ISSN 2308-3840 (online)

World Journal of *Meta-Analysis*

World J Meta-Anal 2014 August 26; 2(3): 49-134

Published by Baishideng Publishing Group Inc

WJMA World Journal of Meta-Analysis

A peer-reviewed, online, open-access journal of meta-analysis

Editorial Board

2013-2018

The *World Journal of Meta-Analysis* Editorial Board consists of 400 members, representing a team of worldwide experts in clinical meta-analysis research. They are from 41 countries, including Argentina (2), Australia (3), Austria (1), Belgium (5), Brazil (11), Canada (16), Chile (3), China (121), Croatia (1), Egypt (1), Finland (4), France (2), Germany (9), Greece (11), Hungary (2), India (12), Iran (2), Ireland (1), Israel (2), Italy (41), Japan (5), Lithuania (1), Malaysia (1), Netherlands (8), New Zealand (1), Norway (1), Peru (1), Poland (4), Portugal (6), Romania (1), Saudi Arabia (4), Singapore (3), South Africa (1), South Korea (7), Spain (8), Sri Lanka (2), Sudan (1), Switzerland (2), Thailand (4), Turkey (3), United Kingdom (22), and United States (64).

EDITOR-IN-CHIEF

Giuseppe Biondi-Zoccai, *Latina*

GUEST EDITORIAL BOARD MEMBERS

Bo-Ying Bao, *Taichung*
Hsing-Yi Chang, *Maoli*
Ching-Chi Chi, *Chiayi*
Kuo-Liong Chien, *Taipei*
Chien-Chang Lee, *Doliou*
Hung-Chang Lee, *Hsinchu*
Henry WC Leung, *Taoyuan*
YC Su, *Chiayi*
Jauyih Tsauo, *Taipei*

MEMBERS OF THE EDITORIAL BOARD

Argentina

J Mariani, *Ciudad Autónoma de Buenos Aires*
Marcelo Signorini, *Provincia de Santa Fe*

Australia

Mark Boschen, *Southport*
Terry Boyle, *Perth*
Andy Kim Ho Lim, *Melbourne*

Austria

Patrick Sadoghi, *Graz*

Belgium

Marc Arbyn, *Brussels*
Christophe Demoulin, *Liege*
Sascha Colen, *Pellenberg*
P Lehert, *Mons*
Steve Majerus, *Liege*

Brazil

Euclides Castilho, *Sao Paulo*
Luciana Tricai Cavalini, *Rio de Janeiro*
Regina El Dib, *Botucatu*
Alexandre Fachini, *Araraquara*
Guilherme Francisco, *Sao Paulo*
Bruno Gualano, *Sao Paulo*
FC Paes-Barbosa, *Campo Grande*
Rachel Riera, *Sao Paulo*
Inajara Rotta, *Curitiba*
Marcos Sousa, *Belo Horizonte*
Felipe Francisco Tuon, *Curitiba*

Canada

Caroline Barakat-Haddad, *Toronto*
A Baranchuk, *Kingston*
Mohammad Bashashati, *Calgary*
Alonso Carrasco-Labra, *Hamilton*
Eugene Crystal, *Toronto*
Ediriweera Desapriya, *Vancouver*
A Fairchild, *Edmonton*
Alejandro Lazo-Langner, *London*
Michel Lucas, *Québec*
Alex Soroceanu, *Halifax*
Mohamed Ali Tagin, *Winnipeg*
Siamak Bashardoust Tajali, *London*
Sam Michael Wiseman, *Vancouver*
Rebecca KS Wong, *Toronto*
Clement Zai, *Toronto*
Konstantine K Zakzanis, *Toronto*

Chile

Alonso Carrasco-Labra, *Hamilton*
Romina Brignardello Petersen, *Santiago*
Luis A Quiñones, *Santiago*

China

Yi-Xi Bao, *Chongqing*
Janita Chau, *Hong Kong*
Jia-Xu Chen, *Beijing*
Shao-Jie Chen, *Chongqing*
Jin-Fei Chen, *Nanjing*
Hao-Yu Chen, *Shantou*
Ching-Lung Cheung, *Hong Kong*
Wen-Peng Cui, *Changchun*
Cong Dai, *Shenyang*
Bo Deng, *Chongqing*
Qiang Du, *Shenyang*
Jian Fei, *Shanghai*
Chun Gao, *Beijing*
Wei-Hong Ge, *Nanjing*
Ai-Hua Gu, *Nanjing*
Xiao-Xiang Guan, *Nanjing*
Zhi-Yong Guo, *Guangzhou*
Chuan-Yong Guo, *Shanghai*
Zhi-Wei He, *Dongguan*
Ben He, *Shanghai*

Guo-Wei He, *Tianjin*
G Huang, *Shanghai*
Bing-Yang Ji, *Beijing*
Jing Jiang, *Changchun*
Joey Sum Wing Kwong, *Hong Kong*
Wei-Dong Leng, *Shiyan*
Le-Qun Li, *Nanning*
Xiao-Ping Li, *Chengdu*
Jian-Sheng Li, *Zhengzhou*
Jing-Cheng Li, *Chongqing*
Jun-Sheng Li, *Nanjing*
Yan-Yan Li, *Nanjing*
Hua Liu, *Nanchong*
Tong Liu, *Tianjin*
Ai-Ping Lu, *Hong Kong*
Ying Luo, *Kunming*
Chao Ma, *Shanghai*
Jie Ma, *Xi'an*
Xiang-Yu Ma, *Chongqing*
Yan-Lei Ma, *Shanghai*
Wei Nie, *Shanghai*
Wen-Quan Niu, *Shanghai*
Wen-Sheng Pan, *Hangzhou*
Shi-Qiang Shen, *Wuhan*
Xiang-Chun Shen, *Guiyang*
Rui-Hua Shi, *Nanjing*
Yong-Bing Shi, *Suzhou*
Ke-Qing Shi, *Wenzhou*
Marcelo Signorini, *Hong Kong*
Zhi-Yuan Song, *Chongqing*
Qing-Min Sun, *Nanjing*
Yi-Hong Sun, *Beijing*
Shi-Qiao Tan, *Chengdu*
SQ Tan, *Chengdu*
Yong Tang, *Tianjin*
Jiu-Lai Tang, *Hefei*
Na-Ping Tang, *Shanghai*
Jian-Cheng Tu, *Wuhan*
Bin Wang, *Beijing*
Dao-Rong Wang, *Yangzhou*
Jing Wang, *Changshu*
Yu-Ting Wang, *Chengdu*
Xi-Shan Wang, *Harbin*
Fu-Zhou Wang, *Nanjing*
Bing Xia, *Wuhan*
Hong-Xia Wang, *Shanghai*
Zhen-Ning Wang, *Shenyang*
Na Wang, *Shijiazhuang*
Shu-Kui Wang, *Wuhan*
Xing-Huan Wang, *Wuhan*
Wei Wang, *Wuxi*
Cong-Xia Wang, *Xi'an*
Feng Xie, *Shanghai*
Zi-Qiang Xin, *Beijing*
Dan Xing, *Tianjin*
Jun Xiong, *Nanchang*
Xi-Ping Xu, *Guangzhou*
Lin Xu, *Nanjing*
Zhuo-Qun Xu, *Wuxi*
Hui-Ping Xue, *Shanghai*
Tian Yang, *Changchun*
Shuan-Ying Yang, *Xi'an*
Yi-Cong Ye, *Beijing*
Yan-Wei Yin, *Beijing*
Zi Yin, *Guangzhou*
Ym Yin, *Nanjing*
Bin Yu, *Guangzhou*
Yun-Xian Yu, *Hangzhou*
Ling Zhang, *Beijing*
Bei-Bei Zhang, *Chengdu*
Li-Li Zhang, *Chongqing*
Qiu Zhang, *Hefei*
Shuo Zhang, *Shenyang*
You-Cheng Zhang, *Lanzhou*
Jian Zhang, *Shanghai*
Jun-Hua Zhang, *Tianjin*
Yu-Rong Zhang, *Xi'an*
Zhong-Heng Zhang, *Jinhua*
Hai-Tao Zhao, *Beijing*
Pan Zhao, *Beijing*
Yu-Lan Zhao, *Shanghai*
Jie-Jiao Zheng, *Shanghai*
Xue-Sheng Zheng, *Shanghai*
Guo-Qing Zheng, *Wenzhou*
Ming-Hua Zheng, *Wenzhou*
Cui-Hong Zheng, *Wuhan*
Lai-Ping Zhong, *Shanghai*
Tian-Biao Zhou, *Nanning*
Peng Zhou, *Shanghai*
Ping Zhou, *Wuhan*
Kun-Ju Zhu, *Zhanjiang*
Xiao-Ping Zou, *Nanjing*

Croatia

Miljenko Franic, *Zagreb*

Egypt

Ashraf Fawzy Nabhan, *Cairo*

Finland

Jouni JK Jaakkola, *Oulu*
Ville Kyto, *Turku*
Jouko Miettunen, *Oulu*
R Quansah, *Oulu*

France

Alain Braillon, *Amiens*
Francesco Fiorica, *Ferrara*

Germany

Tonio Ball, *Freiburg*
Robert Bergholz, *Hamburg*
Jan Brunkwall, *Cologne*
Holger Cramer, *Essen*
Joseph Kambeitz, *Munich*
Sascha Meyer, *Homburg*
Thomas Nickl-Jockschat, *Aachen*
Martin Pinquart, *Marburg*
Robert Schier, *Cologne*

Greece

Vangelis G Alexiou, *Athens*
Stefanos Bonovas, *Athens*
Athanasios Papatsoris, *Athens*
Irini Chatziralli, *Athens*
Dimitrios Daoussis, *Patras*
Pagona Lagiou, *Athens*
John Goudakos, *Thessaloniki*
Savas Grigoriadis, *Thessaloniki*
Konstantinos A Toulis, *Thessaloniki*
Sotirios Tsiodras, *Athens*
Nikolaos Tsoukalas, *Athens*

Hungary

Balazs Gyorffy, *Budapest*
Istvan Wittmann, *Pecs*

India

Ritesh Agarwal, *Chandigarh*
Giridhara R Babu, *Bangalore*
Subho Chakrabarti, *Chandigarh*
Y Madhavi, *New Delhi*
Tanu Midha, *Kanpur*
Kameshwar Prasad, *New Delhi*
Kaushal Kishor Prasad, *Chandigarh*
Krishna Undela, *Mysore*
Singh Rajender, *Lucknow*
Vinod Ravindran, *Kozhikode*
V Shetty, *Mumbai*
R Umaya Suganthi, *Bangalore*

Iran

Nejat Mahdieh, *Ilam*
Sadeghi Ramin, *Mashhad*

Ireland

Ian Conrick-Martin, *Dublin*

Israel

Uri Kopylov, *Ramat Gan*
Meir Lotan, *Kfar-Saba*

Italy

Umberto Aguglia, *Catanzaro*
Alessandro Antonelli, *Pisa*
Annalisa Blasetti, *Chieti*
Francesco Brigo, *Merano*
Emanuele Cereda, *Pavia*
Emanuele Cigna, *Rome*
Roberto Cirocchi, *Terni*
Bernardo Cortese, *Milano*

