevant factor. According to epidemiological data on depression in older adults, depression affects twice as many women as men[28]. Hip fracture is an important and debilitating condition in older people, especially women. Epidemiological data vary between countries, but it is globally estimated that hip fractures will have an influence on approximately 18% of women and 6% of men[8]. An essential reason is that functional outcomes reported between males and females in survivors and functional outcomes between men and women are most often reported as equivalent or favoring men, although men are consistently reported as having more postoperative complications[32]. For example, at discharge, men were more independent in locomotion, transfers, and sphincter control[33]. It is also possible that the difference is artificial and that sex differences in help-seeking behavior or symptom reporting exist to the extent that women are more likely to admit and complain about
their dysphoric feelings than men, who are more likely to deny and instead of act them out[34]. However, there is some controversy. For example, although the incidence is higher in women than in men, the severity of postoperative depression in men is often higher than in women, and the ratio of men to women varies by type of depression [35].

Types of hip fracture: Surgical treatment of hip fractures is mainly divided into internal fixation, hemiarthroplasty, and total arthroplasty. Studies have shown that sex and age are associated with different hip fractures[36]. Extracapsular-intertrochanteric fractures were found to increase dramatically with age in women. Since sex and age are risk factors for the development of PHFDs, the type of fracture also plays an important role. Moreover, the postoperative status and rehabilitation requirements for all kinds of T fractures vary, thus representing the differences in the recovery of patients’ postoperative activity ability, which has become one of the leading factors for the development of PHFDs. Elderly patients who receive internal fixation surgery for femoral intertrochanteric fracture do not recommend weight bearing of the affected hip joint at an early time of recovery. A considerable number of patients are still in bed or sitting in a chair for a long time within 1 mo after discharge. Therefore, displaced femoral neck fractures, treated with total hip arthroplasty or hemiarthroplasty, provide a better functional outcome than treatment with internal fixation[37].

Preoperative status and comorbidity: The preoperative status determines the maximum degree of recovery from surgery to some extent. In the study by Shyu et al [38], the authors found that those who were female and had lower prefracture activities of daily living performance were more likely to be at higher risk for depression. Poor nutritional status can also induce postoperative pain[39]. In addition to poor motor function and nutritional status being risk factors for the development of PHFDs, some articles found that chronic preoperative opioid use, history of smoking, drug or alcohol use disorder, and preoperative smoking were also risk factors[13,40]. Patients with underlying conditions are often more vulnerable to another condition.
The presence of comorbidities such as preoperative anxiety disorder[13], cardiovascular disease, diabetes, and cognitive dysfunction was associated with the incidence of postoperative depression[41]. In Miguel-Diez et al[42], the authors concluded that the incidence of hip fracture is higher in subjects with chronic obstructive pulmonary disease. Liu et al[43] discovered that predisposed patients with fewer comorbidities belong to the progressively lower-risk group rather than the fluctuating higher-risk group, which suggests that more comorbidities increase the risk of long-term depression following hip fracture. Unfortunately, after extensive literature research, no scale can be found to quantify the association between PHFD and preoperative comorbidities. Given that this may provide better risk stratification, this scale should be created in subsequent studies.

Status of prognosis: The quality of prognosis can be divided into acute and long-term outcomes, and both are sure to be risk factors accelerating depression. In the prognosis of the acute phase, pain is an important risk factor for PHFD. Numerous articles[44-46] have examined the mechanisms linking pain and depression, indicating that pain can remodel neural pathways to trigger negative symptoms of depression. Voshaar et al [15] identified postoperative pain and baseline anxiety as the strongest independent risk factors for depression. Bruggemann et al[47] provides tentative evidence that postinjury beliefs and hopelessness influence levels of depression and anxiety in hip fracture patients in the acute phase of their injury. Long-term risk factors such as loss of walking ability and reduced self-care ability can contribute to the progression of depression. Studies have also indicated a significant difference in the incidence of depression between homebound and semi-homebound elderly adults[48]. The psychological changes of the patient as they recover are also important[49]. Prolonged bedtime, reduced range of activity, and severe effects on the ability to perform daily activities are all susceptible to forming a sense of helplessness. In conclusion, each stage of the patient’s psychological and physical changes after surgery can affect the development of depression.

Psychosocial factors: Patients’ education, which is often overlooked, is linked to depression. In Liu et al[43], researchers distinguished groups of postoperative depressive-symptom trajectories, and found that patients in the progressively lower-risk group were less likely to be illiterate than those in the fluctuating higher-risk group (P = 0.02). Emotional support from family, friends, and society plays an important role in one’s psychological development, especially those who have experienced impaired health-related QOL[50]. Low socioeconomic status[51], living in a rural area, being unmarried, and housing circumstances promote depression disorder, although no direct experimental data have demonstrated an association between PHFD and these factors.

PHFD-associated clinical outcomes
Depression alone or hip fracture alone can lead to a decrease in QOL. The purpose of this part is to summarize the serious clinical outcomes of PHFDs. Exploring the impact of PHFDs on patients can help doctors provide more comprehensive treatment strategies.

PHFD has a great impact on physical function. There was a unanimous conclusion that during the first months after a hip fracture, the physical functioning of all the patients was seriously affected. At the same time, depression symptoms were associated with worse patient-reported outcomes[9,52,53]. The trial by Kohring et al [54] quantified the effect of depression on physical function recovery; untreated depressed patients demonstrated significantly smaller gains in physical function computerized adaptive test scores than nondepressed patients. The changes in preoperative and postoperative adjusted mean scores ranged from 37.8 to 45.1 in the nondepression group compared to 38.3 to 45.1 in the untreated depression group (P = 0.020). These results suggest that the recovery of physical function in patients with postoperative depression is very limited. Moreover, Gialanella et al[55] found that neuropsychiatric symptoms made the rehabilitation process slower and less efficient in hip fracture patients. The decline in physical function is bound to affect the QOL, in turn affecting the patient’s psychological changes. Therefore, it is necessary to put an end to this vicious cycle.

Affect postoperative physiological and psychological reactions: Some experiments have reported a statistically significant effect of depression on postoperative pain, resulting in aggravation of both conditions[56,57]. Afton et al[58] found that
improvements in depression and anxiety symptoms after surgery were associated with enhancements in pain severity and physical dysfunction. In Cristancho et al[59], researchers found that a group experiencing a persistently high level of depressive symptoms had poorer functional recovery and higher rates of pain in the year following hip fracture. However, there are new experimental data[60] showing that the relationship between pain and the geriatric depression scale (GDS) after hip fracture is not statistically significant. The effect of depression on postoperative pain needs to be further determined. In addition to causing pain, depression triggers complications after surgery. The analysis by Pan et al[61] of the relationship between psychiatric disorders and postoperative outcomes in patients undergoing primary total knee arthroplasty showed that patients with psychiatric disorders had higher ORs for most complications. Second fracture, rate of intensive care unit inpatient[62], postoperative infections[63], and chronic medical illness[64] are all linked to depression. In terms of psychological reaction, patients often show poor adherence, cognitive impairment[65], and poor motivation, which directly affect the rehabilitation and medication process.

**More medical support is required:** Sullivan et al[66] found that depression was significantly correlated with longer hospital longer limits of stability (LOS) while controlling for demographics and injury severity ($P = 0.023$) by looking at 460 subjects. Longer hospital stays[67] mean increased health care costs and a reduced flow of inpatients. Shelley et al[57] also statistically analyzed the relationship between injury- and hospital-related variables and posttraumatic stress disorder and depression at baseline and 12 mo after orthopedic trauma, showing that depression at baseline and 12 mo was associated with mechanical ventilation and LOS. In addition, the cost of treating a variety of complications mentioned in the previous section associated with postoperative depression accounts for a large part of the cost. Given increasing health care costs and the condition in which hospital beds are scarce, it is necessary to reduce the incidence of postoperative depression as much as possible to alleviate this situation.

**Increase the possibility of rehospitalization:** Gold et al[68] showed that overall, 90-d readmission rates were approximately 8% for total hip arthroplasty (THA) and total knee arthroplasty (TKA), while the readmission for subjects with depression was 21%-24% higher overall. Therefore, depression is related to a markedly higher risk of readmission after THA and TKA. Reduced readmissions can facilitate the rationalization of medical resources and exert apparent impacts on cost and quality. Hospital readmission rates are increasingly used as an indicator of health care quality and a potential source of regulatory penalties for hospitals[69]. At present, there are few studies on the direct effect of PHFDs on readmission, and further studies are needed.

Hip fracture is a common cause of mortality in older adults[70] and has a 1-year mortality of approximately 25%[71]. Older people with depression also have a higher mortality than average[28]. Both factors were independent predictors of mortality in elderly individuals. Liu et al[72] listed the risk factors for mortality in elderly patients with hip fractures, and depression was included. Nightingale et al[73] concluded that depressive symptoms increase the risk of mortality in patients with hip fracture in the following 2 years. By contrast, Hershkovitz et al[74] demonstrated that only dementia and age were independent predictors of mortality in hip fracture patients during the first 2 years after discharge from a post-acute rehabilitation program, while the presence of depression was not. As the most important clinical concern, patient mortality should be given sufficient attention.

**Why does PHFD have such clinical outcomes?** Patients with depression may feel hopeless, lack motivation, and be apathetic predisposing them to not participate actively in rehabilitation (adverse effects of depression and cognitive impairment on rehabilitation participation and recovery from hip fracture)[75]. Therefore, depression may exert its negative effect as a mediator via differential response to rehabilitation as well as other behaviour with a potential negative effect such as decreased activity, decreased social behaviour, and impaired diet.

Duggal et al[16,76-78] performed a series of studies and found that neutrophil bactericidal properties, monocyte and natural killer cell functioning, and T regulatory immune cells were impaired only in hip fracture patients who developed depressive symptoms. The dysregulation of these immune cells lead to a decline in immune system function and easily lead to postoperative infection.
People with depression tend to have higher cortisol levels than normal people. The higher the cortisol levels in older people, the frailer they are and the worse their physical function[79]. Low levels of serum anti-gluocorticoid dehydroepiandrosterone (DHEAS) sulfate have been associated with poorer physical function. Phillips et al[18] found that the ratio of cortisol to DHEAS was significantly higher in patients with PHFDs and was also associated with poorer physical function. The effects of depression on physical functioning may also be mediated by regulating these two substances.

DISCUSSION

Interventions for PHFD
To prevent the negative effects of depression on postoperative recovery and allow patients to recover to the preoperative physical condition as much as possible with less time, effort and cost, PHFD intervention is necessary. The interventions, inclusion criteria, exclusion criteria, groups, and main results included in each trial are listed in Figure 3[80-83,43].

One of the major reasons we focused on postfracture depression rather than preoperative depression is that the treatment of preoperative depression is similar to that of general depression. Depression in patients with PHFDs can be treated by both physical and psychological approaches because improvements in treating physical function can lead to better relief of depression, and early identification and intervention of risk factors for PHFDs is apparently an optimal choice compared with general depression treatments.

Nurse-led treatment and prevention for PHFDs are not significantly effective: Burns et al[80] wanted to determine whether a structured psychiatric intervention (nurse-led) is more effective than treatment as usual (TAU) in reducing depressive symptoms in older people who have depression after hip fracture surgery. Moreover, they explored whether a psychological intervention (cognitive behavior therapy) was more effective than TAU in preventing the development of depression in older people who were free from depressive symptoms immediately after surgery. This was a trial to determine if intervention for depression can treat and prevent its occurrence and progression. In the treatment study, they found a slight reduction in depressive symptoms. In prevention studies, there was no significant difference in the incidence of depression between psychological intervention and conventional treatment. There was also no difference in function or pain outcomes.

In the treatment-of-depression trial, 121 participants scored 7 or higher on the GDS and were allocated to nurse intervention (n = 61) or treatment as usual (n = 60). There were no differences in variables of interest between the two groups at baseline. After six weekly treatments, 28 (52%) in the control group remained depressed, compared with 17 (34%) in the intervention group. This univariate comparison was of only marginal statistical significance (P = 0.10). GDS scores were simultaneously slightly lower (adjusted 6 wk mean 6.3 [intervention group] vs 7.1 [control group], mean difference = 0.8; P = 0.11).