Alessandro Cucchetti, *Bologna*
Gianfranco Damiani, *Rome*
Fabrizio D'Ascenzo, *Turin*
Massimo Del Fabbro, *Milano*
Valeria Fadda, *Florence*
Alessandro Fancellu, *Sassari*
Giuseppe Ferrante, *Milan*
Virginia Festa, *Rome*
Francesco Fiorica, *Ferrara*
Guglielmo Giraldi, *Rome*
Jenny Guidi, *Bologna*
Lorenzo Loffredo, *Rome*
Andrea Messori, *Firenze*
Eliano Pio Navarese, *Bydgoszcz*
Stefano Omboni, *Solbiate Arno*
Alvisa Palese, *Udine*
Stefano Palomba, *Reggio Emilia Stefano*
Carlo Perricone, *Rome*
Mario Petretta, *Naples*
Alessandro Pezzini, *Brescia*
Gianluca Pontone, *Milan*
Paolo Emilio Puddu, *Rome*
Andrea Rognoni, *Novara*
Giuseppe Scalabrino, *Milan*
Città della Salute e della Scienza, *Turin*
Fabrizio Sgolastra, *L'Aquila*
Maria Lucia Specchia, *Rome*
Stefano Trastulli, *Terni*
Fabio Tine, *Palermo*
Nereo Vettoretto, *Chiari*
Alberto Vianello, *Perugia*
Luigi Zorcolo, *Cagliari*

Japan

Nguyen Tien Huy, *Nagasaki*
Hiroharu Kamioka, *Tokyo*
Koji Kawakami, *Kyoto*
Keitaro Matsuo, *Nagoya*
Kazushi Okamoto, *Nagoya*

Lithuania

Edmundas Kadusevicius, *Kaunas*

Malaysia

SP Pani, *Ipoh*

Netherlands

Michel van den Bekerom, *Amsterdam*
Dimitra Dodou, *Delft*
D Haverkamp, *Amsterdam*
Vassilios Koussoulas, *Drachten*
BJ Polder, *Nijmegen*
Theo Stijnen, *Leiden*
RNM Weijers, *Amsterdam*
Joost de Winter, *Delft*

New Zealand

Shaofeng Li, *Auckland*

Norway

Eivind Berge, *Oslo*

Peru

Rafael Bolaños Díaz, *Lima*

Poland

Maciej Banach, *Lodz*
Krzysztof Jonderko, *Sosnowiec*
Jolanta Lissowska, *Warsaw*
Maciej Plaszewski, *Warsaw*

Portugal

Daniel Caldeira, *Lisboa*
J Costa, *Lisbon*
Ana Miguel, *Coimbra*
Manuel Morgado, *Covilh*
Bárbara Peleteiro, *Porto*
Rui Torres, *Paredes*

Romania

Fratila Ovidiu, *Oradea*

Saudi Arabia

Hazem M Al-Mandeel, *Riyadh*
Ezzeldin M Ibrahim, *Jeddah*
Mutahir A Tunio, *Riyadh*
Hayfaa A Wahabi, *Riyadh*

Singapore

Nikos LD Chatzisarantis, *Singapore*
Edwin Choon Wyn Lim, *Singapore*
Roger Ho, *Singapore*

South Africa

Alaine Umubyeyi Nyaruhirira, *Pretoria*

South Korea

Jung-Hee Kim, *Cheonan*
Hyangsook Lee, *Seoul*
Myeong Soo Lee, *Daejeon*
Chi-Un Pae, *Bucheon*
Yong Hyun Park, *Yangsan*
Jae Hong Seo, *Seoul*
Yong Sang Song, *Seoul*

Spain

Pablo Avanzas, *Oviedo*
Joan Cid, *Barcelona*
Joaquin de Haro, *Madrid*
Joan Guardia-Olmos, *Barcelona*
Nabil Halaihlel, *Zaragoza*
JA Monge-Argilés, *Alicante*
Raul Moreno, *Madrid*
Inés Velasco, *Aracena*

Sri Lanka

Ranil Jayawardena, *Colombo*
Priyanga Ranasinghe, *Colombo*

Switzerland

Jay P Singh, *Zurich*
Giorgio Treglia, *Bellinzona*

Sudan

Samir MH Shaheen, *Khartoum*

Thailand

Chuenjid Kongkaew, *Phitsanulok*
Manop Pithukpakorn, *Bangkok*
Piyamitr Sritara, *Bangkok*
Surasak Saokaew, *Phayao*

Turkey

Nese Demirturk, *Afyonkarahisar*
Nilüfer Ozabaci, *Eskisehir*
Ilke Sipahi, *Istanbul*

United Kingdom

Omar M Aboumarzouk, *Wales*
Abeer Al-Namankany, *London*
Ernest A Azzopardi, *Cardif*
Umberto Benedetto, *Cambridge*
Joanne Brooke, *London*
Noriko Cable, *London*
David Chan, *Cardiff*
YC Cheong, *Southampton*
Andrew Currie, *Harrow*
G Nabi, *Dundee*

Lesley A Anderson, *Belfast*
Valentina Gallo, *London*
Gianpiero Gravante, *Leicester*
Peter N Lee, *Surrey*
Igho Onakpoya, *Oxford*
Ashish Pradhan, *Huntingdon*
Evridiki Patelarou, *London*
Jian-Qing Shi, *Newcastle*
Surendra P Singh, *Wolverhampton*
Natalie Taylor, *Leeds*
Yousef Shahin, *Hull*
Zheng Ye, *Cambridge*

United States

L Joseph Su, *Rockville*
Olusola Adesope, *Pullman*
Mike Allen, *Milwaukee*
Bhupinder Anand, *Bellaire*
Stephen Aronoff, *Philadelphia*
KoKo Aung, *San Antonio*
William L Baker, *Farmington*
Moritz C Wyler von Ballmoos, *Milwaukee*
Matthew L Bechtold, *Columbia*
Atul Bhardwaj, *Hershey*
Somjot S Brar, *Los Angeles*
Hui Cai, *Nashville*
Subhash Chandra, *Towson*
Wen-Pin Chang, *Omaha*
Yong Chen, *North Wales*
Myunghan Choi, *Phoenix*
John Coverdale, *Houston*
Prakash C Deedwania, *Fresno*
Eugene Demidenko, *Hanover*
Hong-Wen Deng, *New Orleans*
Eric M Deshaies, *Syracuse*
Tao Fan, *Whitehouse Station*
Shinga Feresu, *Bloomington*
Janvier Gasana, *Miami*
Kaveh Hajifathalian, *Boston*
Mohammad Obaidul Hoque, *Baltimore*
Larissa R Brunner Huber, *Charlotte*
Imran H Iftikhar, *Columbia*
Vijayvel Jayaprakash, *Buffalo*
Xuezhi Jiang, *Weat Reading*
Shuo Jiao, *Seattle*
Evelyn Johnson, *Boise*
Le Kang, *Silver Spring*
SR Kapadia, *Cleveland*
Lior Katz, *Houston*
Daniel M Laskin, *Richmond*
Yu Liang, *Foster City*
Paul Ellis Marik, *Norfolk*
Lynne V McFarland, *Seattle*
Marcovalerio Melis, *New York*
Brian J Miller, *Augusta*
Pavlos Msaouel, *New York*
Joshua E Muscat, *Hershey*
Chee Yuan Ng, *Loma Linda*
Nghi C Nguyen, *Saint Louis*
Brandi S Niemeier, *Whitewater*
Thomas D Parsons, *Denton*
Nidal Abi Rafeh, *New Orleans*
Praveen Roy, *Marshfield*
Ali Salavati, *Philadelphia*
Ankur Sethi, *Chicago*
Tatyana A Shamliyan, *Minneapoli*
Qian Shi, *Rochester*
Zhongjie Shi, *Philadelphia*
Param Puneet Singh, *Chicago*
KV Slavin, *Chicago*
Ali El Solh, *Buffalo*
Jielin Sun, *Winston-Salem*
Richard G Trohman, *Chicago*
Laurah Turner, *Cincinnati*
Sheila Wilhelm, *Detroit*
Alex K Wong, *Los Angeles*
Xiaohui Xu, *Gainesville*
Lu Yin, *Nashville*

WJMA World Journal of Meta-Analysis

Contents

Quarterly Volume 2 Number 3 August 26, 2014

Contents

World Journal of Meta-Analysis
Volume 2 Number 3 August 26, 2014

AIM AND SCOPE

World Journal of Meta-Analysis (*World J Meta-Anal*, *WJMA*, online ISSN 2308-3840, DOI: 10.13105) is a peer-reviewed open access academic journal that aims to guide clinical practice and improve diagnostic and therapeutic skills of clinicians, with a specific focus on meta-analysis, systematic review, mixed-treatment comparison, meta-regression, overview of reviews.

WJMA covers a variety of clinical medical fields including allergy, anesthesiology, cardiac medicine, clinical genetics, clinical neurology, critical care, dentistry, dermatology, emergency medicine, endocrinology, family medicine, gastroenterology and hepatology, geriatrics and gerontology, hematology, immunology, infectious diseases, internal medicine, obstetrics and gynecology, oncology, ophthalmology, orthopedics, otolaryngology, pathology, pediatrics, peripheral vascular disease, psychiatry, radiology, rehabilitation, respiratory medicine, rheumatology, surgery, toxicology, transplantation, and urology and nephrology, while maintaining its unique dedication to systematic reviews and meta-analyses.

EDITORS FOR THIS ISSUE

Responsible Assistant Editor: *Xiang Li*
Responsible Electronic Editor: *Ya-Jing Lu*
Proofing Editor-in-Chief: *Lian-Sheng Ma*
Responsible Science Editor: *Yuan Qi*
Proofing Editorial Office Director: *Jin-Lei Wang*

NAME OF JOURNAL
World Journal of Meta-Analysis

ISSN
ISSN 2308-3840 (online)

LAUNCH DATE
May 26, 2013

FREQUENCY
Quarterly

EDITORS-IN-CHIEF
Giuseppe Biondi-Zoccai, MD, Assistant Professor, Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina 04100, Italy

EDITORIAL OFFICE
Jin-Lei Wang, Director
Xiu-Xia Song, Vice Director
World Journal of Meta-Analysis
Room 903, Building D, Ocean International Center, No. 62 Dongsihuan Zhonglu, Chaoyang District, Beijing 100025, China
Telephone: +86-10-85381891
Fax: +86-10-85381893
E-mail: editorialoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com

PUBLISHER
Baishideng Publishing Group Inc
8226 Regency Drive,
Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com

PUBLICATION DATE
August 26, 2014

COPYRIGHT
© 2014 Baishideng Publishing Group Inc. Articles published by this Open-Access journal are distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license.

SPECIAL STATEMENT
All articles published in journals owned by the Baishideng Publishing Group (BPG) represent the views and opinions of their authors, and not the views, opinions or policies of the BPG, except where otherwise explicitly indicated.

INSTRUCTIONS TO AUTHORS
Full instructions are available online at http://www.wjgnet.com/2308-3840/g_info_20100722180909.htm

ONLINE SUBMISSION
http://www.wjgnet.com/esps/

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.49

World J Meta-Anal 2014 August 26; 2(3): 49-63
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

REVIEW

Stroke and depression: A bidirectional link

Elisabetta Del Zotto, Paolo Costa, Andrea Morotti, Loris Poli, Valeria de Giuli, Alessia Giossi, Irene Volonghi, Antonio Callea, Alessandro Padovani, Alessandro Pezzini

Elisabetta Del Zotto, U.O. di Recupero e Rieducazione Funzionale, IRCCS Fondazione Don Gnocchi, 20148 Milano, Italy
Paolo Costa, Andrea Morotti, Loris Poli, Valeria de Giuli, Irene Volonghi, Alessandro Padovani, Alessandro Pezzini, Dipartimento di Scienze Cliniche e Sperimentali, Clinica Neurologica, Università degli Studi di Brescia, 25123 Brescia, Italy
Alessia Giossi, U.O Neurologia, Istituto Clinico "S. Anna", 25127 Brescia, Italy
Antonio Callea, Dipartimento Fisiopatologia Medico-Chirurgica e dei Trapianti, Sezione di Neuroscienze, Università degli Studi di Milano, 20122 Milano, Italy
Author contributions: Del Zotto E, Costa P, Morotti A, Poli L, de Giuli V, Giossi A, Volonghi I, Callea A, Padovani A and Pezzini A contributed to this paper; with substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and with final approval of the version to be published.
Correspondence to: Alessandro Pezzini, MD, Dipartimento di Scienze Cliniche e Sperimentali, Clinica Neurologica, Università degli Studi di Brescia, P.le Spedali Civili, 1, 25123 Brescia, Italy. ale_pezzini@hotmail.com
Telephone: +39-30-3384086 Fax: +39-30-3384086
Received: March 10, 2014 Revised: May 23, 2014
Accepted: June 10, 2014
Published online: August 26, 2014

Abstract

A number of studies have assessed the influence of depression on the risk of cardiovascular disease. A growing literature indicates a link between depression and cerebrovascular events, although the direction of this association remains unclear. Numerous data have emerged suggesting an association between depressive symptoms and subsequent risk of stroke, thus leading to the hypothesis that a direct causality between depression and stroke exists. Notwithstanding, how depression may act as a risk factor for stroke is still unclear. Depression might be linked to stroke *via* neuroendocrine and inflammation effects, through correlation with major comorbidities such as hypertension and diabetes or by intervention of lifestyle behavioral mediators. Finally, antidepressant medications have recently drawn attention for a possible association with increased risk of stroke, although such findings remain uncertain. Depression has been also established as an important consequence after stroke, exerting a significant adverse impact on the course of motor recovery, social functioning and, overall, on quality of life. Post stroke depression occurs in nearly one third of stroke cases, but the exact mechanism leading to depression after stroke is still incompletely understood. In this article, we will review contemporary epidemiologic studies, discuss potential mechanisms and specific aspects of the complex relation between depression and stroke.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Depression; Mood disorders; Stroke; Post-stroke depression; Antidepressant medications; Cerebrovascular disease

Core tip: A number of studies have assessed the influence of depression on the risk of cardiovascular disease. A growing literature indicates a link between depression and cerebrovascular events, although the direction of this association remains unclear. Numerous data have emerged suggesting an association between depressive symptoms and subsequent risk of stroke, thus leading to the hypothesis that a direct causality between depression and stroke exists. Depression has been also established as an important consequence after stroke, affecting functional recovery and quality of life. Contemporary epidemiologic studies, potential mechanisms and specific aspects of the complex relation between depression and stroke will be discussed.

Del Zotto E, Costa P, Morotti A, Poli L, de Giuli V, Giossi A, Volonghi I, Callea A, Padovani A, Pezzini A. Stroke and depression: A bidirectional link. *World J Meta-Anal* 2014; 2(3): 49-63 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/49.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.49

INTRODUCTION

According to definition of the Diagnostic and Statistical Manual of Mental Disorders-Ⅳ (DSM-Ⅳ), major depressive disorder consists of two or more episodes of depressed mood, loss of interest or diminished sense of pleasure in usual activities for more than 2 wk, in addition to other depressive features sufficient to cause clinically important psychological or physical distress, or functional impairment, that is atypical for usual behavior and not attributable to a medical condition or bereavement[1].

Depression is a highly prevalent condition worldwide, more common among women, with a lifetime prevalence of more than 16% in the general population, and a consequent impact on public health[2]. In the last decades, a number of studies have investigated the influence of depression on the risk of developing chronic condition such as diabetes[3] and hypertension[4], and also cardiovascular disease[5]. Adding to these findings, a growing literature indicates a link between depression and cerebrovascular events, although the exact mechanisms of this association remain unclear.

The presence of depression has been established as an important consequence after stroke[6], affecting functional recovery and quality of life[7]. In addition to this evidence, numerous data have emerged pointing toward a relation between depressive symptoms and subsequent risk of stroke as well as common predisposing conditions such as hypertension, diabetes or lifestyle behavioral factor[8], thus leading to the hypothesis that a direct causality between depression and stroke exists[9].

IS DEPRESSION A RISK FACTOR FOR STROKE?

Several observational studies investigated the relation between depression and the risk of subsequent stroke, with conflicting results (Table 1).

Data from 10 studies published before 2005 were pooled in a first meta-analysis which detected an association between depression and risk of stroke, but with a significant heterogeneity among the studies included[38]. Subsequently, many other studies were published and recently summarized in two more detailed meta-analyses[39,40], strengthening the evidence of a possible role of depression as a modifiable risk factor for stroke. According to the meta-analysis of Pan and coworkers, including prospective cohort studies, the pooled HR of stroke among patients with depression is 1.45 (95%CI: 1.29-1.63). Stratifying analysis by pathological stroke subtype, the pooled HR for ischemic stroke is 1.25 (95%CI: 1.11-1.40), while there is no significant influence of depression on the pooled risk of hemorrhagic stroke[39]. However, since very few studies have analyzed the association by stroke subtype we cannot draw any conclusion in this regard.

Likewise, a predisposing effect of depression was also observed by Dong *et al*[40] in their meta-analysis (pooled RR = 1.34; 95%CI: 1.17-1.54 in depressed subjects compared to non-depressed). Even, the INTERSTROKE study, an international multicenter case-control study designed to establish the association of traditional and emerging risk factors with stroke in countries of high, middle, and low income, reported a similar magnitude of the association between depression and an increased risk of all stroke and ischemic stroke (OR = 1.35; 99%CI: 1.10-1.66, and 1.47; 99%CI: 1.19-1.83, respectively), but not intracerebral haemorrhagic stroke[41].

Both meta-analyzes found no difference in pooled risk stratifying data by gender[39,40]. Although few studies reported age-stratified results, the age difference in the depression-stroke association was also evaluated, resulting in a increased risk in younger subjects [mean age < 65 years, HR = 1.77 (95%CI: 1.30-2.41); mean age ⩾ 65 years, HR = 1.30 (95%CI: 1.18-1.44)][39]. In this regard, the Framingham study was the first to find evidence of effect modification by age, examining the elderly and non-elderly groups separately and documenting the association between depressive symptoms and stroke risk in those aged ⩽ 65 years[26]. Similarly, the Established Populations for Epidemiologic Studies of the Elderly (EPESE) which examined an older population with subgroups aged 65-74 years, and 75 years or older, observed this relation with age, with depression associated with increased stroke risk in younger but not in older participants[21]. Recently, in line with these data, the Intervention Project on Cerebrovascular Diseases and Dementia in the District of Ebersberg (INVADE trial), a population-based longitudinal study, corroborated the association between depression and the risk of ischemic stroke, particularly in women and patients younger than 65 years[35]. These findings suggest the hypothesis that differences may exist in the depression-associated stroke risk in sub-groups of subjects defined by age and sex, which needs to be investigated and confirmed in further studies.

Is this enough to establish a causal association between depression and subsequent risk of stroke? As pointed out by many authors, most of the studies reported so far present several methodological limitations, and there is evidence of a significant heterogeneity.

First, this depends on differences in study design, sample size and population characteristics. Exclusion of patients with history of stroke is important to avoid the possibility of reverse causality and bias in risk estimation. Both meta-analyzes observed a temporal relationship between depression and stroke by including only first-time stroke events that occurred after baseline assessments of depression[39,40]. However the possibility that undiagnosed stroke may have caused depression remains, despite efforts to exclude enrollment of participants with preexisting stroke at study entry. Furthermore, the reverse causality hypothesis might apply to the association with stroke as well, because depressive symptoms might be markers of preexisting cerebrovascular disease, as suggested by the vascular depression hypothesis[42]. In this regard, the results of two recent analyses from the Rotterdam

Table 1 Risk of stroke according to depression status in prospective studies

Ref.	Population	Follow-up years, (period)	No. of stroke cases	Risk of total stroke (95%CI)	Risk of ischemic stroke (95%CI)	Risk of hemorrhagic stroke (95%CI)	Adjustment for covariates and confounders	Depression assessment	Stroke ascertainment
Vogt *et al*[10]	2573 Men and women aged ⩾ 18 yr	15 (1970/71-1985)	NA	HR 0.84 (0.57-1.22)			Age, sex, socioeconomic status, length of health plan membership, subjective health status, smoking	Depressive index	Death index or vital records
Wassertheil-Smoller *et al*[11]	4367 Men and women aged ⩾ 60 yr	4.5 (1985-1990)	204	RR 0.85 (0.45-1.64)			Age, baseline depression, race, education, smoking, diabetes, history of CVD, activities of daily living	CES-D	Medical records
Everson *et al*[12]	6676 Men and women aged 16-94 yr	29 (1965-1983)	169	RR 1.54 (1.06-2.22)			Age, sex, race, education, hypertension, diabetes, smoking status, alcohol, BMI	HPL Depression Scale	Death certificates
Whooley *et al*[13]	7518 Women aged ⩾ 65 yr	6 (1988-1994)	94	HR 1.7 (0.8-3.5)			Age, history of MI, stroke, hypertension, diabetes, smoking, perceived health, cognitive function	GDS	Medical records
Jonas *et al*[14]	6095 Men and women aged 25-74 yr	16 (1971/75-1992)	483	RR 1.73 (1.30-2.31)			Age, SBP, education, smoking, BMI, alcohol, physical activity, cholesterol, diabetes, history of heart disease	GWS	Hospital records and death certificates
Larson *et al*[15]	1703 Men and women aged ⩾ 18 yr	13 (1980/1983-1993/1996)	95	OR 2.67 (1.08-6.63)			Age, sex, race, educational attainment, smoking status, history of diabetes, heart problems, high blood pressure	DIS	Self-report and death certificates
Ohira *et al*[16]	879 Men and women aged 40-78 yr	10.3 (1985-1996)	69	RR 1.9 (1.1–3.5)	RR 2.7 (1.2–6.0)	RR 0.9 (0.3–3.1)	Age, sex, smoking status, alcohol use, BMI, SBP, serum total cholesterol level, current treatment with antihypertensive medication, and history of diabetes	Zung SDS	National insurance claims, medical records, clinical diagnosis and death certificates
Ostir *et al*[17]	2478 Men and women aged ⩾ 65 yr	6 (1986-1992)	340	RR 1.30 (0.85-1.99)			Age, sex, marital status, household income, education, smoking status, BMI, heart attack, diabetes, hypertension	CES-D	Self-report and death certificates
May *et al*[18]	2124 Men aged 45-59 yr	14 (1984/1988-1998)	130		HR 1.26 (0.85–1.85)		Age, social class, marital status, smoking status, alcohol use, BMI, SBP, comorbidity (ischemic heart disease, diabetes, respiratory disease, or retirement due to ill health)	GHQ	Medical records
Wassertheil-Smoller *et al*[19]	93676 Women aged 50-79 yr	4.1 (1993/1998-2000)	751	RR 1.01 (0.78-1.3)[1]; 1.45 (1.11-1.9)[2]			Age, race, education, income, smoking status, BMI, physical activity, hormone use, high cholesterol level requiring medications, diabetes, and hypertension	CES-D and DIS	Self-report and medical redords
Gump *et al*[20]	11216 Men aged 35-57 yr	18.4 (1979/1981-1999)	167	HR 1.48 (0.93-2.36)			Age, intervention group, race, educational attainment, smoking at baseline and visit 6, SBP, alcohol consumption, fasting cholesterol and the occurrence of nonfatal cardiovascular events during the trial	CES-D	Death certificates
Avendano *et al*[21]	2812 Men and women aged ⩾ 65 yr	12 (1982-1994)	270	HR age 65-74: 3.05 (1.63–5.7); age ⩾ 75: 0.95 (0.46–1.98)			Age, sex, race, education and income	CES-D	Self report and medical records