In a prevention-of-depression trial, 172 participants scored 6 or lower on the GDS, and 170 were allocated to the cognitive behaviour therapy (CBT) intervention or treatment as usual (n = 85 in each, two participants dropped out before randomization). There were no differences in variables of interest between the two groups at baseline. After the 6 wk follow-up, 4 (6%) of the intervention group had depression, compared with 11 (16%) in the treatment-as-usual group (P = 0.15).

In conclusion, the findings from the abovementioned trials have some implications for clinical practice. The findings of these two trials indicate that interventions do not have a distinct benefit in treating or preventing depression in older people after hip surgery. Neither therapy had a distinct effect on other variables of interest such as pain, functional recovery, or length of stay. Although the nurse-led intervention is ineffective in either preventing or treating depressive symptoms, this study confirmed the rates of depression after surgery for hip fracture found in previous studies and found that poor clinical outcome was related to factors other than depressed mood. Because they may cause lower human and financial expenditures, the therapeutic methods contained in it also have certain reference significance.

Interdisciplinary intervention decreases depressive symptoms, especially the risk of being persistently depressive: Interdisciplinary treatment involves several professionals who independently treat various issues a patient may have, focusing on the
Qin HC et al. Outcomes and interventions of PHFD

<table>
<thead>
<tr>
<th>Author</th>
<th>Intervention</th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
<th>Groups</th>
<th>Main results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns et al. 2007</td>
<td>Nurse-led treatment</td>
<td>1. Written two weeks of surgery for a femoral neck fracture</td>
<td>1. Significant cognitive impairment</td>
<td>240 Hospitalized</td>
<td>1. There was a slight reduction in depressive symptoms in the acute care of patients who misused</td>
<td>Nonsensical treatment do not have a significant benefit for treating prevention of depression in older adults after hip surgery.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Age over 60 yrs</td>
<td>2. Severe hearing impairment</td>
<td>200 Control group</td>
<td>2. In the prevention study, there was no significant difference in incident depression between psychological intervention and treatment as usual.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Patients with GDS scores for depression</td>
<td>3. Severe physical health</td>
<td>170 Risk of depression</td>
<td>3. There were no differences in the functional and mental outcomes.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Live too far away</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Language problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shyu et al. 2010</td>
<td>Interdisciplinary intervention</td>
<td>1. 60 yr older</td>
<td>1. Severely cognitively impaired and unable to follow orders</td>
<td>200 Randomized intervention group</td>
<td>The interdisciplinary intervention for hip fracture beneficial elderly persons with hip fracture by improving clinical outcomes, self-esteem, and physical Health.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Accidental slip-and-fall fracture</td>
<td>2. Unable to Communicate</td>
<td>200 Control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Recurrent hip fracture</td>
<td>3. Terminally ill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. CBI score ≥ 70</td>
<td>4. Admitted from a nursing home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shyu et al. 2012</td>
<td>Comprehensive interdinary care</td>
<td>1. 60 yr old</td>
<td>1. Severely cognitively impaired and unable to follow orders</td>
<td>200 Randomized intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Accidental slip-and-fall fracture</td>
<td>2. Unable to Communicate</td>
<td>200 Control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Recurrent hip fracture</td>
<td>3. Terminally Ill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. CBI score ≥ 70</td>
<td>4. Admitted from a nursing home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tseng et al. 2015</td>
<td>Interdisciplinary intervention</td>
<td>1. 60 yr old</td>
<td>1. Severely cognitively impaired and unable to follow orders</td>
<td>200 Randomized intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Accidental slip-and-fall fracture</td>
<td>2. Unable to Communicate</td>
<td>200 Control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Recurrent hip fracture</td>
<td>3. Terminally Ill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. CBI score ≥ 70</td>
<td>4. Admitted from a nursing home</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al. 2018</td>
<td>Comprehensive interdinary care</td>
<td>1. 60 yr old</td>
<td>1. Severely cognitively impaired and unable to follow orders</td>
<td>200 Randomized intervention group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Accidental slip-and-fall fracture</td>
<td>2. Unable to Communicate</td>
<td>200 Control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Recurrent hip fracture</td>
<td>3. Terminally Ill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. CBI score ≥ 70</td>
<td>4. Admitted from a nursing home</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3 General summary of interventions for postoperative hip fracture depression. The figures in the main results section above refer to Burns et al. [80], Shyu et al. [81,82], Tseng et al. [83] and Liu et al. [43]. CBI: Caring behavior inventory; GDS: Geriatric depression scale.

issues in which they specialize. Previous studies [84,85] have shown that interdisci- nancy interventions can effectively reduce depressive symptoms after hip fracture surgery in the short term. Shyu et al. [81] reported its long-term effects on variables such as self-reported and performance-based outcomes until the 2nd year after patient discharge. They hypothesized that the clinical benefits of the interdisciplinary intervention would persist into the 2nd year.

The results showed that depressive symptoms decreased significantly over the first 12 mo after discharge, especially within the 6th mo after discharge, which is consistent with the conclusion of previous experiments. Subjects in the intervention group obviously had fewer depressive symptoms than those in the control group (P = 0.05). Over the next 12 mo, although the two groups did not differ significantly in risk for depression at months 12, 18, and 24 after discharge, the severity of depressive symptoms at these time points was greater in the control group than in the intervention group.

In most studies of depressive symptoms among older adults with hip fracture, the focus was on the average change in depressive symptoms over time and how individuals deviate from this average. However, there are actually several broad categories of changes in each person’s depression trajectory. Tseng et al. [83] wanted to determine the distinct trajectories of depressive symptoms over time and the effects of the interdisciplinary intervention on these distinct courses of changes in depressive symptoms. Tseng performed a secondary analysis using data from previous clinical trials to confirm this hypothesis.
After the classification and correlation analyses of patients with different depression trajectories, it was found that participants in the experimental group were less likely to be persistently depressive over the 2-year period of observation, which means that interdisciplinary intervention reduces the risk of being persistently depressive among older patients with hip fracture. However, the interdisciplinary intervention did not apparently reduce the risk of being marginally depressive.

For interdisciplinary care, although there are multidisciplinary combination therapies, the actual benefits are controversial in some studies. Some experiments[86-88] have shown that compared with conventional geriatric care and rehabilitation, interdisciplinary home rehabilitation for older individuals following hip fracture resulted in similar proportions of complications, readmissions, and total days spent in the hospital after discharge. Interdisciplinary approaches to some aspects of patient recovery after hip surgery remain controversial.

**Comprehensive care is the optimal solution at present.** In recent years, increasing attention has been given to comprehensive nursing after hip fracture in the elderly [89]. At the same time, some articles[90-92] have proven that the comprehensive hip fracture program reduced the rate of in-hospital postoperative complications and mortality and improved physical function, role physical, general health, and mental health more than those who received usual care. Improvements in function and a focus on the psyche of the patient with comprehensive treatment can restore the patient’s mental state.

Shyu et al[82] filled a gap in research on the effectiveness of comprehensive therapy directly on depressive symptoms after hip fracture. They designed a study to compare the consequences of the comprehensive care model with those of the author’s previously developed interdisciplinary care model and usual care. They hypothesized that elderly patients who received the comprehensive intervention would have fewer depressive symptoms. After deriving ORs from regression coefficients, they found that the comprehensive care group was 52% less likely to be at risk for depression than the usual care group (OR = 0.48, P < 0.01). Similarly, the comprehensive care group was 49% less likely than the interdisciplinary care group (OR = 0.51, P < 0.05) to be at risk for depression. In conclusion, researchers’ comprehensive care programs with depression management appeared to be more beneficial than interdisciplinary care alone for older persons with hip fracture.

After determining that comprehensive care can reduce the probability risk for depression, Liu and Yang et al[43] conducted a secondary analysis of data on a subsample of hip fracture patients at risk for depression. The goal was to examine the associations of membership depressive symptom trajectory groups with receiving the comprehensive care model. Participants belonging to the better postoperative depressive-symptom trajectory group were hypothesized to have received the comprehensive care model. After using a series of analyses to fit trajectory models with two to five trajectory groups, the researchers found that the best fit to the observed data was the model containing two distinct depressive-symptom trajectory groups. To identify the predictors of group membership, they used binary logistic regression modelling and found that patients were likely to be in the progressively lower-risk group if they received comprehensive care than those who received interdisciplinary or usual care (OR = 2.78, 95% CI: 1.00–7.85, P = 0.05). Patients in the progressively lower-risk group (approximately 71%) experienced low levels of depressive symptoms that diminished over the 12 mo following surgery.

This analysis is based on a previous experiment, and the conclusion is drawn that comprehensive care featuring both psychological and physical interventions can not only reduce the potential risk for depression but also alleviate depressive symptoms in elderly hip fracture patients.

Both comprehensive intervention and interdisciplinary intervention include a postoperative bedside rehabilitation plan and a postdischarge rehabilitation plan at home. Rehabilitation treatment focuses on the recovery of body function. This is important because depressed patients’ functional recovery is mutually affected. Depression worsens the effectiveness of rehabilitation treatment, and poor consequences in turn continue to increase the psychological pressure of patients. This is sure to be a dilemma for most patients after fracture surgery. Therefore, in the treatment of PHFD patients, it is necessary to pay attention to the physiological and psychological recovery of patients. The treatment we give can only be meaningful if both are improved simultaneously. In addition to the rehabilitation program, only the comprehensive care models feature structured management of depressive symptoms, including patients with GDS-s scores > 10 (n = 4). Psychiatric consultation was arranged during hospitalization for diagnosis, suggested psychosocial interventions,
and medication management, and a home nurse followed up on patient appointment attendance and adherence to medication regimens. This nurse provided individual consultations, emotional support, and engaged family support for patients with GDS-s scores > 5 (n = 28). However, we cannot confirm the effect of depression management because of the difficulties of teasing out the impact of a single care component within a complex intervention program. Early assessment also plays an important role in both intervention models, and geriatric assessment in the early stage can yield information on developing tailored interventions by identifying the type of depression trajectory. For example, those whose GDS scores at discharge are within the 95% CI (6.5–8.5) may need further interventions for interventions for managing depressive symptoms.

The common shortcoming of both models is that the regions we included in the study cannot generalize the global situation, and the effects of the intervention in this study might not be observed if implemented in Western countries or in Japan, limiting the generalizability of the findings to these nations. A comparison of the three treatment strategies is shown in Table 1.

Management of medication: In terms of medication management, concerns here are that antidepressants can increase the risk of fractures and reduce physical function. Agarwal et al[93] compared physical function in current female antidepressant users and non-users and found that the grip strength of non-users was 13.3% lower than that of current users (P = 0.04). Wu et al[94] demonstrated that depression and antidepressant use are independently linked to an increased risk of fracture. Thus, antidepressant treatment might increase the risk of secondary fractures in patients with PHFDs. Specifically, selective serotonin reuptake inhibitor (SSRI) antidepressants have been associated with reduced bone mineral density, and an increased risk of fragility fracture and tricyclic antidepressants may increase the risk of fracture via mechanisms rather than an influence on bone mineral density. However, data on the effects of other antidepressants on bone is insufficient[95]. Older patients, who are particularly vulnerable to osteoporosis, may be preferentially prescribed non-SSRI antidepressants. In conclusion, determining whether antidepressant treatment is required and which antidepressant should be prescribed for specific PHFD patients remains unknown. However, at minimum, PHFD patients should be tested for bone mineral density before medication is prescribed. Subsequent studies should compare the risk-benefit ratio for PHFD patients who take antidepressants to those who do not.