Stürmer *et al*[22]	3920 Men and women aged 40-65 yr	8.5 (1992/1995-2002/2003)	62	RR 1.53 (0.83-2.8)		Age, sex, education, smoking status, alcohol consumption, BMI, exercise, family histories of MI, stroke and cancer, comorbidity	Standardized personality questionnaires	Medical records and death certificates
Kamphuis *et al*[23]	799 Men aged 70-90 yr	7.4 (1989/1991-2000)	66	HR 3.41 (1.69-6.9)		Age, country, education, BMI, smoking, alcohol intake, blood pressure, cholesterol, and physical activity	Zung SDS	Death certificates
Arbelaez *et al*[24]	5525 Men and women aged ⩾ 65 yr	11 (1989-2000)	611	HR 1.25 (1.02–1.53)		Age, sex, race, occupation, income, education level, marital status, smoking status, BMI, blood cholesterol, highdensity lipoprotein cholesterol, lowdensity lipoprotein cholesterol, triglycerides levels, comorbidity (hypertension, diabetes, CHD), CRP	CES-D	Self report and medical records
Kawamura *et al*[25]	535 Men and women aged ⩾ 65 yr	6.3 (1985-2000)	103	RR 1.25 (0.82–1.90)		NA	Zung SDS and physician diagnosis	Death certificates
Salaycik *et al*[26]	4120 Men and women aged ⩾ 29 yr	8 (1990/1998-1998/2006)	228	HR Age < 65: 3.59 (1.76–7.3); Age ⩾ 65: 0.93 (0.59–1.47)		Age, sex, smoking status, blood pressure, diabetes, atrial fibrillation, CVD, left ventricular hypertrophy on the ECG	CES-D or ADM use	Medical records (Stroke+TIA)
Bos *et al*[27]	4424 Men and women aged ⩾ 61 yr	5.8 (1997/1999-2005)	291	HR 1.21 (0.80–1.83)	HR 1.43 (0.87-2.35)	Age, sex, smoking status, SBP, intima–media thickness, medication use (antithrombotic drug, antihypertensive drug, cholesterol lowering drug, psycholeptic drug and psychoanaleptics drug), comorbidity (diabetes, MI, PTCA or CABG, TIA)	CES-D	Medical records
Lee *et al*[28]	4962 Men and women 18-44 yr	5 (1998-2003)	98	HR 5.43 (3.47-8.5)		Age, sex, geographic location, income, urbanization, substance abuse, comorbidity (diabetes, hypertension, hyperlipidemia, renal disease)	Physician diagnosis	Medical records
Liebetrau *et al*[29]	401 Men and women aged 85 yr	3 (1986/1987-1989/1990)	56	HR 2.6 (1.5-4.6)		NA	DSMMD-III	Self report and medical records
Surtees *et al*[30]	20627 Men and women aged 41-80 yr	8.5 (1996/2000-2006)	595	HR 1.08 (0.67–1.75)		Age, sex, social class, education, smoking status, obesity, SBP, total cholesterol, diabetes, hypertension treatment, MI, family history of stroke, and ADM use	HLEQ	Medical records
Whooley *et al*[8]	20627 Men and women aged mean 63 yr and 68 yr[3]	4.8 (2000/2002-2008)	47	HR 1.47 (0.70-3.11)		Age, smoking, medication adherence, physical activity, CRP, left ventricular ejection fraction, history of MI, stroke, heart failure, diabetes	PHQ	Self report and medical records
Glymour *et al*[31]	19087 Men and women aged ⩾ 50 yr	8.1 (1996-2006)	1864	HR 1.25 (1.12- 1.39)		Age, overweight, ethnicity, birth place, parental educations, self education, income, wealth and marital status, year of enrollment, alcohol and smoking status, and self-reported diagnoses of hypertension, diabetes, or heart disease	CES-D	Self report
Nabi *et al*[32]	23282 Men and women aged 20-54 yr	7 (1998-2005)	129	HR 0.87 (0.57–1.32)		Age, sex, education, smoking status, alcohol use, sedentary lifestyle, obesity,comorbidity (hypertension, diabetes, incident CHD)	BDI	Medical records

Peters *et al*[33]	2656 Men and women aged ⩾ 80 yr	2.1 (2001-2007)	97	HR 1.82 (1.19–2.78)			Age, sex, treatment group, treatment allocation, country area, educational level, living status, number of comorbidities, previous CVD, previous treatment for hypertension and hypertension	GDS	Self report and medical records
Pan *et al*[34]	80574 Women aged 54-79 yr	6 (2000-2006)	1033	RR 1.29 (1.13-1.48)	RR 1.11 (0.91-1.35)	RR 1.20 (0.80-1.79)	Age, marital status, ethnicity, smoking status, alcohol consumption, BMI, physical activity level, menopausal status, postmenopausal hormone therapy, current aspirin or multivitamin use, Dietary Approaches to Stop Hypertension dietary score, parental history of MI	MHI or self reported diagnosis or ADM use	Self report, medical records and death certificates
Seifert *et al*[35]	3852 Men and women aged > 55 yr	6.1 (2001/2003-2009)	156	HR 1.3 (0.9-1.89)			Age, sex, BMI, smoking, hypertension, diabetes, hyperlipidemia, previous MI, previous TIA, previous stroke, history of atrial fibrillation, physical activity	GDS or ADM use	Medical reports
Majed *et al*[36]	9601 Men aged 48-64 yr	10 (1991/1993-2003)	136	HR 1.41 (0.95-2.11)	HR 1.65 (1.07-2.55)		Age; study centers; socioeconomic factors, including marital status, education level, employment status, physical activity, smoking status, daily alcohol intake, SBP, use of anti-hypertensive drugs, BMI, total and high-density lipoprotein cholesterol, treatment for diabetes, ADM use.	CES-D	Hospital records and general practitioner records
Jackson *et al*[37]	10547 Women aged 47-52 yr	12 (1998-2010)	177	OR 1.94 (1.37- 2.74)			Age, education, homeownership, hypertension, diabetes mellitus, heart disease, hysterectomy/oophorectom, smoking, alcohol use, physical activity, BMI	CES-D or ADM use	Self report and death certificates

[1]No baseline CVD; [2]With baseline CVD; [3]Group with depression mean age 63 years, group without depression mean age 68 years. n Scale: Human Population Laboratory Depression Scale; GDS: Geriatric. CES-D: Center for Epidemiologic Studies Depression Scale; HPL Depression Depression Scale; GWS: General Well-Being Schedule; DIS: diagnostic interview schedule; Zung SDS: Zung Self-Rating Depression Scale; GHQ: General Health Questionnaire; ADM: Antidepressant medication; DSMMD: Diagnostic and Statistical Manual of Mental Disorders; HLEQ: Health and Life Experiences Questionnaire; PHQ: Patient Health Questionnaire; BDI: Beck Depression Inventory; MHI: Mental Health Index; CI: Confidence interval; NA: Not available; HR: Hazard ratio; RR: Relative risk; OR: Odds ratio; CVD: Cardiovascular disease; BMI: Body mass index; MI: Myocardial infarction; CHD: Coronary heart disease; CRP: C reactive protein; SBP: Systolic blood pressure; ECG: Electrocardiogram; PTCA: Percutaneous transluminal angioplasty; CABG: Coronary artery bypass graft.

Study[43] and the Health and Retirement Study[31] did not support this hypothesis.

Second, in most studies the influence of several confounders on the final results was not considered. These include, for example, behavioral and lifestyle factors such as alcohol consumption, dietary factors and physical activity, as well as socioeconomic status, which raise the possibility that residual confounders may be partially responsible for the relation between depression and stroke. In this regard, data from a prospective study including a cohort of 1017 subjects with stable coronary disease, aimed at evaluating the influence of baseline disease severity and potential biological or behavioral mediators on the association of depressive symptoms with subsequent cardiovascular events, point towards a role of poor health behaviors, particularly physical inactivity[8].

Third, in most of the studies reported so far there are differences in stroke and depression measures. Stroke case ascertainment is based on a variety of sources, including medical records, death certificates, clinical diagnoses and self-reports. The assessment of depression is more often based on self-rating questionnaires than on a psychiatric structured interview and the DSM criteria.

Although the link between depression and stroke is evident using diagnostic measures of depression[15,28], it is more consistent when using measures of depressive symptoms[14,16-17,26-27,30]. Most studies used the Center for Epidemiological Studies Depression scale (CES-D), a self-rating screening test for depression with an adequate internal consistency and reliability. The pooled RR for studies that used CES-D for depression assessment has been estimated to be 1.23 (95%CI: 1.13-1.34)[40]. However, we should also consider other types of mood disorder, such as bipolar disorder, which are not measured. In this regard, bipolar disorder has been suggested to be a potential risk factor for vascular disease[44] and stroke (HR = 1.24, 95%CI: 1.12-1.38) after adjusting for patients' demographic characteristics, comorbid medical disorders and socioeconomic status[45].

These drawbacks support the need for further studies aimed at investigating the link between affective disorders and stroke. Moreover, both stroke and depression have a high prevalence and incidence. It was estimated that approximately 3.9% (n = 273000) of stroke cases in the United States could be related to depression[39]. This raises important clinical and public health implications because of the potential of reducing stroke risk by prevention and treatment of mood disorders.

HOW CAN DEPRESSION LEAD TO STROKE?

So far, various and no fully convincing evidence has been produced to explain how depression may act as a risk factor for vascular disease, including stroke.

A first hypothesis is based on the physiological disturbances linked with depression. These changes have mostly been investigated in relation to cardiovascular disease, and not specifically to stroke. Depression has known neuroendocrine effects, that is, an enhanced activity of the hypothalamic pituitary adrenocortical (HPA) axis and sympathoadrenal hyperactivity[46,47]. HPA axis disturbances predict increased circulating catecholamines, endothelial dysfunction, platelet activation and reduced heart rate variability, which could influence stroke risk[46,48]. The evidence for an involvement of autonomic cardiovascular dysregulation in depressed patients is revealed also by an increased heart rate response to physical stressors, baroreceptor sensitivity and ventricular instability[46,48]. Autonomic dysfunction is also known to influence the risk of atrial fibrillation and depression has been shown to predict atrial fibrillation recurrence after cardioversion[49], suggesting a possible challenging link between depression and stroke that deserves further investigations.

Another mechanism whereby depression can affect the risk of stroke is an inflammation effect[50]. Inflammatory markers such as C-reactive protein, IL-1 and IL-6 have been suggested to be associated with depression[51,52]. Sparse results also pointed towards a role of some genetic[53] and biological markers of thrombotic risk, such as increased levels of fibrinogen[54] and increased serotonin and platelet activation[55]. The hypercoagulable, platelet-activating and inflammatory effects of depression may all be operant in increasing the risk of cardiovascular events in depressed patients. Notwithstanding, the degree to which these predisposing conditions might explain a significant alteration of the risk of stroke in depressed patients is unknown.

Since depression is correlated with major comorbidities, such as hypertension[4] and diabetes[3] probably trough increased adrenergic activity, a third hypothesis is that depression influences stroke risk through the development of hypertension or diabetes or both. In this regard, some authors have suggested that depression may be a sign of preexisting cerebrovascular disease[56]. According to the vascular depression hypothesis, a small-vessel disease secondary to hypertension or diabetes might predispose or at least exacerbate depressive symptoms through impairment of brain regions involved in the regulation of emotions as a direct result of disruption of frontal-subcortical circuits[42]. Data from Rotterdam Study indicate an association between the presence of depressive symptoms and the risk of stroke in the general elderly population, but only in men and not in women and the authors discussed the possibility that depressive symptoms could be the expression of a cerebral vascular damage[27]. In this view, the association between depressive symptoms and later onset ischemic stroke can be considered an epiphenomenon. The vascular depression hypothesis is also supported by the presence and severity of white matter hyperintensities in elderly groups with depression[57,58]. By contrast, data from the Health and Retirement Study, a large national study, provide evidence that depressive symptoms predicted an increased risk of stroke independently of memory impairment, considered a probable early manifestation of cerebral vascular injury[31]. This finding suggests that depression is independently associated with stroke rather than a marker of cerebrovascular damage, possibly through other mechanisms that increase stroke risk.

A fourth hypothesis is that the depression-stroke pathway is modulated by the intervention of behavioral mediators, such as smoking, physical inactivity, poor diet and lack of medication adherence, all of which are modifiable factors. In a prospective cohort study of more than 1000 outpatients with stable coronary heart disease, the association between depressive symptoms and cardiovascular events, including stroke, resulted non significant after adjustment for physical activity and other health behaviors (HR = 1.05; 95%CI: 0.79-1.40), suggesting that the increased risk of cardiovascular events associated with depression could potentially be prevented by behavior modifications, especially physical exercise[8]. Moreover, in depressed patients the control of vascular risk factors may be suboptimal because of non-adherence to medical treatment[59]. In this regard, a meta-analysis showed depression to be a risk factor for reduced medication compliance, with an OR = 3.03 (95%CI: 1.96-4.89)[59].

Finally, depression and stroke might be linked *via* effects of antidepressant medications (ADM).

ANTIDEPRESSANT MEDICATION USE AND STROKE RISK

The trend of antidepressant use has increased in many countries, including the United States and Europe[60,61]. ADM use has recently drawn attention for a possible association with increased risk of cardiovascular events, although such findings remain uncertain[62-65]. There is evidence that ADM exposure is correlated with bleeding complication[66], increased inflammation[67], weight gain[68], cardiac toxic effects[69] and hypertension[70] thus resulting in a possible effects on vascular outcome. The link of antidepressants with stroke development has also been investigated in several studies, with different results. A 20% to 50% increased risk of stroke associated with ADM was shown in a large case-control study[71] and in a case-crossover study[72]. Recently, data from 6 prospective studies published before 2011 were pooled in a secondary analysis of the meta-analysis of Pan and coworkers, which showed a positive association between ADM use and stroke risk, with an estimated HR = 1.41 (95%CI: 1.25-1.59)[39]. Negative findings in randomized trials[73] and case-control studies[74,75] had also been reported. However, misclassification of depression and the absence of information on dose and duration of treatment in

many studies limit the possibility to understand the link between antidepressant use and stroke. A further confounder is the prescription of ADM for conditions other than depression such as insomnia, headaches and neuropathic pain.

Overall, these findings point out a "paradox" related to the fact that depression is a potential risk factor for stroke, but so it appears to be the use of ADM prescribed to treat depression[76]. A key point is whether antidepressants exposure may be considered a surrogate for depression severity, rather than a causal mechanism. The drug-disease association may be expression of underlying differences in vascular risk factor, including depression, among the exposed patients. A recent population-based cohort study found a significant association between depression and risk of stroke regardless of exposure to antidepressant while in patients using only antidepressants an increased risk of stroke was not observed[77]. The authors suggested as explanation the possibility of "confounding by indication", that is the situation in which "the indication for drug use could confound the drug-disease association so that it appears as if the drugs causes the disease"[77].

Among the different classes of antidepressants, selective serotonin reuptake inhibitors (SSRIs) are the first line agents for management of depression today, owing to their relative safety in overdose, tolerability and well-established efficacy. Nevertheless, data from a well-characterized cohort of > 80000 United States middle-aged and elderly women of the Nurses' Health Study during 6 years of follow-up showed that ADM use was associated with an increased stroke risk (HR = 1.30; 95%CI: 1.08-1.55), with a significant association for selective serotonin reuptake inhibitors (HR = 1.39; 95%CI: 1.13-1.72), but not for other ADM[34], suggesting different effects for SSRIs on the risk of stroke in line with what observed in other studies[63,72].

Despite the antiplatelet effect, SSRIs exposure was associated with an increased risk of ischemic stroke[71,72,77]. A possible biological explanation for this association is that serotonin receptors may act on smooth muscle cells leading to a vasoconstriction and, thus, favoring the thrombotic process in atherosclerotic cerebral arteries[78-80].

Consistent with the evidence that SSRIs may increase the risk of bleeding complications owing to the blockade of serotonin reuptake and secondary depletion of platelet serotonin, which may inhibit platelet aggregation[81], a possible association between SSRIs exposure and risk of hemorrhagic stroke has been also reported[63,72], though with mixed results[71]. Recently, a systematic review and meta-analysis of 16 observational studies was performed to determine the association between SSRIs use and risk of brain hemorrhage, showing a significant association with both intracranial (RR = 1.51; 95%CI: 1.26-1.81) and intracerebral bleeding (RR = 1.42; 95%CI: 1.23-1.65)[82]. Moreover, SSRIs use increased significantly the risk of hemorrhagic stroke in patients concomitantly using oral anticoagulants compared with patients receiving only oral anticoagulants (RR = 1.56, 95%CI: 1.33-1.83)[82]. This raises the issue, when considering SSRIs prescription, of an appropriate patient selection. In particular, caution should be used in those subjects with intrinsic risk factors for intracerebral hemorrhage, such as oral anticoagulant exposure, cerebral amyloid angiopathy or severe alcohol abuse[83]. However, more data are needed in this setting.

POST-STROKE DEPRESSION

Mood disorders are common and important sequelae of stroke. Depression, anxiety disorder, apathy, catastrophic reactions and psychosis are frequently observed after stroke[84]. In stroke patients, neuropsychiatric complications may exert a significant adverse impact on the course of motor recovery, social functioning and, overall, on quality of life[84].

DIAGNOSIS OF PSD AND METHODOLOGICAL PROBLEMS

Post stroke depression (PSD) has been associated with increased disability[85,86], impaired rehabilitation outcomes[87,88], and mortality[88,89]. Although many studies have investigated the occurrence of depression among patients with stroke, a real estimate of its prevalence is difficult as a consequence of the wide variability across studies. This is due in part to methodological aspects, such as study population and timing of assessment, and in part to complexity in recognition, assessment, and diagnosis of depression. Furthermore, a contribution to reported differences in the prevalence of PSD may also arise from diagnostic tools used for detection of this disorder. The diagnosis of PSD was assessed on the basis of structured interviews using the diagnostic standards defined by the DSM, while in other studies the assessment is based on the use of cutoff scores in different rating scales. A recent review suggested that, among depression scales, the CES-D, the Hamilton Depression Rating Scale and the Patient Health Questionnaire-9 are adequate options, but they should not be used without a more detailed clinical assessment for an accurate identification of depression in stroke patients[90]. Moreover, there is an obvious risk of under or overestimation in the diagnosis of PSD[91,92]. In fact, stroke may produce somatic symptoms such as fatigue, sleep disturbance, appetite disturbances that might lead to an overdiagnosis of PSD, while post-stroke neurological disabilities, including aphasia or cognitive impairment, may cause under-recognition of PSD. An under-diagnosis of PSD may be also observed when its assessment is made by non-psychiatrists[93].

As a consequence of this, the issue of how to diagnose depression in stroke patients has been the focus of a large number of studies. The presence of physiological symptoms such as psychomotor retardation, and disturbances in appetite, sleep, and sexual interest that can be

related either to stroke or PSD may affect the diagnosis. Stroke is one of the few conditions listed in the DSM-IV as "directly" causing depression. In this case, PSD is classified within the group of "Mood disorders due to stroke, with depressive features" or "with major depressive-like episode"[1].

The validity of DSM diagnostic criteria for depression among patients with stroke has been assessed in different studies. Depressive syndrome in patients with post-stroke major depression is similar to that observed in patients with major depression without a known medical cause[94]. Furthermore, all the symptoms used for the diagnosis of major depression following stroke are significantly more common among depressed patients compared with non-depressed[91,95,96]. However, differences between symptoms in major PSD and primary major depression were also described, with more likely catastrophic reactions, hyper-emotionalism, and diurnal mood variations in patients with PSD[97,98]. In a recent study, Cumming and coworkers confirmed that PSD has a phenomenological profile similar to that of depression unrelated to brain injury regarding psychological and somatic symptoms. This suggests that the diagnosis of depression based on DSM criteria is valid in patients with stroke[99].

EPIDEMIOLOGICAL ASPECTS AND COURSE OF PSD

According to the meta-analysis of Hackett *et al*[6] including studies published up to 2004, the pooled estimated frequency of depression was of 33% (95%CI: 29%-36%) at any time after acute stroke. The assessment of prevalence rates of PSD is complicated by a considerable variation across studies because of the variability in mood assessment, difference in the selection of cases (*i.e.*, variation in stroke features, clinical characteristics, source of patient recruitment), and timing of assessment. A lower prevalence rates has been generally observed in population studies than in studies conducted in acute hospital setting, rehabilitation hospitals or outpatients clinics, suggesting a potential selection bias. Moreover, the risk of depression is expected to be higher in the first few months after stroke. Quite surprisingly, the meta-analysis showed consistency in the overall frequency of depression across population-based, hospital-based and rehabilitation-based studies and different time intervals from stroke onset[6] (Table 2).

A more recent meta-analysis, including data from studies conducted between 1983 and 2011, confirmed that PSD occurs in nearly one third of cases and that this prevalence is independent of time-interval after stroke and study setting[100]. The incidence of PSD has been poorly investigated. Recent data from the South London Stroke Register showed an incidence of depression of 16% in the first year after stroke, of 7%-21% in the 15 years after stroke, and a cumulative incidence of 55%[101]. Moreover, the few studies comparing the incidence of depression in cohorts of stroke patients with that in appropriately matched non-stroke cohorts reported a doubled risk in the former group[102,103].

The natural course of PSD seems to be dynamic and dependent on the timing of onset. Longitudinal studies observed that most patients who have depression after stroke became depressed shortly after the acute event, with a greatest increase in the prevalence of PSD during the first months post-stroke despite the overall disability decreases over time[101,104-106]. Moreover, although a significant proportion of these patients recovered from depression, the occurrence of new cases made the overall prevalence of depression stable over time. About 15%-50% of patients with early onset PSD has been reported to recover in subsequent assessments within 1 year[100], and to have a higher probability of remission in comparison to patients with later onset depressive episodes[106]. Data from a rehabilitation-based study indicates that, at 1 year, 60% of the patients with early depression (0 to 3 mo) had recovered and that those who had not recovered at this follow-up time had a high risk of developing chronic depression[107]. In line with these observations, data from a longitudinal study with a 15-year follow-up showed that half of the patients who were depressed at 3 mo had recovered from depression at 1 year, while the other half recovered gradually between years 2 and 9 and that the proportion of recurrent cases rose from 38% at 2 years to 100% at 15 years[101]. Therefore, depression is often persistent after stroke, with high risk of relapse even after remission over a long period of time.

Stroke survivors have more than six-fold higher risk of developing clinically overt depression even two or more years after index stroke compared to age-matched controls[108]. This suggests that stroke survivors remain at elevated risk for clinically significant depressive symptoms for years after the incident stroke.