Psychological intervention: To date, we have not found an article demonstrating a specific psychotherapy for PHFDs or a controlled trial to verify its effectiveness. However, given the characteristics of patients after hip fracture (i.e. most patients are frail and elderly), some psychological treatment methods for elderly patients with depression can be appropriately adopted. Commonly used and proven psychotherapeutic methods include CBT, behavioral activation therapy, interpersonal psychotherapy, problem-solving therapy (PST), life-review therapy, and nondirective counseling[96]. CBT is a widely used technique in the treatment of depression, and its methods are constantly being innovated to apply to different types of depression patients. However, its core is the therapist through therapeutic interactions and topics of discussion, using a psychoeducational approach, and teaching patients new ways of coping with stressful situations to modify the patient’s dysfunctional beliefs (cognitive restructuring) and change their current behaviour as well as future functioning. Huang et al[97] demonstrated the effectiveness of CBT through a three-step process in elderly individuals: behavioral scheduling, cognitive assessment and restructuring, and altering dysfunctional beliefs and performing problem-solving strategies. In PST, patients learn how to systematically solve their problems in a number of steps. Kisses et al[98] integrated a problem-solving approach (situation selection, situation modification, attentional deployment, cognitive change, and response modulation) with compensatory strategies, environmental adaptations, and caregiver participation to improve emotional regulation in 74 older adults with depression and found that the participants in problem adaptation therapy had a significantly greater reduction in depression. Life-review therapy is mainly used in older adults and is aimed at balancing positive and negative reminiscences, redefining negative experiences, and elaborating memory. The results of Preschl et al[99] indicated that life-review therapy can be recommended for depressive older adults. In the field of psychotherapy for PHFDs, there is still an urgent need for a large amount of experimental data to support the effectiveness of these methods.
Table 1 Comparison of three care models

<table>
<thead>
<tr>
<th>Content</th>
<th>Usual care</th>
<th>Interdisciplinary care</th>
<th>Comprehensive care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geriatric consultation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rehabilitation program</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Range of motion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Muscle strength and endurance</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Proprioceptive enhancement</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Balance challenges</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Aerobic capacity</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>On-protocol in-hospital rehab</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Nurse rehab visit</td>
<td>Primary Registered nurse</td>
<td>Once per day</td>
<td>Once per day</td>
</tr>
<tr>
<td>Physical therapist</td>
<td>Two times</td>
<td>Two times</td>
<td>Two times</td>
</tr>
<tr>
<td>On-protocol in-home rehab</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Nurse rehab visit</td>
<td>8 visits in 4 mo</td>
<td>13 visits in 1 yr</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>3 visits in 4 mo</td>
<td>4 visits in 6 mo</td>
<td></td>
</tr>
<tr>
<td>Nutritional consultation/education</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Depression screening and management</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Fall prevention</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Discharge planning with post-hospital services</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Modifications (medication recommendations, personal or environmental safety)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

CONCLUSION

Our study found that PHFD, if left untreated, has a significant impact on postoperative recovery. Therefore, preoperative identification of risk factors and postoperative intervention on depressive symptoms themselves are important. Given the current research, patients who have PHFDs may benefit from a comprehensive postoperative care model that features both psychological and physical interventions and an interdisciplinary care model. In addition, interdisciplinary approaches may also be a reasonable choice. More high-quality randomized controlled studies are required to validate current models and explore more effective interventions to further intervene in PHFD.

In this review, we introduced PHFDs and summarized their risk factors and clinical outcomes. In addition, we presented a summary of all previous interventions for depression after hip fracture and summarized the methods and experimental results of interdisciplinary treatment, comprehensive treatment, nurse-led treatment, and usual treatment for doctors' reference. Hip fractures are very common in older people and are associated with high rates of disability and mortality. However, depression following joint fracture surgery is also common and places a greater burden on the patient's recovery. Therefore, it is necessary to have a deep understanding of this kind of concomitant disease and to make correct diagnoses and interventions for a better prognosis.

Serious clinical outcomes may occur if PHFD is not dealt with in a timely manner, which emphasizes more importance in the assessment of risk factors to prevent PHFD. If multiple risk factors exist, then patients should be classified as high-risk, and the depression scale should be evaluated after surgery. This is of great significance for the early identification and intervention of PHFD patients. Therefore, the PHFD scale for preoperative risk factors should be reasonably used in patients. Since there is no unified version of the scale at present, future studies can focus on this aspect.

Patients with high-risk factors should receive appropriate intervention before surgery, rather than after the onset of postoperative depressive symptoms, when it may be too late, the cost of treatment increases, and the effectiveness of treatment
Figure 4 Hypothesis of clinical diagnosis and treatment of postoperative hip fracture depression.

decreases. Subsequent studies may focus on preoperative intervention on the mechanism of PHFD-induced poor outcomes.

What if, after all that, depression still occurs? At present, it seems that interdisciplinary treatment and comprehensive treatment combined with medication therapy are appropriate choices. (See the hypothesis of clinical diagnosis and treatment of PHFD in Figure 4).

However, future research should explore whether global effectiveness means fitting different countries and regions well. To further verify that the treatment regimen is suitable outside of China, it can be tested in countries with similar national conditions, such as Cambodia and India. At the same time, it can also be tried in smaller developed countries such as Singapore to verify the effect of this treatment on countries with different economic development. In this process, treatment patterns will evolve based on local characteristics and eventually produce a version that works for the entire world. A final point should not be overlooked: we need to expand the scope of the research to explore more effective interventions to further treat PHFDs.

ARTICLE HIGHLIGHTS

Research background
The number of hip patients is on the rise, while the number of postoperative depressions is also increasing. The coexistence of the two diseases makes it more difficult for patients to recover their mental and physical functions after surgery, making this a worldwide medical problem.

Research motivation
Clinically, we observed that depression after hip fracture in the elderly is very common and has caused a great burden on patients’ families and society. Therefore, we wondered if we could find the best intervention to solve the problem.

Research objectives
The aim was to make doctors more aware of the serious consequences of postoperative hip fracture depression (PHFD). In addition, we hoped to provide a basis for the early detection of PHFDs and recommend treatment methods after occurrence of the disease.
Research methods
We searched four databases for articles related to new depression after hip fracture and classified similar articles.

Research results
In this article, we summarize the risk factors for PHFD and the associated adverse outcomes, which can help to raise awareness of the disease and help with its early diagnosis. At the end of this paper, we summarize the practical and effective intervention methods in clinical practice: Interdisciplinary treatment and comprehensive treatment.

Research conclusions
Preoperative identification of risk factors and postoperative intervention on depressive symptoms themselves are important. In addition, PHFDs may benefit from a comprehensive postoperative care model that features both psychological and physical interventions and an interdisciplinary care model.

Research perspectives
Future research should explore global effectiveness in working with different countries and regions. Moreover, the following need should not be overlooked: we need to expand the scope of the research to explore more effective interventions in order to further treat PHFDs.

REFERENCES


12 Mitchell PB, Harvey SB. Depression and the older medical patient—when and how to intervene. Maturitas 2014; 79: 153-159 [PMID: 24939806 DOI: 10.1016/j.maturitas.2014.05.010]


14 Ciro CA, Ottenbacher KJ, Graham JE, Fisher S, Berges I, Ostir GV. Patterns and correlates of


Qin HC et al. Outcomes and interventions of PHFD


77 Duggal NA, Upton J, Phillips AC, Hampson P, Lord JM. NK cell immunosenescence is increased by psychological but not physical stress in older adults associated with raised cortisol and reduced perforin expression. *Age (Dordr)* 2015; 37: 9748 [PMID: 25663421 DOI: 10.1007/s11357-015-9748-2]


Qin HC et al. Outcomes and interventions of PHFD

1. Qin HC, WJP. Outcomes and interventions of PHFD


15. Wu Q. Depression and antidepressant medications: both are linked to increased fracture risk. Osteoporos Int 2019; 30: 695-696 [PMID: 30488276 DOI: 10.1007/s00198-018-4785-1]


Effectiveness of cognitive behavioral therapy-based interventions on health outcomes in patients with coronary heart disease: A meta-analysis

Yan-Ni Li, Nicholas Buys, Samantha Ferguson, Zhan-Jiang Li, Jing Sun

Abstract

BACKGROUND
Recently, the efficacy of cognitive behavioral therapy (CBT)-based intervention on health outcomes in patients with coronary heart disease (CHD) has been recognized in randomized controlled trials (RCTs), but no comprehensive systematic review has been conducted. To address this research gap, our study aimed to evaluate whether comprehensive CBT-based interventions positively affect health outcomes in CHD patients. It was hypothesized that CBT-based interventions are effective in: (1) Reducing depression, anxiety, and stress symptoms; (2) Reducing body mass index, blood pressure, and lipid levels; and (3) Improving quality of life, and exercise endurance.

AIM
To verify the effectiveness of CBT-based interventions on CHD patients through a meta-analysis of previous publications.

METHODS
Relevant RCTs published in English were obtained by searching electronic databases, including PubMed, Embase, Cochrane Central Register of Controlled Trials, Scopus, and Proquest, with the retrieval time from inception to August 2020. The primary outcomes were psychological factors (depression, anxiety, and stress symptoms), physiological factors (body mass index, blood pressure, blood
Coronary heart disease (CHD), also known as ischemic heart disease, refers to arterial blockage and functional changes of the coronary artery, which then causes myocardial ischemia and hypoxia[1]. CHD patients are prone to have severe physical and mental problems and show the impaired health-related quality of life (QOL)[2], related to high mortality and additional cardiac events[3]. It has been reported that 7.4 million deaths worldwide were due to CHD[4]. Body mass index (BMI), blood pressure, total cholesterol (TC), low-density lipoprotein cholesterol, and glycosylated hemoglobin are important physiological factors observed in patients with CHD[5].

An increasing number of studies have shown that CHD is comorbid with the psychosomatic disease, and mental health plays a significant role in the development and progression of atherosclerosis and heart disease[6]. Doering et al[7] undertook a longitudinal cohort study with 2325 patients with stable CHD and found that depression and anxiety symptoms increased the risk of death in participants. Pająk et al[8] undertook a longitudinal cohort study with 2325 patients with stable CHD and found that depression and anxiety symptoms increased the risk of death in participants.
al[8] found a high prevalence of depression and anxiety in CHD patients, and that these psychological factors were negatively associated with favorable lifestyle changes, which were not conducive to CHD rehabilitation. A review by Cohen et al[9] indicated that chronic anxiety and exposure to daily stressors harmed cardiovascular health. The aggressive treatment of depression and anxiety is imperative because of the association of these factors with decreased medication compliance[10], lifestyle change adherence, and an increased risk of mortality[11].

Cognitive behavioral therapy (CBT) has been increasingly used to modify cardiovascular risk factors in psychosocial intervention programs[12,13] and to reduce psychosocial risk factors, including depression, anxiety, and loneliness, in patients with heart disease[14]. In general, CBT is a psychological therapy based on the hypothesis that cognitive processes affect emotion and behavior. Therefore, changing non-adaptive thought patterns and beliefs will lead to positive behavioral changes [15]. However, the effects of CBT on improving health outcomes in CHD patients are inconsistent. For instance, a cardiac rehabilitation program[16] showed that CBT was associated with significant improvement in depressive symptoms and high-density lipoprotein cholesterol (HDL-C). Similarly, Magán et al[17] found that psychological outcomes in CHD patients were improved when a CBT approach was implemented. However, Turner et al[18] found no differences between CBT and control groups on change in depression or anxiety symptoms in cardiac patients. Additionally, in most meta-analyses conducted to date, there has been no detailed analysis of the critical components of CBT affecting patients with CHD.

To address the research gap, this study aimed to evaluate whether CBT-based interventions are effective in reducing depression, anxiety and stress symptoms, BMI, blood pressure, and lipids level, improving QOL, and exercise endurance in CHD patients.

MATERIALS AND METHODS
We conducted this meta-analysis following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines. In addition, we registered the research protocol at the PROSPERO International Prospective Register of Systematic Reviews (Registration ID: CRD42020213587 PROSPERO 2020 website: https://www.crd.york.ac.uk/prospero/#recordDetails). We did not need to get Ethical Approval or Patient Consent for this study, because it was a systematic review and meta-analysis based on the previous publications.