According to the results of the Depression in Stroke patients multicenter observational study group, early onset depression appears to be distinct from later onset depression (after 6th month) regarding not only time course but also clinical features. Actually, patients with early occurrence of PSD presented more severe symptoms of depression, assessed using the Montgomery-Asberg Depression Rating Scale, than those developing PSD later[106].

WHAT IS THE PATHOGENESIS OF POSTSTROKE DEPRESSION?

The exact mechanism leading to depression after stroke is still incompletely understood (Figure 1).

In the mid-seventies, the hypothesis that PSD might depend on the anatomic location of brain lesions led to the view of this disorder as a clinical condition related to the interruption of specific pathways involved in mood regulation[109]. Subsequently, many reports suggested that left hemispheric lesions involving frontal region, basal ganglia and those the frontal pole are correlated with an increased risk of PSD[110-114]. The association between le-

Table 2 Pooled prevalence of post stroke depression stratified by study setting and timing assessment in the meta-analysis of Hackett *et al*[6,147] and Ayerbe *et al*[100,101]

	Hackett *et al*[6,147]	Ayerbe *et al*[100,101]
Publication period of included studies	1977-2002	1983-2011
Number of included studies (population: *n*, hospital: *n*, rehabilitation: n)	51 (population: 6, hospital: 16, rehabilitation: 29)	43 (population: 6, hospital: 15, rehabilitation: 22)
Pooled prevalence		
Overall	33% (95%CI: 29%-36%)	29% (95%CI: 25%-32%)
Study setting		
Population-based	AP: 33% (95%CI: 29%-37%) MTP: 33% (95% CI: 0%-72%) LTP: 33% (95%CI: 29%-36%)	22% (95%CI: 17%-28%)
Hospital-based	AP: 36% (95%CI: 0%-73%) MTP: 32% (95% CI: 23%-41%) LTP: 34% (95%CI: 24%-45%)	30% (95%CI: 24%-36%)
Rehabilitation-based	AP: 30% (95%CI: 16%-44%) MTP: 36% (95% CI: 20%-39%) LTP: 34% (95%CI: 26%-42%)	30% (95%CI: 25%-36%)
Timing of assessment		
Acute phase (< 1 mo)	32% (95%CI: 19%-44%)	28% (95%CI: 23%-34%)
Medium-term phase (1-6 mo)	34% (95%CI: 20%-39%)	31% (95%CI: 24%-39%)
Long-term phase (6 mo or more)[1] (6 mo to 1 yr)[2]	34% (95%CI: 29%-39%)	33% (95%CI: 23%-43%)
Very long-term phase (> 1 yr)	n.d.	25% (95%CI: 19%-32%)

AP: Acute phase; MTP: Medium-term phase; LTP: Long-term phase; n.d.: Not determined.

sions involving these anatomic regions and depression was even stronger during the first months after an acute stroke[110,115]. Though interesting, these data have not been consistently replicated and some studies were also reported showing that depression might be associated with right-hemisphere lesions[106,115-118]. Moreover, the hypothesis that depression is influenced by the site of the cerebral lesion was not confirmed in a systematic review by Carson and coworkers. Thus, there is no definitive conclusion on the hypothesis of lateralization and risk for depression[119].

Methodological limitations have been considered to explain these inconsistencies. Boghal and coworkers suggested that the heterogeneity of the results regarding lesions located in the left hemisphere and PSD might depend on whether patients were sampled as inpatients or from the community[120]. Moreover, a significant association of PSD with lesions located in the left hemisphere was found in the first month after stroke, in the right hemisphere after 6 mo[120]. Thus, it is possible that acute PSD and late PSD might be due to different mechanisms. In the acute phase variations in biogenic amines and modulations of serotonin (5HT) receptor may be involved, while in chronic phase PSD may reflect a failure to adapt to changes secondary to stroke, such as impairment in daily activities[121,122].

Several studies have also explored the potential impact of small vascular lesions and chronic ischemic damage in triggering PSD. Based on the concept of vascular depression, chronic ischemic damage could predispose, precipitate or perpetuate depression in the elderly as a consequence of affecting frontal-subcortical circuits responsible for mood control[42]. White matter lesions are conceptualized as a marker of underlying cerebral vascular pathology and they have been described associated with late-onset depression, possibly affecting severity and outcome[123]. In line with this, Brodaty and coworkers found that PSD may be related to accumulation of vascular lesions rather than site and severity of single stroke, supporting the view that biological factors might be an important determinant of PSD[124]. Moreover, in a neuropathological studies of 41 consecutive autopsy cases of patients with stroke it was observed that the vascular burden depending on progressive accumulation of lacunar infarcts within the thalamus and basal ganglia, and of microvascular lesions in deep white matter might have a role in the prediction of PSD[125].

Further support to these findings comes from the growing evidence supporting an association between cerebral microbleeds (CMBs) and occurrence of PSD[126-129]. CMBs are common in ischemic stroke and considered as an indicator of underlying small vessel vasculopathy[130]. There is some evidence indicating that CMBs could affect not only the risk of PSD but also its severity[128]. Furthermore in a recent study, the possibility of non-remission of depression at 1 year follow-up was associated to the presence of lobar CMBs in patients with well-established cerebrovascular disease, suggesting also an influence on outcome[129]. Research on the biology of CBMs may provide useful information on the mechanisms of PSD.

Vascular risk factors might also influence the risk of PSD. In particular, hypertension was found to impact significantly the development of post-stroke depressive symptoms[131,132]. In line with previous findings, hypertension may be linked to depression following stroke through a development of small vessel vasculopathy.

Additionally, several lines of evidence have shown that stroke determines a perturbation of proinflammatory cytokines, which might influence the inflammatory responses implicated in the pathophysiology of depression[133,134], through a physiological dysfunction of brain structures involved in mood control, such as the limbic system[133,135].

However, feeble and often contrasting results have

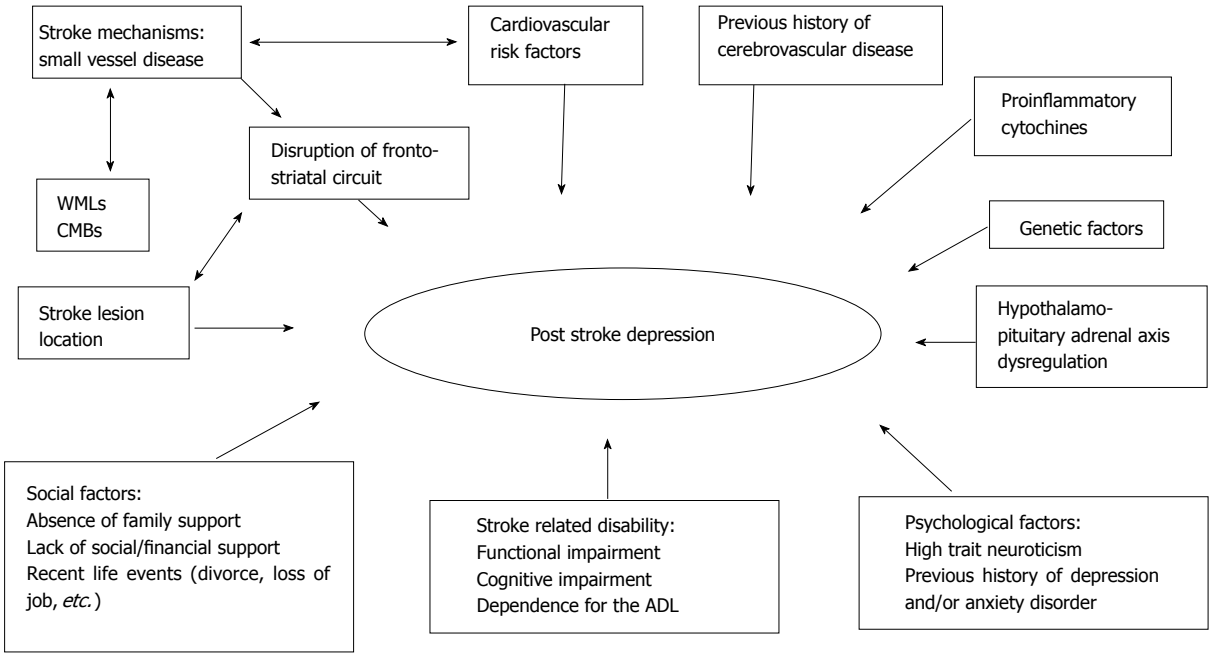


Figure 1 Potential pathogenic pathways of post-stroke depression. WMLs: White matter lesions; CMBs: Cerebral microbleeds; ADL: Activities of daily living.

been reported for the association of PSD with a variety of inflammatory mediators, in particular interleukin (IL)-1beta, IL-6, IL 18, tumor necrosis factor alpha or C-reactive protein[136-139]. A significant association between high serum leptin levels and PSD have also been found at 1-month after stroke[138].

Inconsistent results have also been reported for association between neurotrophic factors such as brain-derived neurotrophic factor (BDNF) and the development of PSD at the acute stage of ischemic stroke[138,140].

The contribution of genetic factors has been also investigated in PSD. A common genetic variant in the promoter region of the serotonin transporter (5-HTT) gene, the short variation length in the 5-HTT-linked polymorphic region (5-HTTLPR, s-allele) has been found to be significantly associated to PSD[141]. Patients carrying the s/s genotype have been also reported to have a better response to psychological intervention for PSD[142]. The val66met polymorphism of BDNF is another variant which may be implicated in PSD[143].

Other predictive factors for PSD, including female gender[144] and previous stroke[118], have been considered with inconsistent results.

Psychosocial factors have been reported to play a role in development of PSD. These include specific personality traits such as premorbid neuroticism[145], previous history of depression[118], living alone, and social isolation with lack of support[107,126]. Major recent life events seem to be a strong risk factor for PSD[118]. In these cases, the overwhelming psychological nature of stroke can trigger a depressive episode in predisposed individuals or in subjects with inadequate social relations. One year after stroke the persistence of few social contacts outside the immediate family contributes to depression[107]. Dependence in the activities of daily living is another important predictor of depression after the first three months[107,146]. In a systematic review of observational studies, Hackett and coworkers found that, despite a wide range of predictive factors, only physical disability, stroke severity, and cognitive impairment resulted consistently associated to PSD[147]. As observed by the authors, however, methodological heterogeneity and the limited number of studies on this topic do not allow firm conclusions on how to identify those patients at the greatest risk of developing depression following a stroke. Depression and physical and cognitive impairment in stroke patients may be associated by a bidirectional causal link. Stroke-related disability may trigger depression which, in turn, may reduce patients' compliance to rehabilitation treatments leading to unfavorable functional outcome[126,147-149].

OUTCOME OF PSD

Depression can exert significant negative impacts on stroke recovery and impair outcome leading to a worsening of cognitive functions, motor abilities and quality of life. It also increases mortality. In a recent prospective population-based study, Ayerbe and coworkers recently reported that the occurrence of depression 3 mo after acute stroke was significantly associated with higher disability, anxiety and a lower quality of life up to 5 years after stroke[150]. Moreover, mortality rate during the 5 years following stroke was higher for patients depressed at 3 mo, and recovering from depression at 1 year did not improve prognosis. These patients, in fact, showed a higher mortality risk during the 5 years after stroke (HR = 1.69; 95%CI: 1.09-2.62) compared with those non-depressed[150]. Additionally, patients with acute PSD were 3.4 times more likely to die during a 10-year follow-up, compared to patients who were non-depressed after

acute stroke[151].

CONCLUSION

Strong evidence supports the view that depression and vascular diseases are deeply related, especially in the elderly. The link between depression and stroke observed in epidemiological analyses appears to be bidirectional, being depression both a precursor and an important consequence of stroke. In spite of a wide literature, the mechanisms underlying this association have not been completely clarified. In this regard, it is necessary that future studies use common methodological approaches, based on accurate description and validated scales for depression. Understanding how depression leads to stroke would allow the development of targeted prevention strategies and interventions aimed at reducing depression-related morbidity and mortality. Identifying the subgroup of stroke patients at highest risk of depression should be the first step. Early identification and treatment of PSD may improve stroke rehabilitation outcomes and decrease mortality.

REFERENCES

1 **American Psychiatric Association.** Diagnostic and statistical manual of mental disorders. Washington (DC): APA, 1994

2 **Kessler RC**, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; **289**: 3095-3105 [PMID: 12813115]

3 **Mezuk B**, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. *Diabetes Care* 2008; **31**: 2383-2390 [PMID: 19033418 DOI: 10.2337/dc08-0985]

4 **Patten SB**, Williams JV, Lavorato DH, Campbell NR, Eliasziw M, Campbell TS. Major depression as a risk factor for high blood pressure: epidemiologic evidence from a national longitudinal study. *Psychosom Med* 2009; **71**: 273-279 [PMID: 19196807 DOI: 10.1097/PSY.0b013e3181988e5f]

5 **Nicholson A**, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *Eur Heart J* 2006; **27**: 2763-2774 [PMID: 17082208]

6 **Hackett ML**, Yapa C, Parag V, Anderson CS. Frequency of depression after stroke: a systematic review of observational studies. *Stroke* 2005; **36**: 1330-1340 [PMID: 15879342]

7 **Carod-Artal J**, Egido JA, González JL, Varela de Seijas E. Quality of life among stroke survivors evaluated 1 year after stroke: experience of a stroke unit. *Stroke* 2000; **31**: 2995-3000 [PMID: 11108762]

8 **Whooley MA**, de Jonge P, Vittinghoff E, Otte C, Moos R, Carney RM, Ali S, Dowray S, Na B, Feldman MD, Schiller NB, Browner WS. Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. *JAMA* 2008; **300**: 2379-2388 [PMID: 19033588 DOI: 10.1001/jama.2008.711]

9 **Evans DL**, Charney DS, Lewis L, Golden RN, Gorman JM, Krishnan KR, Nemeroff CB, Bremner JD, Carney RM, Coyne JC, Delong MR, Frasure-Smith N, Glassman AH, Gold PW, Grant I, Gwyther L, Ironson G, Johnson RL, Kanner AM, Katon WJ, Kaufmann PG, Keefe FJ, Ketter T, Laughren TP, Leserman J, Lyketsos CG, McDonald WM, McEwen BS, Miller AH, Musselman D, O'Connor C, Petitto JM, Pollock BG, Robinson RG, Roose SP, Rowland J, Sheline Y, Sheps DS, Simon G, Spiegel D, Stunkard A, Sunderland T, Tibbits P, Valvo WJ. Mood disorders in the medically ill: scientific review and recommendations. *Biol Psychiatry* 2005; **58**: 175-189 [PMID: 16084838]

10 **Vogt T**, Pope C, Mullooly J, Hollis J. Mental health status as a predictor of morbidity and mortality: a 15-year follow-up of members of a health maintenance organization. *Am J Public Health* 1994; **84**: 227-231 [PMID: 8296945]

11 **Wassertheil-Smoller S**, Applegate WB, Berge K, Chang CJ, Davis BR, Grimm R, Kostis J, Pressel S, Schron E; SHEP Cooperative Research Group (Systolic Hypertension in the elderly). Change in depression as a precursor of cardiovascular events. SHEP Cooperative Research Group (Systoloc Hypertension in the elderly). *Arch Intern Med* 1996; **156**: 553-561 [PMID: 8604962]

12 **Everson SA**, Roberts RE, Goldberg DE, Kaplan GA. Depressive symptoms and increased risk of stroke mortality over a 29-year period. *Arch Intern Med* 1998; **158**: 1133-1138 [PMID: 9605786]

13 **Whooley MA**, Browner WS; Study of Osteoporotic Fractures Research Group. Association between depressive symptoms and mortality in older women. Study of Osteoporotic Fractures Research Group. *Arch Intern Med* 1998; **158**: 2129-2135 [PMID: 9801180]

14 **Jonas BS**, Mussolino ME. Symptoms of depression as a prospective risk factor for stroke. *Psychosom Med* 2000; **62**: 463-471 [PMID: 10949089]

15 **Larson SL**, Owens PL, Ford D, Eaton W. Depressive disorder, dysthymia, and risk of stroke: thirteen-year follow-up from the Baltimore epidemiologic catchment area study. *Stroke* 2001; **32**: 1979-1983 [PMID: 11546884]

16 **Ohira T**, Iso H, Satoh S, Sankai T, Tanigawa T, Ogawa Y, Imano H, Sato S, Kitamura A, Shimamoto T. Prospective study of depressive symptoms and risk of stroke among japanese. *Stroke* 2001; **32**: 903-908 [PMID: 11283390]

17 **Ostir GV**, Markides KS, Peek MK, Goodwin JS. The association between emotional well-being and the incidence of stroke in older adults. *Psychosom Med* 2001; **63**: 210-215 [PMID: 11292267]

18 **May M**, McCarron P, Stansfeld S, Ben-Shlomo Y, Gallacher J, Yarnell J, Davey Smith G, Elwood P, Ebrahim S. Does psychological distress predict the risk of ischemic stroke and transient ischemic attack? The Caerphilly Study. *Stroke* 2002; **33**: 7-12 [PMID: 11779881]

19 **Wassertheil-Smoller S**, Shumaker S, Ockene J, Talavera GA, Greenland P, Cochrane B, Robbins J, Aragaki A, Dunbar-Jacob J. Depression and cardiovascular sequelae in postmenopausal women. The Women's Health Initiative (WHI). *Arch Intern Med* 2004; **164**: 289-298 [PMID: 14769624]

20 **Gump BB**, Matthews KA, Eberly LE, Chang YF; MRFIT Research Group. Depressive symptoms and mortality in men: results from the Multiple Risk Factor Intervention Trial. *Stroke* 2005; **36**: 98-102 [PMID: 15569872]

21 **Avendano M**, Kawachi I, Van Lenthe F, Boshuizen HC, Mackenbach JP, Van den Bos GA, Fay ME, Berkman LF. Socioeconomic status and stroke incidence in the US elderly: the role of risk factors in the EPESE study. *Stroke* 2006; **37**: 1368-1373 [PMID: 16690902]

22 **Stürmer T**, Hasselbach P, Amelang M. Personality, lifestyle, and risk of cardiovascular disease and cancer: follow-up of population based cohort. *BMJ* 2006; **332**: 1359 [PMID: 16687457]

23 **Kamphuis MH**, Kalmijn S, Tijhuis MA, Geerlings MI, Giampaoli S, Nissinen A, Grobbee DE, Kromhout D. Depressive symptoms as risk factor of cardiovascular mortality in older European men: the Finland, Italy and Netherlands Elderly

(FINE) study. *Eur J Cardiovasc Prev Rehabil* 2006; **13**: 199-206 [PMID: 16575273]
24 **Arbelaez JJ**, Ariyo AA, Crum RM, Fried LP, Ford DE. Depressive symptoms, inflammation, and ischemic stroke in older adults: a prospective analysis in the cardiovascular health study. *J Am Geriatr Soc* 2007; **55**: 1825-1830 [PMID: 17916124]
25 **Kawamura T**, Shioiri T, Takahashi K, Ozdemir V, Someya T. Survival rate and causes of mortality in the elderly with depression: a 15-year prospective study of a Japanese community sample, the Matsunoyama-Niigata suicide prevention project. *J Investig Med* 2007; **55**: 106-114 [PMID: 17481379]
26 **Salaycik KJ**, Kelly-Hayes M, Beiser A, Nguyen AH, Brady SM, Kase CS, Wolf PA. Depressive symptoms and risk of stroke: the Framingham Study. *Stroke* 2007; **38**: 16-21 [PMID: 17138952]
27 **Bos MJ**, Lindén T, Koudstaal PJ, Hofman A, Skoog I, Breteler MM, Tiemeier H. Depressive symptoms and risk of stroke: the Rotterdam Study. *J Neurol Neurosurg Psychiatry* 2008; **79**: 997-1001 [PMID: 18208858 DOI: 10.1136/jnnp.2007.134965]
28 **Lee HC**, Lin HC, Tsai SY. Severely depressed young patients have over five times increased risk for stroke: a 5-year follow-up study. *Biol Psychiatry* 2008; **64**: 912-915 [PMID: 18718571 DOI: 10.1016/j.biopsych.2008.07.006]
29 **Liebetrau M**, Steen B, Skoog I. Depression as a risk factor for the incidence of first-ever stroke in 85-year-olds. *Stroke* 2008; **39**: 1960-1965 [PMID: 18451342 DOI: 10.1161/STROKEAHA.107.490797]
30 **Surtees PG**, Wainwright NW, Luben RN, Wareham NJ, Bingham SA, Khaw KT. Psychological distress, major depressive disorder, and risk of stroke. *Neurology* 2008; **70**: 788-794 [PMID: 18316690 DOI: 10.1212/01.wnl.0000304109.18563.81]
31 **Glymour MM**, Maselko J, Gilman SE, Patton KK, Avendaño M. Depressive symptoms predict incident stroke independently of memory impairments. *Neurology* 2010; **75**: 2063-2070 [PMID: 21135381 DOI: 10.1212/WNL.0b013e318200d70e]
32 **Nabi H**, Kivimäki M, Suominen S, Koskenvuo M, Singh-Manoux A, Vahtera J. Does depression predict coronary heart disease and cerebrovascular disease equally well? The Health and Social Support Prospective Cohort Study. *Int J Epidemiol* 2010; **39**: 1016-1024 [PMID: 20360321 DOI: 10.1093/ije/dyq050]
33 **Peters R**, Pinto E, Beckett N, Swift C, Potter J, McCormack T, Nunes M, Grimley-Evans J, Fletcher A, Bulpitt C. Association of depression with subsequent mortality, cardiovascular morbidity and incident dementia in people aged 80 and over and suffering from hypertension. Data from the Hypertension in the Very Elderly Trial (HYVET). *Age Ageing* 2010; **39**: 439-445 [PMID: 20497949 DOI: 10.1093/ageing/afq042]
34 **Pan A**, Okereke OI, Sun Q, Logroscino G, Manson JE, Willett WC, Ascherio A, Hu FB, Rexrode KM. Depression and incident stroke in women. *Stroke* 2011; **42**: 2770-2775 [PMID: 21836097 DOI: 10.1161/STROKEAHA.111.617043]
35 **Seifert CL**, Poppert H, Sander D, Feurer R, Etgen T, Ander KH, Pürner K, Brönner M, Sepp D, Kehl V, Förstl H, Bickel H. Depressive symptoms and the risk of ischemic stroke in the elderly--influence of age and sex. *PLoS One* 2012; **7**: e50803 [PMID: 23226388 DOI: 10.1371/journal.pone.0050803]
36 **Majed B**, Arveiler D, Bingham A, Ferrieres J, Ruidavets JB, Montaye M, Appleton K, Haas B, Kee F, Amouyel P, Ducimetiere P, Empana JP; PRIME Study Group. Depressive symptoms, a time-dependent risk factor for coronary heart disease and stroke in middle-aged men: the PRIME Study. *Stroke* 2012; **43**: 1761-1767 [PMID: 22517599 DOI: 10.1161/STROKEAHA.111.645366]
37 **Jackson CA**, Mishra GD. Depression and risk of stroke in midaged women: a prospective longitudinal study. *Stroke* 2013; **44**: 1555-1560 [PMID: 23686976 DOI: 10.1161/STROKEAHA.113.001147]
38 **Van der Kooy K**, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. *Int J Geriatr Psychiatry* 2007; **22**: 613-626 [PMID: 17236251]
39 **Pan A**, Sun Q, Okereke OI, Rexrode KM, Hu FB. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. *JAMA* 2011; **306**: 1241-1249 [PMID: 21934057 DOI: 10.1001/jama.2011.1282]
40 **Dong JY**, Zhang YH, Tong J, Qin LQ. Depression and risk of stroke: a meta-analysis of prospective studies. *Stroke* 2012; **43**: 32-37 [PMID: 22020036 DOI: 10.1161/STROKEAHA.111.630871]
41 **O'Donnell MJ**, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, Rangarajan S, Islam S, Pais P, McQueen MJ, Mondo C, Damasceno A, Lopez-Jaramillo P, Hankey GJ, Dans AL, Yusoff K, Truelsen T, Diener HC, Sacco RL, Ryglewicz D, Czlonkowska A, Weimar C, Wang X, Yusuf S. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010; **376**: 112-123 [PMID: 20561675 DOI: 10.1016/S0140-6736(10)60834-3]
42 **Alexopoulos GS**, Meyers BS, Young RC, Campbell S, Silbersweig D, Charlson M. 'Vascular depression' hypothesis. *Arch Gen Psychiatry* 1997; **54**: 915-922 [PMID: 9337771]
43 **Newson RS**, Hek K, Luijendijk HJ, Hofman A, Witteman JC, Tiemeier H. Atherosclerosis and incident depression in late life. *Arch Gen Psychiatry* 2010; **67**: 1144-1151 [PMID: 21041615 DOI: 10.1001/archgenpsychiatry.2010.142]
44 **Fiedorowicz JG**, He J, Merikangas KR. The association between mood and anxiety disorders with vascular diseases and risk factors in a nationally representative sample. *J Psychosom Res* 2011; **70**: 145-154 [PMID: 21262417 DOI: 10.1016/j.jpsychores.2010.07.010]
45 **Wu HC**, Chou FH, Tsai KY, Su CY, Shen SP, Chung TC. The incidence and relative risk of stroke among patients with bipolar disorder: a seven-year follow-up study. *PLoS One* 2013; **8**: e73037 [PMID: 24023667 DOI: 10.1371/journal.pone.0073037]
46 **Musselman DL**, Evans DL, Nemeroff CB. The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Arch Gen Psychiatry* 1998; **55**: 580-592 [PMID: 9672048]
47 **Pariante CM**, Lightman SL. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci* 2008; **31**: 464-468 [PMID: 18675469 DOI: 10.1016/j.tins.2008.06.006]
48 **Carney RM**, Freedland KE, Veith RC. Depression, the autonomic nervous system, and coronary heart disease. *Psychosom Med* 2005; **67** Suppl 1: S29-S33 [PMID: 15953797]
49 **Lange HW**, Herrmann-Lingen C. Depressive symptoms predict recurrence of atrial fibrillation after cardioversion. *J Psychosom Res* 2007; **63**: 509-513 [PMID: 17980224]
50 **Shimbo D**, Chaplin W, Crossman D, Haas D, Davidson KW. Role of depression and inflammation in incident coronary heart disease events. *Am J Cardiol* 2005; **96**: 1016-1021 [PMID: 16188535]
51 **Vaccarino V**, Johnson BD, Sheps DS, Reis SE, Kelsey SF, Bittner V, Rutledge T, Shaw LJ, Sopko G, Bairey Merz CN; National Heart, Lung, and Blood Institute. Depression, inflammation, and incident cardiovascular disease in women with suspected coronary ischemia: the National Heart, Lung, and Blood Institute-sponsored WISE study. *J Am Coll Cardiol* 2007; **50**: 2044-2050 [PMID: 18021871]
52 **Howren MB**, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med* 2009; **71**: 171-186 [PMID: 19188531 DOI: 10.1097/PSY.0b013e3181907c1b]
53 **Kendler KS**, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *Am J Psychiatry* 2006; **163**: 109-114 [PMID: 16390897]