Search strategy
The databases, including PubMed, Embase, Cochrane Central Register of Controlled Trials, Scopus, Proquest, were searched separately by two researchers. The keywords of "coronary heart disease" or "unstable angina" or "percutaneous coronary intervention" or "coronary artery bypass graft" or "acute coronary syndrome" or "myocardial infarction" and "cognitive behavioral therapy" were used to do the literature retrieval. We used the subject words + free words with Boolean operators AND/OR to search for target articles in abstracts, keywords, or titles, and the language was limited to English. In the search process, we modified the search terms appropriate for different search rules of various databases. The specific retrieval strategy was in the supplementary file. The reference lists of the publications were searched to find additional relevant articles. The retrieved time was restricted from its inception to August 2020. We used the Endnote X8 Software (Clarivate Analytics, PA, United States) to manage the collected literature. The literature search was conducted separately by two researchers. When there was a disagreement between the two researchers, a third researcher was invited to join in the discussion to resolve the disagreement.

Study selection
Inclusion criteria were developed using the PICOS approach: (1) P: The target population was adult patients aged 18 and above and with diagnosed CHD, regardless of disease stage and severity. Eligible patients included those with unstable angina, those who had undergone percutaneous coronary intervention or coronary artery bypass graft, and patients with the acute coronary syndrome or those who had suffered a myocardial infarction; (2) I: The principles of CBT had to be applied to the intervention, including both cognitive and behavioral techniques, such as recognition and diagnosis of automatic thinking, behavior experiment, behavior activation,
cognitive restructuring, and emotional modulation. A separate CBT or CBT combined with other methods was acceptable. The delivery of intervention could be face-to-face or remote, and the intervention form could be individual or group-based; (3) C: The control groups could be the non-CBT interventions, or usual care or waiting list. The usual care group was chosen if there were multiple comparison groups; (4) O: The main outcomes were psychological factors, including depression, anxiety, and stress symptoms; physiological factors, including BMI, blood pressure, lipids level: Low-density lipoprotein cholesterol, HDL-C, triglycerides, TC. The secondary outcomes were QOL and exercise endurance. 6-Minute Walk Distance Test was used to assess the exercise endurance. The article reporting one or more of the above outcomes was included; (5) S: We included the randomized controlled trial (RCT), with no limitation on follow-up time; (6) Only English articles; and (7) Accessibility of full-text publication.

Studies were excluded based on the following criteria: (1) Comprising solely cognitive or behavioral therapies; (2) No report on quantitative analyses; (3) All reviews and protocols; (4) Literature with incomplete data and inconsistent statistical methods; (5) Duplicate publications, we only included the most relevant study; (6) Low-quality studies, whose score of Physiotherapy Evidence Database tool (PEDro) ≤ 3; (7) Not English articles; or (8) Not a peer-reviewed journal article.

Data extraction
We extracted data by using a data-collection form, which included the following information: The name of the first author, publication year, research location, research subjects (sample size, age, sex, the type of diseases), research design, intervention method, control method, and outcome factors. Where data was incomplete, the corresponding author was contacted to obtain the data.

Quality assessment
The full texts of included articles were read by two reviewers independently and PEDro tool[19] was used to evaluate their methodological quality. There were ten items in PEDro tool: Subjects were randomly grouped; randomization was concealed; the baseline information between groups was similar; subjects, researchers, and assessors were blinded; low drop-out rate; “intention to treat” principle was used; the variability measures were used; between-group comparison methods were used. The quality of studies is categorized into three levels based on the ten items of PEDro tool: Research with eight points and above is high quality, with four to seven is medium quality, with three and less than three is low quality. The disagreements on the quality rating of the included study were resolved through consultation with a third researcher.

Statistical analysis
Review Manager 5.3 was used for the meta-analysis. We did a separate meta-analysis for each outcome variable. We used the pooled mean difference (MD) with a 95% confidence interval (CI) for continuous outcome variables and used the standardized MD (SMD) with 95%CI to measure their effect size. The effect size was small if SMD was between 0.2 and 0.5. When SMD ranged from 0.5 to 0.7, the effect size was medium. And the effect size was large when SMD was more than 0.7. The threshold was set as 0.05 and two-sided.

We used the Chi-square test and $I^2$ for heterogeneity testing. If $P < 50\%$ or $P > 0.05$, homogeneity was considered to be good among the included studies, and the fixed-effect model was adopted. If $P > 50\%$, studies were assessed as having a high level of heterogeneity, and the random-effect model was adopted.

Publication bias was evaluated using the Egger test using Stata Statistical Software 15. There was a statistical publication bias if the $P$ value was less than 0.05. In addition, we used the funnel plot to analyze whether there was publication bias if the number of meta-analysis studies exceeded 10.

Subgroup analyses were conducted for depression and anxiety symptoms in CHD patients to find the underlying factors that led to heterogeneity and to figure out the effective components of CBT. We evaluated the pooled MD for each subgroup and used the Q statistics to examine the differences in MD. The subgroups were as follow: The treatment form (individual or group); treatment course (< 12 wk or ≥ 12 wk); duration of a session (< 50 min or ≥ 50 min); the number of sessions (< 10 or ≥ 10); mode of delivery (face-to-face or remote); use of a CHD-specific manual (yes or no); drop-out rate (< 20% or ≥ 20%); the intervention group used specific techniques of CBT (yes or no). We categorized the specific techniques for CBT based on the Compre-
hensive Psycho-therapeutic Intervention Rating Scale and previous studies[20-23], which included psycho-education, behavioral strategies, cognitive strategies, affective strategies, interpersonal strategies, exposure, body-directed strategies, behavior experiment, mindfulness and attention, homework assignment. If the intervention technique was mentioned as an important technique, it was identified as "yes", otherwise it was identified as "no".

RESULTS

Literature search
As shown in Figure 1, we found 1293 articles, including 1289 articles from the database search and four articles from references in the publications. After removing duplicates, there were 916 articles left for the title and abstract screening. Of these, we discarded 793 publications that did not meet the inclusion criteria. Then, 123 articles were left for the full-text screening. Two researchers independently did the full-text evaluation in strict accordance with the inclusion rules. Through full-text screening, 101 articles were excluded, and the specific process of identifying relevant articles for inclusion in the systematic review and meta-analysis was described in Figure 1, resulting in 22 studies being included in this meta-analysis.

Study characteristics
Full details of the included studies were displayed in Table 1. Four thousand nine hundred and ninety-one participants were included in the 22 RCTs: 2443 in the intervention group and 2548 in the control group. The mean age of the intervention group was 59.57 and 59.85 in the control group. Five[24-28] of these studies were carried out in developing countries and 17[14,16,18,29-42] in developed countries. The intervention group used CBT or CBT-based intervention and the control group adopted usual care methods activity methods. Five studies[34,38-41] used remote interventions, such as through the internet or telephone, and 17[14,16,18,24-33,35-37,42] used traditional face-to-face interventions. Twelve[16,24-26,29,32-34,38-40,42] studies used an individual CBT intervention, eight studies[18,27,30,31,35-37,41] used a group-based CBT intervention, and one[14] used both. Only six studies[18,24,29,33,37,41] reported using an instruction manual. Five types of outcomes were included in the analysis: physiological factors, psychological factors, QOL, exercise endurance, and cardiac rehabilitation factors. Six studies[16,31,34,37,41,42] reported physiological factors, 20 studies[14,16,18,25-26,29,32-34,39,40,42] reported psychological factors, 10 studies[24,25,30-34,39,40,42] reported QOL, two[37,41] reported exercise endurance, and four[16,24,31,37] reported cardiac rehabilitation factors. However, due to the differences in cardiac rehabilitation factors, a meta-analysis could not be performed. The mean drop-out rate for the intervention group was 12.49% and 11.09% for the control group.

Post- to pre-treatment effects of CBT-based intervention
Effects on psychological factors: Eighteen studies[14,16,18,25,27,29-36,38-42] analyzed the effect of CBT-based intervention on depression symptoms in 3579 participants. The number of intervention and control group participants was 1798 and 1781, respectively. Results showed a significant improvement in depression symptoms, with a reduction of depression symptoms of -2.08 (95%CI: -2.50 to -1.66, P < 0.001), and a small effect size of SMD -0.42 (95%CI: -0.60 to -0.24, P < 0.001). There was statistical heterogeneity across the 18 studies (I² = 83%, P < 0.001) (Table 2). The forest plot of the effect was shown in Figure 2A.

Subgroup analyses (Table 3) found that CBT-based intervention was more effective on improving depression symptoms when it was individual treatment form, when there was no CHD-specific manual, and when the drop-out rate was < 20%. CBT-based interventions statistically reduced depression symptoms when they emphasized psycho-education, behavioral and cognitive strategies as the core approaches. In addition, CBT-based intervention was more effective when it did not highlight affective or interpersonal strategies.

The effect on anxiety symptoms was analyzed in 14 studies[16,18,25,27-30,32,33,36,38,40-42] with 1293 patients. There were 639 patients in the intervention group and 654 in the control group. The meta-analysis showed that CBT-based interventions significantly alleviated anxiety symptoms, with a reduction of symptom mean of -2.07 (95%CI: -3.39 to -0.75, P = 0.002), and a small effect size of -0.42 (95%CI: -0.72 to -0.11, P = 0.008). The heterogeneity was significant among the studies (I² = 91%, P < 0.001).
Table 1 Characteristics of randomized controlled trials included in this meta-analysis

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Country</th>
<th>Sample size (IC)</th>
<th>Gender: female, n (%)</th>
<th>Age (mean ± SD)</th>
<th>CHD diagnosis</th>
<th>Intervention</th>
<th>Format</th>
<th>Manual</th>
<th>Control group</th>
<th>Outcome</th>
<th>Measurement</th>
<th>Drop-out rate (%)</th>
<th>Quality of article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barth et al[29], 2005</td>
<td>Germany</td>
<td>59 (27/32)</td>
<td>I: 5 (18.50) C: 9 (28.10)</td>
<td>I: 60.81 ± 11.06 C: 55.62 ± 10.05</td>
<td>MI; CABG; PTCA; Unstable Angina Pectoris</td>
<td>CBT based intervention</td>
<td>Individual</td>
<td>Yes</td>
<td>UC</td>
<td>(2)</td>
<td>BDI; HADS</td>
<td>0.00 C: 12.50</td>
<td>7 (moderate)</td>
</tr>
<tr>
<td>Beresnevaitė et al[24], 2016</td>
<td>Lithuania</td>
<td>150 (79/71)</td>
<td>I: 13 (30.20) C: 14 (30.40)</td>
<td>I: 56.70 ± 10.20 C: 59.10 ± 11.10</td>
<td>CABG; valve replacement; valve repair surgery</td>
<td>CBT</td>
<td>Individual</td>
<td>Yes</td>
<td>UC</td>
<td>(3), (5)</td>
<td>SF36</td>
<td>45.57 C: 35.21</td>
<td>4 (moderate)</td>
</tr>
<tr>
<td>Berkman et al[14], 2003</td>
<td>United States</td>
<td>2481 (1238/1243)</td>
<td>I: 532 (43.00) C: 552 (44.00)</td>
<td>I: 61.00 ± 12.60 C: 61.00 ± 12.50</td>
<td>MI</td>
<td>CBT</td>
<td>Individual + group</td>
<td>No</td>
<td>UC</td>
<td>(2)</td>
<td>BDI</td>
<td>7.51 C: 7.64</td>
<td>7 (moderate)</td>
</tr>
<tr>
<td>Blumenthal et al[13], 2016</td>
<td>United States</td>
<td>151 (76/75)</td>
<td>I: 31 (41.00) C: 24 (32.00)</td>
<td>I: 61.80 ± 10.80 C: 60.40 ± 10.60</td>
<td>ACS; stable angina; coronary revascularization</td>
<td>CR + SMT</td>
<td>Group</td>
<td>No</td>
<td>CR</td>
<td>(2), (3)</td>
<td>BDI-II; STAI; PSS; CHQ</td>
<td>3.95 C: 4.00</td>
<td>7 (moderate)</td>
</tr>
<tr>
<td>Claesson et al[31], 2006</td>
<td>Sweden</td>
<td>198 (101/97)</td>
<td>I: 101 (100.00) C: 97 (100.00)</td>
<td>I: 59.40 ± 9.30 C: 62.20 ± 7.70</td>
<td>AMI; CABG; coronary angioplasty; angina pectoris</td>
<td>CBT based intervention</td>
<td>Group</td>
<td>No</td>
<td>UC</td>
<td>(1), (2), (3), (5)</td>
<td>ELSS; CPRS-S-A</td>
<td>23.76 C: 15.46</td>
<td>6 (moderate)</td>
</tr>
<tr>
<td>Dao et al[32], 2011</td>
<td>United States</td>
<td>100 (50/50)</td>
<td>I: 11 (22.90) C: 10 (20.40)</td>
<td>I: 62.80 ± 11.80 C: 64.20 ± 11.90</td>
<td>A CAD diagnosis and were scheduled to undergo a first-time CABG</td>
<td>CBT</td>
<td>Individual</td>
<td>No</td>
<td>UC</td>
<td>(2), (3)</td>
<td>BDI-II; STAI SF12</td>
<td>4.00 C: 4.00</td>
<td>6 (moderate)</td>
</tr>
<tr>
<td>Freedland et al[33], 2009</td>
<td>United States</td>
<td>123 (41/42/40)</td>
<td>I: 23 (56.00) C1: 21 (50.00) C2: 17 (43.00)</td>
<td>I: 62.00 ± 11.00 C1: 59.00 ± 10.00 C2: 61.00 ± 9.00</td>
<td>CABG</td>
<td>CBT + UC</td>
<td>Individual</td>
<td>Yes</td>
<td>SSM+ UC; UC</td>
<td>(2), (3)</td>
<td>BDI; BAI PSS; SF36</td>
<td>2.44 C1: 11.90 C2: 7.50</td>
<td>7 (moderate)</td>
</tr>
<tr>
<td>Javahehri et al[34], 2020</td>
<td>United States</td>
<td>34 (18/16)</td>
<td>I: 7 (38.89) C: 2 (12.50)</td>
<td>I: 70.30 ± 10.00 C: 72.90 ± 9.20</td>
<td>MI; coronary artery revascularization; angiographically documented stenosis of a major coronary artery</td>
<td>General sleep education + web-based CBT</td>
<td>Individual</td>
<td>No</td>
<td>Wait-list</td>
<td>(1), (2), (3)</td>
<td>PHQ-8; Duke Health Profile</td>
<td>16.67 C: 11.11</td>
<td>5 (moderate)</td>
</tr>
<tr>
<td>Koertge et al[38], 2008</td>
<td>Sweden</td>
<td>247 (119/128)</td>
<td>I: 119 (100.00) C: 128 (100.00)</td>
<td>I: 61.36 ± 9.10 C: 62.73 ±</td>
<td>AMI; PCI; CABG</td>
<td>CBT based intervention</td>
<td>Group</td>
<td>No</td>
<td>UC</td>
<td>(2)</td>
<td>BDI</td>
<td>17.65 C: 20.31</td>
<td>7 (moderate)</td>
</tr>
</tbody>
</table>
The efficacy of cognitive behaviour therapy on CHD was studied by various researchers in different countries and settings.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Participants</th>
<th>Intervention</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
<th>Baseline 3</th>
<th>Baseline 4</th>
<th>Baseline 5</th>
<th>Baseline 6</th>
<th>Outcome 1</th>
<th>Outcome 2</th>
<th>Outcome 3</th>
<th>Outcome 4</th>
<th>Outcome 5</th>
<th>Outcome 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li et al [25], 2016</td>
<td>China</td>
<td>75 (38/37)</td>
<td>PCI</td>
<td>I: 52.40 ± 6.30</td>
<td>52.00 ± 6.20</td>
<td>MI; angiographically documented CHD</td>
<td>CBT + UC</td>
<td>Individual</td>
<td>No UC</td>
<td>HAM-D</td>
<td>HAM-A</td>
<td>CROQ-PTCA-POST</td>
<td>0.00</td>
<td>6 (moderate)</td>
<td></td>
</tr>
<tr>
<td>Merswolken et al [36], 2011</td>
<td>Germany</td>
<td>62 (30/32)</td>
<td>CBT based intervention</td>
<td>I: 62.50 ± 8.30</td>
<td>59.80 ± 7.50</td>
<td>CBTG</td>
<td>Group</td>
<td>No Blank control</td>
<td>HADS</td>
<td>I: 16.67</td>
<td>5 (moderate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murphy et al [37], 2013</td>
<td>Australia</td>
<td>275 (139/136)</td>
<td>MI; CABG; PCI</td>
<td>I: 58.02 ± 8.87</td>
<td>59.92 ± 9.27</td>
<td>MI; AP</td>
<td>CBT</td>
<td>Group</td>
<td>Yes UC</td>
<td>6 (moderate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nehra et al [26], 2012</td>
<td>India</td>
<td>50 (25/25)</td>
<td>N/A</td>
<td>I: 59.00 ± 8.30</td>
<td>58.00 ± 7.50</td>
<td>CABG; coronary angiography</td>
<td>CBTG</td>
<td>Individual</td>
<td>No UC</td>
<td>PSS</td>
<td>0.00</td>
<td>5 (moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nejati et al [27], 2020</td>
<td>Iran</td>
<td>51 (17/17)</td>
<td>CBTG</td>
<td>I: 62.50 ± 8.30</td>
<td>59.80 ± 7.50</td>
<td>MI; AP</td>
<td>CBT</td>
<td>Individual</td>
<td>No UC</td>
<td>HADS</td>
<td>16.91</td>
<td>7 (moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norlund et al [38], 2018</td>
<td>Sweden</td>
<td>239 (117/122)</td>
<td>MI</td>
<td>I: 58.00 ± 8.87</td>
<td>59.92 ± 9.27</td>
<td>MI</td>
<td>iCBT</td>
<td>Individual</td>
<td>No UC</td>
<td>BDI-II</td>
<td>17.95</td>
<td>7 (moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Neil et al [39], 2014</td>
<td>Australia</td>
<td>121 (61/60)</td>
<td>ACS; unstable angina CBT based intervention</td>
<td>I: 61.00 ± 10.20</td>
<td>58.90 ± 10.70</td>
<td>ACS; unstable angina</td>
<td>CBTG</td>
<td>Group</td>
<td>No CTG; UC</td>
<td>6 (moderate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rafanelli et al [16], 2020</td>
<td>Italy</td>
<td>100 (50/50)</td>
<td>MI; unstable angina CBT + WBT</td>
<td>I: 57.64 ± 9.99</td>
<td>60.02 ± 10.94</td>
<td>MI; unstable angina</td>
<td>CBTG</td>
<td>Individual</td>
<td>No CM</td>
<td>PHQ-9; GAD-7; DASS-21</td>
<td>13.11</td>
<td>9 (high)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schneider et al [40], 2020</td>
<td>Canada</td>
<td>53 (25/28)</td>
<td>MI; unstable angina iCBT</td>
<td>I: 56.72 ± 11.90</td>
<td>59.29 ± 6.93</td>
<td>MI; unstable angina</td>
<td>iCBT</td>
<td>Individual</td>
<td>No Wait-list</td>
<td>12.00</td>
<td>7 (moderate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turner et al [18], 2013</td>
<td>Australia</td>
<td>57 (25/32)</td>
<td>ACS; PCI; CABG CBT</td>
<td>I: 61.00 ± 11.00</td>
<td>62.00 ± 9.00</td>
<td>ACS; PCI; CABG</td>
<td>CBT</td>
<td>Group</td>
<td>Yes BI</td>
<td>BDI-II; HADS-A</td>
<td>16.00</td>
<td>7 (moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turner et al [41], 2014</td>
<td>Australia</td>
<td>42 (21/21)</td>
<td>AMI; CABG; PCI iCBT</td>
<td>I: 55.60 ± 8.80</td>
<td>C: 6 (moderate)</td>
<td>AMI; CABG; PCI</td>
<td>iCBT</td>
<td>Group</td>
<td>Yes UC</td>
<td>BDI-II; HADS-A</td>
<td>28.57</td>
<td>5 (moderate)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Efficacy of cognitive behaviour therapy on CHD

Zeighami et al. [28], 2018
Iran
90 (30/30/30) 1
I: 10 (33.33)
C1: 9 (29.00)
C2: 12 (40.10)
MI
57.00 ± 11.20
CBT
57.00 ± 11.20
EMDR (2)
BAI
57.00 ± 11.20
Zeighami et al [28], 2018 Iran 90 (30/30/30) 1 I: 10 (33.33) C1: 9 (29.00) C2: 12 (40.10) MI 57.00 ± 11.20 CBT 57.00 ± 11.20 EMDR (2) BAI 57.00 ± 11.20

Zetta et al. [42], 2011
United Kingdom
233 (116/117)
I: 31 (28.00)
C: 38 (35.00)
Angina
64.80 ± 10.04
C: 65.94 ± 9.96
CBT based intervention
Individual
No
UC
6.03 ± 1.40
7 (moderate)

1 I = Cognitive behavioral therapy (CBT) + usual care (UC); C1 = Supportive stress management (SSM) + UC; C2 = UC.
2 I = Cognitive-behavioral group intervention; C1 = Cognitive therapy in groups; C2 = UC.
3 I = CBT; C1 = Eye movement desensitization reprocessing; C2 = control group.

Outcome: (1) Physiological factors (BMI, blood pressure, blood lipid profile: LDL-C, HDL-C, TG, TC); (2) Psychological factors (depression, anxiety, and stress symptoms); (3) QOL; (4) Exercise endurance; and (5) Cardiac rehabilitation. UC: Usual care; SSM: Supportive stress management; EMDR: Eye movement desensitization reprocessing; CBTG: Cognitive-behavioral group intervention; MI: Myocardial infarction; AMI: Acute myocardial infarction; PCI: Percutaneous coronary intervention; CABG: Coronary artery bypass grafting; ACS: Acute coronary syndrome; PTCA: Percutaneous transluminal coronary angioplasty; SMT: Stress management training; WBT: Well-being therapy; CM: Clinical management; BI: No further intervention; NR: No report; BDI: Beck Depression Inventory; HADS: Hospital Anxiety and Depression Scale; SF-36: 36-Item Short Form Health Survey; BDI-II: Beck Depression Inventory II; STAI: State-Trait Anxiety Inventory; PSS: Perceived Stress Scale; GHQ: General Health Questionnaire; ELSS: The Everyday Life Stress Scale; CPRS-S-A: Comprehensive Psychopathological Rating Scale Self-Affective; SF-12: Short Form-12 Health Survey; BAI: Beck Anxiety Inventory; PHQ-8: Patient Health Questionnaire-8; HAM-D: Hamilton Depression Rating Scale; HAM-A: Hamilton Anxiety Scale; CROQ-PTCA-POST: Coronary Revascularization Outcome Questionnaire; 6MWT: 6-Minite Walk Distance Test; DASS-21: Depression Anxiety, Stress Scales; PHQ-9: Patient Health Questionnaire-9; SQ: Symptom Questionnaire; GAD-7: General Anxiety Disorder 7-Item.

(Table 2).

Similarly, subgroup analyses (Table 4) found that CBT-based intervention was more effective for anxiety symptoms when it was individual treatment form, longer duration of a session (≥ 50 min), less than 10 sessions (< 10), face-to-face treatment delivery, and no use of CHD-specific manual. As for CBT technique components, CBT-based intervention was more effective when it applied the following techniques as the core strategies: psycho-education, behavioral and cognitive strategies. In addition, CBT-based interventions that did not use interpersonal strategy as the core approach showed a better effect on reducing anxiety symptoms.

Five studies [26,27,30,33,40] with a total participant number of 369 analyzed the effect of CBT on stress symptoms. There were 184 patients in the intervention group and 185 in the control group. CBT-based interventions improved stress symptoms significantly with a pooled MD of -3.33 (95%CI: -4.23 to -2.44, \( P < 0.001 \)) and a medium effect size of -0.62 (-0.83 to -0.40, \( P < 0.001 \)). There was no significant heterogeneity between these studies (\( I^2 = 22\% \), \( P = 0.280 \)) (Table 2). Figure 2C presented the forest plot of the effect.

**Effects on physiological factors:** The effect of CBT-based intervention on BMI was analyzed in two studies [31,42] with 377 patients. There were 186 participants in the intervention group and 191 in the control group. Results showed that CBT-based
Table 2: Total effect of cognitive behavioral therapy-based interventions on depression, anxiety, stress symptoms, body mass index, blood pressure, total cholesterol, high-density lipoprotein cholesterol, quality of life, and 6-min walk test distance

<table>
<thead>
<tr>
<th>Index</th>
<th>Outcomes: Post-to pre-treatment effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Studies, n</td>
</tr>
<tr>
<td>Depression</td>
<td>18</td>
</tr>
<tr>
<td>Anxiety</td>
<td>14</td>
</tr>
<tr>
<td>Stress</td>
<td>5</td>
</tr>
<tr>
<td>BMI</td>
<td>2</td>
</tr>
<tr>
<td>Systolic pressure</td>
<td>4</td>
</tr>
<tr>
<td>Diastolic pressure</td>
<td>4</td>
</tr>
<tr>
<td>TC</td>
<td>2</td>
</tr>
<tr>
<td>HDL-C</td>
<td>2</td>
</tr>
<tr>
<td>Physical function</td>
<td>7</td>
</tr>
<tr>
<td>Mental function</td>
<td>5</td>
</tr>
<tr>
<td>6-minute walk test distance</td>
<td>2</td>
</tr>
</tbody>
</table>

*p < 0.05  
*p < 0.01  
*p < 0.001.

MD: Mean difference; SMD: Standard mean difference; BMI: Body mass index; TC: Total cholesterol; HDL-C: High-density lipoprotein cholesterol; CI: Confidence interval.

The effect of CBT-based intervention on blood pressure was analyzed in four studies[31,34,41,42]. There were 216 patients in the intervention group and 221 in the control group. The meta-analysis did not show a significant reduction either in systolic blood pressure (-1.82, 95%CI: -5.12 to 1.49, P = 0.280) or diastolic pressure (-0.95, 95%CI: -2.75 to 0.86, P = 0.300). Additionally, there was not significant heterogeneity for systolic blood pressure (I² = 0%, P = 0.550) (Table 2). Figure 3A presented the forest plot of the effect. Two studies[16,41] with 131 patients measured the effect on TC and HDL-C. There were 65 and 66 participants in the intervention and control group. We did not find a significant reduction in TC (0.15, 95%CI: -0.81 to 1.12, P = 0.760). Similarly, we did not find a significant improvement in HDL-C (0.07, 95%CI: -0.09 to 0.23, P = 0.390). There was not significant heterogeneity for TC (I² = 0%, P = 0.390) or HDL-C (I² = 0%, P = 0.870) (Table 2). Figure 3B and C presented the forest plot of the effect.

**Effects on QOL and exercise endurance:** The effect on physical function of the QOL was analyzed in seven studies[24,25,33,34,39,40,42] with 666 subjects. There were 332 and 334 patients in the intervention and control groups. Results showed that CBT-based interventions significantly improved physical function across seven studies with improved MD of 3.36 (95%CI: 1.63 to 5.10, P = 0.000) and a small effect size of 0.27 (0.12 to 0.43, P = 0.001). The heterogeneity between these studies was not significant (I² = 49%, P = 0.070) (Table 2). Figure 4A showed the forest plot of the effect.

The effect on the mental function of the QOL was analyzed in five studies[24,25,33,39,40] with 419 patients. There were 208 patients in the intervention group and 211 in the control group. The pooled MD of improvement of mental function for CBT was 6.91 (95%CI: 4.10 to 9.73, P < 0.001) and a large effect size of 0.71 (0.41 to 1.00, P <
### Table 3 Subgroup analysis on the effect of cognitive behavioral therapy-based interventions on depression symptoms

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Depressive symptom: Post- to pre-treatment effect</th>
<th>Treatment form</th>
<th>Duration of a session</th>
<th>Number of sessions</th>
<th>Treatment delivery way</th>
<th>Manual available</th>
<th>Drop-out</th>
<th>Psycho-education strategies</th>
<th>Behavioral strategies</th>
<th>Cognitive strategies</th>
<th>Affective strategies</th>
<th>Interpersonal strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Studies (n)</td>
<td>Participants (n)</td>
<td>P % (P value)</td>
<td>Q-test</td>
<td>MD (95%CI)</td>
<td>SMD (95%CI)</td>
<td>P (between)</td>
<td>Studies (n)</td>
<td>Participants (n)</td>
<td>P % (P value)</td>
<td>Q-test</td>
<td>MD (95%CI)</td>
</tr>
<tr>
<td>Treatment form</td>
<td>8</td>
<td>2581</td>
<td>72 (0.001)</td>
<td>24.86</td>
<td>-0.96 (-2.16, 0.24)</td>
<td>-0.17 (-0.32, -0.02) ^a</td>
<td>0.030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>10</td>
<td>1061</td>
<td>88 (&lt; 0.001)</td>
<td>75.51</td>
<td>-2.97 (-4.33, -1.62)</td>
<td>-0.66 (-0.97, -0.35)</td>
<td>0.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>11</td>
<td>3194</td>
<td>81 (&lt; 0.001)</td>
<td>51.97</td>
<td>-1.28 (-2.07, -0.50)</td>
<td>-0.34 (-0.54, -0.14) ^a</td>
<td>0.070</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 12 wk</td>
<td>6</td>
<td>288</td>
<td>68 (0.008)</td>
<td>15.69</td>
<td>-2.83 (-5.08, -0.57)</td>
<td>-0.54 (-1.05, -0.02) ^c</td>
<td>0.330</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 12 wk</td>
<td>11</td>
<td>2788</td>
<td>84 (&lt; 0.001)</td>
<td>62.43</td>
<td>-2.25 (-3.68, -0.81)</td>
<td>-0.34 (-0.55, -0.13) ^b</td>
<td>0.410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50 min</td>
<td>3</td>
<td>439</td>
<td>13 (0.320)</td>
<td>2.30</td>
<td>-0.86 (-1.56, -0.35)</td>
<td>-0.46 (-1.11, 0.19)</td>
<td>0.610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 50 min</td>
<td>11</td>
<td>2788</td>
<td>84 (&lt; 0.001)</td>
<td>62.43</td>
<td>-2.25 (-3.68, -0.81)</td>
<td>-0.34 (-0.55, -0.13) ^b</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 10</td>
<td>9</td>
<td>2876</td>
<td>82 (&lt; 0.001)</td>
<td>43.75</td>
<td>-1.62 (-2.76, -0.49)</td>
<td>-0.28 (-0.45, -0.12) ^b</td>
<td>0.440</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 10</td>
<td>6</td>
<td>288</td>
<td>68 (0.008)</td>
<td>15.69</td>
<td>-2.83 (-5.08, -0.57)</td>
<td>-0.54 (-1.05, -0.02) ^c</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote</td>
<td>5</td>
<td>473</td>
<td>81 (0.000)</td>
<td>20.63</td>
<td>-2.73 (-4.78, -0.67)</td>
<td>-0.58 (-0.97, -0.19) ^b</td>
<td>0.610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face to face</td>
<td>13</td>
<td>3106</td>
<td>85 (&lt; 0.001)</td>
<td>77.45</td>
<td>-1.77 (-2.78, -0.77)</td>
<td>-0.37 (-0.59, -0.16) ^c</td>
<td>0.410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>14</td>
<td>3367</td>
<td>83 (&lt; 0.001)</td>
<td>76.32</td>
<td>-1.76 (-2.56, -0.96)</td>
<td>-0.42 (-0.62, -0.23) ^c</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>197</td>
<td>34 (0.220)</td>
<td>1.51</td>
<td>-1.31 (-4.21, 1.59)</td>
<td>-0.23 (-0.51, 0.05)</td>
<td>0.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 20%</td>
<td>16</td>
<td>3382</td>
<td>85 (&lt; 0.001)</td>
<td>100.58</td>
<td>-2.10 (-3.04, -1.17)</td>
<td>-0.44 (-0.64, -0.24) ^c</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20%</td>
<td>2</td>
<td>212</td>
<td>84 (0.000)</td>
<td>19.24</td>
<td>-3.27 (-9.09, 2.55)</td>
<td>-0.39 (-1.07, 0.30)</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>15</td>
<td>3367</td>
<td>82 (&lt; 0.001)</td>
<td>76.48</td>
<td>-1.78 (-2.59, -0.97)</td>
<td>-0.40 (-0.59, -0.21) ^c</td>
<td>0.670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not important</td>
<td>3</td>
<td>212</td>
<td>92 (&lt; 0.001)</td>
<td>24.52</td>
<td>-4.60 (-11.79, 2.58)</td>
<td>-0.54 (-1.37, 0.28)</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>13</td>
<td>3315</td>
<td>87 (&lt; 0.001)</td>
<td>95.77</td>
<td>-2.48 (-3.47, -1.46)</td>
<td>-0.52 (-0.73, -0.31) ^c</td>
<td>0.360</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not important</td>
<td>5</td>
<td>264</td>
<td>11 (0.340)</td>
<td>4.50</td>
<td>-0.30 (-1.45, 0.85)</td>
<td>-0.05 (-0.30, 0.20)</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>16</td>
<td>3448</td>
<td>85 (&lt; 0.001)</td>
<td>99.64</td>
<td>-2.11 (-2.99, -1.23)</td>
<td>-0.45 (-0.64, -0.26) ^c</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not important</td>
<td>2</td>
<td>131</td>
<td>53 (0.150)</td>
<td>2.11</td>
<td>-1.18 (-5.38, 3.03)</td>
<td>-0.14 (-0.65, 0.37)</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>2</td>
<td>413</td>
<td>0 (0.680)</td>
<td>0.18</td>
<td>-0.49 (-1.32, 0.34)</td>
<td>-0.09 (-0.29, 0.10)</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not important</td>
<td>16</td>
<td>3166</td>
<td>85 (&lt; 0.001)</td>
<td>100.55</td>
<td>-2.32 (-3.30, -1.33)</td>
<td>-0.47 (-0.68, -0.27)</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td>4</td>
<td>2422</td>
<td>80 (0.002)</td>
<td>15.31</td>
<td>-1.29 (-2.67, 0.09)</td>
<td>-0.21 (-0.37, -0.05) ^b</td>
<td>0.670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not important</td>
<td>14</td>
<td>1157</td>
<td>84 (&lt; 0.001)</td>
<td>80.72</td>
<td>-2.40 (-3.55, -1.25)</td>
<td>-0.52 (-0.79, -0.24)</td>
<td>0.230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Body directed strategies

<table>
<thead>
<tr>
<th></th>
<th>Important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>2876</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>70 (0.001)</td>
<td>26.46</td>
</tr>
<tr>
<td></td>
<td>-1.20 (-1.91, -0.49)</td>
<td>-0.33 (-0.53, -0.12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homework assignment strategies</td>
<td>0.070</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Important</th>
<th>Not important</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>2314</td>
<td>1265</td>
</tr>
<tr>
<td></td>
<td>83 (&lt; 0.001)</td>
<td>80 (&lt; 0.001)</td>
</tr>
<tr>
<td></td>
<td>29.48</td>
<td>55.64</td>
</tr>
<tr>
<td></td>
<td>-2.07 (-3.75, -0.38)</td>
<td>-1.93 (-2.96, -0.90)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[^a]: P < 0.05.
[^b]: P < 0.01.
[^c]: P < 0.001.

MD: Mean difference; SMD: Standard mean difference; CI: Confidence interval.

0.001). There was significant heterogeneity between these studies (I² = 52%, P = 0.080) (Table 2). The forest plot of the effect was presented in Figure 4B.

Two studies[^37,^41] with 279 patients measured the change in exercise endurance. The number of intervention and control group participants was 146 and 133, respectively. We did not find a significant distance change between CBT-based interventions and control conditions (9.64, 95%CI: -11.80 to 31.08, P = 0.380). There was no significant heterogeneity between these studies (I² = 0%, P = 0.550). The forest plot of the effect was presented in Figure 4C.

Risk of bias and quality assessment

The PEDro tool was used to evaluate the methodology quality of included studies, with one being high quality, accounting for 4.55%, and the remaining 21 being medium quality, accounting for 95.45%. Specifically, one study[^39] scored 9, 11 studies[^14,^16,^18,^29,^30,^33,^35,^37,^38,^40,^42] scored 7, four studies[^25,^27,^31,^32] scored 6, four studies[^26,^34,^36,^41] scored 5, and two studies[^24,^28] scored 4.

Publication bias

There was no T or P value of Egger analysis for BMI, TC, and HDL-C, because only two studies reported the results of CBT-based interventions on these variables. In Table 5, there was a minimal publication bias in the following outcome variables: depression symptoms (P = 0.467), anxiety symptoms (P = 0.196), systolic pressure (P = 0.516), diastolic pressure (P = 0.836), physical function (P = 0.829), and mental function (P = 0.135). However, the publication bias for stress symptoms (P = 0.021) was significant. We used the one-study-removed methods to evaluate the sensitivity, and we found that the overall results for all outcome variables were not changed when removing one study at a time. The funnel plots were in the Figure 5.

DISCUSSION

The results of our meta-analysis showed that CBT-based interventions were superior to the usual care controls. They reduced depression, anxiety, and stress symptoms, BMI levels, and improved physical and mental functioning in relation to the QOL. In addition, our subgroup analyzed several key identified components of CBT interventions that influenced their effect.

Post- to pre-treatment effects of CBT-based intervention

Effects on Psychological factors: Similar to previous studies[^30,^40], our study found that CBT-based intervention effectively reduced depression, anxiety, and stress symptoms in CHD patients compared with the control group. CBT-based interventions improve patients’ knowledge about CHD through psychological education, and the corresponding cognitive and behavioral techniques contribute to the changes in non-adaptive beliefs of patients[^43]. As for CHD patients with depression symptoms, their negative thinking mode can be altered through cognitive reconstruction[^44]. For anxiety and stress emotions, the excitability of sympathetic
### Table 4 Subgroup analysis on the effect of cognitive behavioral therapy-based interventions on anxiety symptoms

<table>
<thead>
<tr>
<th>Subgroups</th>
<th>Anxiety symptom: Post- to pre-treatment effect</th>
<th>Studies (n)</th>
<th>Participant (n)</th>
<th>I² (%) (P value)</th>
<th>Q-test</th>
<th>MD (95%CI)</th>
<th>SMD (95%CI)</th>
<th>P value (between)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment form</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td>5</td>
<td>320</td>
<td>50 (0.090)</td>
<td>7.96</td>
<td>-1.24 (-2.63, 0.15)</td>
<td>-0.27 (-0.57, 0.04)</td>
<td>0.100</td>
</tr>
<tr>
<td>Individual</td>
<td></td>
<td>8</td>
<td>913</td>
<td>94 (&lt; 0.001)</td>
<td>123.83</td>
<td>-3.25 (-5.17, -1.32) b</td>
<td>-0.62 (-1.05, -0.18) b</td>
<td></td>
</tr>
<tr>
<td><strong>Treatment course</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 12 wk</td>
<td></td>
<td>8</td>
<td>935</td>
<td>73 (0.001)</td>
<td>25.76</td>
<td>-0.73 (-1.69, 0.23)</td>
<td>-0.16 (-0.40, 0.08)</td>
<td>0.280</td>
</tr>
<tr>
<td>&lt; 12 wk</td>
<td></td>
<td>5</td>
<td>261</td>
<td>90 (&lt; 0.001)</td>
<td>38.38</td>
<td>-2.47 (-5.48, 0.54)</td>
<td>-0.63 (-1.37, 0.12)</td>
<td></td>
</tr>
<tr>
<td><strong>Duration of session</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50 min</td>
<td></td>
<td>9</td>
<td>652</td>
<td>89 (&lt; 0.001)</td>
<td>71.35</td>
<td>-2.49 (-4.88, -0.09) a</td>
<td>-0.43 (-0.86, -0.00) a</td>
<td>0.050</td>
</tr>
<tr>
<td>&lt; 50 min</td>
<td></td>
<td>2</td>
<td>318</td>
<td>0 (0.540)</td>
<td>0.38</td>
<td>-0.12 (-0.35, 0.11)</td>
<td>-0.13 (-0.35, 0.09)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of session</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 10</td>
<td></td>
<td>6</td>
<td>617</td>
<td>79 (0.000)</td>
<td>23.89</td>
<td>-0.98 (-2.77, 0.80)</td>
<td>-0.18 (-0.54, 0.18)</td>
<td>0.150</td>
</tr>
<tr>
<td>&lt; 10</td>
<td></td>
<td>7</td>
<td>458</td>
<td>92 (&lt; 0.001)</td>
<td>70.66</td>
<td>-3.50 (-6.42, -0.58) a</td>
<td>-0.66 (-1.23, -0.10) a</td>
<td></td>
</tr>
<tr>
<td><strong>Treatment delivery way</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote</td>
<td></td>
<td>4</td>
<td>383</td>
<td>94 (&lt; 0.001)</td>
<td>51.28</td>
<td>-0.59 (-4.68, 3.51)</td>
<td>-0.29 (-1.17, 0.59)</td>
<td>0.390</td>
</tr>
<tr>
<td>Face to face</td>
<td></td>
<td>10</td>
<td>910</td>
<td>90 (&lt; 0.001)</td>
<td>88.49</td>
<td>-2.54 (-4.17, -0.90) b</td>
<td>-0.47 (-0.78, -0.16) b</td>
<td></td>
</tr>
<tr>
<td><strong>Manual available</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td>4</td>
<td>214</td>
<td>50 (0.110)</td>
<td>6.03</td>
<td>-0.86 (-2.73, 1.02)</td>
<td>-0.19 (-0.52, 0.13)</td>
<td>0.200</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>10</td>
<td>1079</td>
<td>93 (&lt; 0.001)</td>
<td>135.85</td>
<td>-2.48 (-4.12, -0.84) c</td>
<td>-0.52 (-0.91, -0.12) c</td>
<td></td>
</tr>
<tr>
<td><strong>Psycho-education strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>10</td>
<td>1021</td>
<td>93 (&lt; 0.001)</td>
<td>128.71</td>
<td>-2.71 (-4.29, -1.13) c</td>
<td>-0.57 (-0.95, -0.19) c</td>
<td>0.140</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>4</td>
<td>272</td>
<td>77 (0.004)</td>
<td>13.30</td>
<td>-0.14 (-3.14, 2.86)</td>
<td>-0.03 (-0.54, 0.48)</td>
<td></td>
</tr>
<tr>
<td><strong>Behavioral strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>9</td>
<td>998</td>
<td>94 (&lt; 0.001)</td>
<td>134.13</td>
<td>-3.46 (-5.33, -1.60) c</td>
<td>-0.67 (-1.07, -0.27) c</td>
<td>0.003</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>5</td>
<td>295</td>
<td>47 (0.110)</td>
<td>7.57</td>
<td>0.13 (-1.27, 1.54)</td>
<td>0.05 (-0.26, 0.36)</td>
<td></td>
</tr>
<tr>
<td><strong>Cognitive strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>11</td>
<td>1102</td>
<td>93 (&lt; 0.001)</td>
<td>134.22</td>
<td>-2.84 (-4.37, -1.31) c</td>
<td>-0.57 (-0.91, -0.22) c</td>
<td>0.020</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>3</td>
<td>191</td>
<td>72 (0.030)</td>
<td>7.26</td>
<td>1.06 (-1.95, 4.06)</td>
<td>0.16 (-0.37, 0.69)</td>
<td></td>
</tr>
<tr>
<td><strong>Interpersonal strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>2</td>
<td>390</td>
<td>46 (0.180)</td>
<td>1.84</td>
<td>-0.59 (-2.50, 1.32)</td>
<td>-0.09 (-0.31, 0.13)</td>
<td>0.190</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>12</td>
<td>903</td>
<td>92 (&lt; 0.001)</td>
<td>140.01</td>
<td>-2.30 (-4.01, -0.59) b</td>
<td>-0.48 (-0.86, -0.10) b</td>
<td></td>
</tr>
<tr>
<td><strong>Body directed strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>5</td>
<td>694</td>
<td>67 (0.020)</td>
<td>12.22</td>
<td>-0.64 (-1.51, 0.22)</td>
<td>-0.19 (-0.43, 0.05)</td>
<td>0.150</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>9</td>
<td>599</td>
<td>91 (&lt; 0.001)</td>
<td>87.09</td>
<td>-2.73 (-5.43, -0.02)</td>
<td>-0.50 (-1.01, 0.00)</td>
<td></td>
</tr>
<tr>
<td><strong>Homework assignment strategies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
<td>5</td>
<td>529</td>
<td>91 (&lt; 0.001)</td>
<td>45.80</td>
<td>-2.60 (-5.42, 0.23)</td>
<td>-0.61 (-1.17, -0.04) c</td>
<td>0.640</td>
</tr>
<tr>
<td>Not important</td>
<td></td>
<td>9</td>
<td>764</td>
<td>91 (&lt; 0.001)</td>
<td>84.86</td>
<td>-1.79 (-3.62, 0.04)</td>
<td>-0.31 (-0.72, 0.09)</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05.

**p < 0.01.

***p < 0.001.

MD: Mean difference; SMD: Standard mean difference; CI: Confidence interval.
Table 5 Egger’s regression analysis on publication bias

<table>
<thead>
<tr>
<th>Variables</th>
<th>T</th>
<th>P value</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression symptom</td>
<td>-0.75</td>
<td>0.467</td>
<td>-2.60, 1.25</td>
</tr>
<tr>
<td>Anxiety symptom</td>
<td>-1.37</td>
<td>0.196</td>
<td>-7.36, 1.68</td>
</tr>
<tr>
<td>Stress symptom</td>
<td>-4.42</td>
<td>0.021</td>
<td>-9.65, -1.57</td>
</tr>
<tr>
<td>BMI</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Systolic pressure</td>
<td>0.78</td>
<td>0.516</td>
<td>-4.71, 6.80</td>
</tr>
<tr>
<td>Diastolic pressure</td>
<td>-0.24</td>
<td>0.856</td>
<td>-3.48, 3.12</td>
</tr>
<tr>
<td>Total cholesterol</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HDL</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Physical function</td>
<td>-0.23</td>
<td>0.829</td>
<td>-5.76, 4.83</td>
</tr>
<tr>
<td>Mental function</td>
<td>2.03</td>
<td>0.135</td>
<td>-3.83, 17.40</td>
</tr>
<tr>
<td>6-min walk distance</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

BMI: Body mass index; HDL: High-density lipoprotein; CI: Confidence interval.

Figure 1 PRISMA flow diagram. CBT: Cognitive behavioral therapy; RCT: Randomized controlled trial.

We did subgroup analyses to find effective characteristics and components of CBT-based interventions in improving the symptoms of depression and anxiety in CHD patients. Firstly, CBT is more effective when it is provided as an individual rather than group treatment. The possible reason is that the individual treatment method is more targeted and more effective in treating the symptoms of depression and anxiety. Secondly, interventions that do not use a CHD-specific manual have a better effect on symptoms of depression and anxiety due possibly to the fact that the manual approach focuses more on the management of CHD rather than on negative emotions. This is consistent with the conclusions of research conducted by Yang et al [23] on diabetic patients. Thirdly, CBT is more efficient when it emphasizes psycho-
education, behavioral and cognitive strategies as the core approaches, consistent with previous research conclusions, indicating CBT should highlight these strategies as the core\[38,39].

Effects on physiological factors: Our meta-analysis found that CBT-based interventions have a significant effect on the reduction of BMI. The possible reason is associated with the improvement of health-related behaviors. Zetta et al\[42] undertook research to analyze the effect of CBT in patients with angina. The intervention group patients reported increased knowledge awareness, fewer misconceptions, and an increase in self-reported exercise. Murphy et al\[37] found a significantly reduced dietary fat intake among intervention group patients, indicating the favorable impact of the CBT and motivational interviewing program on patient nutritional habits. Thus, CHD patients’ knowledge of the disease and their behavior habits may be improved through CBT-based interventions. They are more likely to maintain a good diet and moderate physical exercise to achieve weight loss goals.

Previous studies have found that CBT-based interventions can help to improve blood pressure and lipid profiles by reducing physiological stress caused by negative emotions, improving autonomic nervous system and endocrine system function\[45]. However, we did not find CBT was effective in improving participants' blood pressure

---

**Figure 2** Forest plots of the effects of cognitive behavioral therapy-based interventions. A: Depression symptom; B: Anxiety symptom; C: Stress symptom. CBT: Cognitive behavioral therapy; CI: Confidence interval.
Li YN et al. Efficacy of cognitive behaviour therapy on CHD

Figure 3 Forest plots of the effects of cognitive behavioral therapy-based interventions. A: Body mass index; B: Systolic blood pressure; C: Diastolic blood pressure; D: Total cholesterol; E: High-density lipoprotein cholesterol. CBT: Cognitive behavioral therapy; CI: Confidence interval.

and lipid profiles. The possible reason is that only four studies reported on blood pressure and two studies on blood lipids. Both blood pressure and blood lipids have complex metabolic mechanisms, reflecting the need for a more significant number of studies to be conducted to assess the effect of CBT-based intervention on these factors accurately.

Effects on QOL and exercise endurance: Increasingly, QOL has become the focus of disease treatment and rehabilitation[46]. In this study, seven studies reported physical function, and five reported mental function. We found that CBT-based interventions can effectively improve patients' physical and mental function, consistent with previous studies[40]. It is possible that through CBT, the patient's negative mood is altered, improving their psychological state. At the same time, through techniques, such as behavioral activation, the exercise time of patients is increased, which is conducive to physical health.

The 6-Minute Walk Test is used to measure functional capacity[47]. As physical function improves, the 6-minute walking distance becomes longer. In contrast with our expectation, we found that the 6-min walk distance of CHD patients was not statistically improved through a CBT intervention. A possible reason for this finding is that only two studies reported this outcome variable. The improvement of a 6-minute walking distance is a long-term, complex process. The size of our sample was insufficient to draw conclusions about any changes in functional capacity.

Strengths and limitations of the study

The advantage of this paper is that it included 22 studies of moderate to high quality. In addition, unlike previous research, this paper also conducted subgroup analyses to...
Li YN et al. Efficacy of cognitive behaviour therapy on CHD

A

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>CBT</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Weight</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bereznakevich et al. (2016)</td>
<td>15.3</td>
<td>17.6</td>
<td>43</td>
<td>5.3</td>
<td>20.3</td>
<td>46</td>
<td>4.9%</td>
<td>6.00</td>
<td>[2.12, 7.88]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedland et al. (2009)</td>
<td>3.96</td>
<td>9.41</td>
<td>41</td>
<td>1.93</td>
<td>9.81</td>
<td>40</td>
<td>17.2%</td>
<td>2.05</td>
<td>[-1.18, 5.22]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Javaheri, S et al. (2020)</td>
<td>-6.4</td>
<td>16.5</td>
<td>15</td>
<td>4.4</td>
<td>15.4</td>
<td>14</td>
<td>22.4%</td>
<td>-10.80</td>
<td>[-22.41, 1.81]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lv, J. et al. (2016)</td>
<td>14.6</td>
<td>9.3</td>
<td>38</td>
<td>6.9</td>
<td>9.8</td>
<td>37</td>
<td>16.1%</td>
<td>7.70</td>
<td>[3.37, 12.03]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’ Neill, A. et al. (2014)</td>
<td>4.43</td>
<td>8.67</td>
<td>61</td>
<td>2.09</td>
<td>8.55</td>
<td>60</td>
<td>32.1%</td>
<td>2.34</td>
<td>[0.73, 4.94]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schneider, L. H et al. (2020)</td>
<td>5.51</td>
<td>7.98</td>
<td>25</td>
<td>2.54</td>
<td>6.96</td>
<td>28</td>
<td>19.0%</td>
<td>2.97</td>
<td>[-0.01, 6.99]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zetke et al (2011)</td>
<td>3.69</td>
<td>21.77</td>
<td>109</td>
<td>0.02</td>
<td>23.22</td>
<td>109</td>
<td>8.5%</td>
<td>3.67</td>
<td>[-2.31, 9.65]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>332</td>
<td>334</td>
<td>100.0%</td>
<td>3.36</td>
<td>[1.63, 5.10]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 3.80 (P=0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4 Forest plots of the effects of cognitive behavioral therapy-based interventions. A: Physical function; B: Mental function; C: 6-min walk distance. CBT: Cognitive behavioral therapy; CI: Confidence interval.

B

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>CBT</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Weight</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bereznakevich et al. (2016)</td>
<td>16.4</td>
<td>20.3</td>
<td>43</td>
<td>1.9</td>
<td>23.1</td>
<td>46</td>
<td>7.9%</td>
<td>14.50</td>
<td>[5.48, 23.52]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedland et al. (2006)</td>
<td>16.11</td>
<td>9.96</td>
<td>41</td>
<td>8.46</td>
<td>12.97</td>
<td>40</td>
<td>15.0%</td>
<td>9.60</td>
<td>[4.68, 14.52]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lv, J. et al. (2016)</td>
<td>32.1</td>
<td>5.7</td>
<td>38</td>
<td>25.1</td>
<td>6.4</td>
<td>37</td>
<td>23.0%</td>
<td>7.02</td>
<td>[4.25, 9.78]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’ Neill, A. et al. (2014)</td>
<td>5.66</td>
<td>9.49</td>
<td>61</td>
<td>2.96</td>
<td>10.85</td>
<td>60</td>
<td>24.4%</td>
<td>3.00</td>
<td>[-0.63, 6.63]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schneider, L. H et al. (2020)</td>
<td>8.29</td>
<td>9.27</td>
<td>25</td>
<td>3.17</td>
<td>8.46</td>
<td>28</td>
<td>20.0%</td>
<td>6.11</td>
<td>[1.60, 10.62]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>288</td>
<td>211</td>
<td>100.0%</td>
<td>6.91</td>
<td>[4.19, 9.73]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 5.03 (P=0.00001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>CBT</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Weight</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
<th>Mean Difference</th>
<th>IV, Fixed</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murphy, B. M et al. (2013)</td>
<td>25.12</td>
<td>90.35</td>
<td>132</td>
<td>17.53</td>
<td>90.7</td>
<td>118</td>
<td>50.9%</td>
<td>7.59</td>
<td>[4.89, 10.27]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turner, A. et al. (2014)</td>
<td>-6.08</td>
<td>68.93</td>
<td>14</td>
<td>15.99</td>
<td>131.28</td>
<td>15</td>
<td>9.1%</td>
<td>30.19</td>
<td>[-41.02, 101.39]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>148</td>
<td>133</td>
<td>100.0%</td>
<td>9.64</td>
<td>[-11.80, 31.08]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect: Z = 0.66 (P=0.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

explore specific characteristics of the CBT intervention, finding some critical components as described above. However, this paper also has some limitations. Due to the small number of articles included in the analysis, we must treat the results with some caution and conclude that more RCTs are needed to enhance the reliability of the evidence. Regarding other limitations, we only analyzed the results before and after the intervention and did not discuss the long-term follow-up results due to insufficient data. Therefore, there was a lack of long-term intervention effects of CBT-based interventions on CHD patients. Secondly, some studies were not able to use a double-blinded study design due to ethical reasons. Consequently, the quality of some of the evidence in this study may be not high. Thirdly, due to the limited number of studies reporting physiological factors, we need to be cautious in concluding the impact of CBT on these factors. Fourth, only articles written in English were included. Hence, it is possible that many eligible studies written in other languages were missed.

Implications

An increasing number of studies have applied CBT-based interventions to manage chronic pain[48], hypertension[49], diabetes[50], and cancer[51] and have found CBT interventions have been effective in improving health outcomes in these patients.
Similarly, our meta-analysis found a positive impact of CBT-based interventions on psychological and physiological factors and QOL in patients with CHD. It is, therefore, reasonable to speculate that the use of CBT as a CHD rehabilitation program is effective. CBT can be incorporated into such programs in hospitals or communities. However, given there has been no standardized procedure for CBT-based interventions for CHD until now, it is advisable to infer that the regular supervision of experienced psychotherapists and psychologists is indispensable to guarantee the quality of CBT delivery.

CONCLUSION

This systematic review and meta-analysis findings suggest that CBT-based intervention is effective for CHD patients in reducing symptoms of depression, anxiety, and stress, BMI levels, and improving physical and mental functioning. In addition, the intervention may be more effective for depression and anxiety in CHD patients when it is delivered as an individual treatment method, when there is no CHD-specific manual, and psycho-education, behavioral, and cognitive strategies are applied as the main approaches.

ARTICLE HIGHLIGHTS

Research background
Coronary heart disease (CHD) is the most common coronary atherosclerotic heart disease, leading to adverse health outcomes for patients with CHD. Cognitive behavioral therapy (CBT) is a short-term psychotherapy, which has been increasingly recognized as an effective intervention approach to reduce psychosocial risk factors in CHD patients. CBT is to change an individual's non-adaptive thought patterns and beliefs, leading to positive behavioral changes. However, no comprehensive systematic review has been conducted to evaluate the efficacy of the CBT-based intervention on health outcomes in patients with CHD.

Research motivation
There is an increasing number of randomized control trials to investigate the efficacy of the CBT-based intervention in CHD patients. However, no comprehensive systematic review has been conducted, and there has been no meta-analysis of the effects of key components of CBT on health outcomes in patients with CHD.

Research objectives
This study aimed to synthesize results from previous randomized controlled trials (RCTs) using a meta-analysis method and explore the therapeutic effect of CBT-based intervention on health outcomes in CHD patients. The secondary aim was to analyze several key components of CBT interventions that explained the effects of CBT intervention.

Research methods
Relevant RCTs published in English were obtained by searching PubMed, Embase, Cochrane Central Register of Controlled Trials, Scopus, and Proquest, with the retrieval time from inception to August 2020. Review Manager 5.3 was used to conduct the meta-analysis. We used The Physiotherapy Evidence Database tool to assess the quality of included studies. Mean difference (MD) was used for continuous outcome variables, and standardized MD was used to measure the effect size. The random-effect model was used to estimate outcomes based on the heterogeneity testing. I² was used for heterogeneity testing, and Egger regression was used to assess the publication bias.

Research results
A total of 22 RCTs were included in the systematic review and meta-analysis. The primary analysis revealed that CBT-based intervention significantly reduced depression, anxiety, stress symptoms, and body mass index (BMI) levels, and improved physical and mental functioning of the quality of life. Additionally, subgroup analysis found that CBT-based interventions were particularly effective in
Reducing symptoms of depression and anxiety in CHD patients when individual treatment form was used, and when psycho-education, behavioral and cognitive strategies were applied as the core treatment components.

Research conclusions
CBT-based interventions are effective in improving health outcomes in CHD patients, including reducing depression, anxiety, stress symptoms, BMI levels, and improving quality of life. Subgroup analysis further found CBT was more effective in reducing depression and anxiety symptoms in CHD patients when it is an individual treatment method, and when psycho-education, behavioral and cognitive strategies are emphasized as the core intervention strategies.

Research perspectives
Our research findings suggest that CBT is a promising approach in the health promotion of the population with CHD. However, the current practice in CBT-based interventions for CHD is not standardized, therefore, further research to standardize the CBT approach in clinical and community intervention programs in patients with CHD is needed.

REFERENCES