54 **Kop WJ**, Gottdiener JS, Tangen CM, Fried LP, McBurnie MA, Walston J, Newman A, Hirsch C, Tracy RP. Inflammation and coagulation factors in persons & gt; 65 years of age with symptoms of depression but without evidence of myocardial ischemia. *Am J Cardiol* 2002; **89**: 419-424 [PMID: 11835923]

55 **Schins A**, Hamulyák K, Scharpé S, Lousberg R, Van Melle J, Crijns H, Honig A. Whole blood serotonin and platelet activation in depressed post-myocardial infarction patients. *Life Sci* 2004; **76**: 637-650 [PMID: 15567189]

56 **Verdelho A**, Ferro JM. Late onset depressive symptoms can be a marker of cerebral vascular pathology. *J Neurol Neurosurg Psychiatry* 2008; **79**: 977 [PMID: 18708565 DOI: 10.1136/jnnp.2007.142364]

57 **Taylor WD**, MacFall JR, Payne ME, McQuoid DR, Steffens DC, Provenzale JM, Krishnan RR. Greater MRI lesion volumes in elderly depressed subjects than in control subjects. *Psychiatry Res* 2005; **139**: 1-7 [PMID: 15927454]

58 **Disabato BM**, Sheline YI. Biological basis of late life depression. *Curr Psychiatry Rep* 2012; **14**: 273-279 [PMID: 22562412 DOI: 10.1007/s11920-012-0279-6]

59 **DiMatteo MR**, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. *Arch Intern Med* 2000; **160**: 2101-2107 [PMID: 10904452]

60 **Olfson M**, Marcus SC. National patterns in antidepressant medication treatment. *Arch Gen Psychiatry* 2009; **66**: 848-856 [PMID: 19652124 DOI: 10.1001/archgenpsychiatry.2009.81]

61 **Reid S**, Barbui C. Long term treatment of depression with selective serotonin reuptake inhibitors and newer antidepressants. *BMJ* 2010; **340**: c1468 [PMID: 20348175 DOI: 10.1136/bmj.c1468]

62 **Whang W**, Kubzansky LD, Kawachi I, Rexrode KM, Kroenke CH, Glynn RJ, Garan H, Albert CM. Depression and risk of sudden cardiac death and coronary heart disease in women: results from the Nurses' Health Study. *J Am Coll Cardiol* 2009; **53**: 950-958 [PMID: 19281925 DOI: 10.1016/j.jacc.2008.10.060.]

63 **Smoller JW**, Allison M, Cochrane BB, Curb JD, Perlis RH, Robinson JG, Rosal MC, Wenger NK, Wassertheil-Smoller S. Antidepressant use and risk of incident cardiovascular morbidity and mortality among postmenopausal women in the Women's Health Initiative study. *Arch Intern Med* 2009; **169**: 2128-2139 [PMID: 20008698 DOI: 10.1001/archinternmed.2009.436]

64 **Tata LJ**, West J, Smith C, Farrington P, Card T, Smeeth L, Hubbard R. General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction. *Heart* 2005; **91**: 465-471 [PMID: 15772201]

65 **Sauer WH**, Berlin JA, Kimmel SE. Effect of antidepressants and their relative affinity for the serotonin transporter on the risk of myocardial infarction. *Circulation* 2003; **108**: 32–36 [PMID: 12821544]

66 **Meijer WE**, Heerdink ER, Nolen WA, Herings RM, Leufkens HG, Egberts AC. Association of risk of abnormal bleeding with degree of serotonin reuptake inhibition by antidepressants. *Arch Intern Med* 2004; **164**: 2367-2370 [PMID: 15557417]

67 **Dawood T**, Lambert EA, Barton DA, Laude D, Elghozi JL, Esler MD, Haikerwal D, Kaye DM, Hotchkin EJ, Lambert GW. Specific serotonin reuptake inhibition in major depressive disorder adversely affects novel markers of cardiac risk. *Hypertens Res* 2007; **30**: 285-293 [PMID: 17541206]

68 **Kivimäki M**, Hamer M, Batty GD, Geddes JR, Tabak AG, Pentti J, Virtanen M, Vahtera J. Antidepressant medication use, weight gain, and risk of type 2 diabetes: a population-based study. *Diabetes Care* 2010; **33**: 2611-2616 [PMID: 20823343 DOI: 10.2337/dc10-1187]

69 **Roose SP**. Treatment of depression in patients with heart disease. *Biol Psychiatry* 2003; **54**: 262-268 [PMID: 12893102]

70 **Licht CM**, de Geus EJ, Seldenrijk A, van Hout HP, Zitman FG, van Dyck R, Penninx BW. Depression is associated with decreased blood pressure, but antidepressant use increases the risk for hypertension. *Hypertension* 2009; **53**: 631-638 [PMID: 19237679 DOI: 10.1161/HYPERTENSIONAHA.108.126698]

71 **Chen Y**, Guo JJ, Li H, Wulsin L, Patel NC. Risk of cerebrovascular events associated with antidepressant use in patients with depression: a population-based, nested case-control study. *Ann Pharmacother* 2008; **42**: 177-184 [PMID: 18212255 DOI: 10.1345/aph.1K369]

72 **Wu CS**, Wang SC, Cheng YC, Gau SS. Association of cerebrovascular events with antidepressant use: a case-crossover study. *Am J Psychiatry* 2011; **168**: 511-521 [PMID: 21406464 DOI: 10.1176/appi.ajp.2010.10071064]

73 **Swenson JR**, Doucette S, Fergusson D. Adverse cardiovascular events in antidepressant trials involving high-risk patients: a systematic review of randomized trials. *Can J Psychiatry* 2006; **51**: 923-929 [PMID: 17249635]

74 **Bak S**, Tsiropoulos I, Kjaersgaard JO, Andersen M, Mellerup E, Hallas J, García Rodríguez LA, Christensen K, Gaist D. Selective serotonin reuptake inhibitors and the risk of stroke: a population-based case-control study. *Stroke* 2002; **33**: 1465-1473 [PMID: 12052976]

75 **Kharofa J**, Sekar P, Haverbusch M, Moomaw C, Flaherty M, Kissela B, Broderick J, Woo D. Selective serotonin reuptake inhibitors and risk of hemorrhagic stroke. *Stroke* 2007; **38**: 3049-3051 [PMID: 17901378]

76 **Bushnell C**. Depression and the risk of stroke in women: an identification and treatment paradox. *Stroke* 2011; **42**: 2718-2719 [PMID: 21921282 DOI: 10.1161/STROKEAHA.111.626895]

77 **Rahman I**, Humphreys K, Bennet AM, Ingelsson E, Pedersen NL, Magnusson PK. Clinical depression, antidepressant use and risk of future cardiovascular disease. *Eur J Epidemiol* 2013; **28**: 589-595 [PMID: 23836399 DOI: 10.1007/s10654-013-9821-z]

78 **Singhal AB**, Caviness VS, Begleiter AF, Mark EJ, Rordorf G, Koroshetz WJ. Cerebral vasoconstriction and stroke after use of serotonergic drugs. *Neurology* 2002; **58**: 130-133 [PMID: 11781419]

79 **Ramasubbu R**. Cerebrovascular effects of selective serotonin reuptake inhibitors: a systematic review. *J Clin Psychiatry* 2004; **65**: 1642-1653 [PMID: 15641869]

80 **Noskin O**, Jafarimojarrad E, Libman RB, Nelson JL. Diffuse cerebral vasoconstriction (Call-Fleming syndrome) and stroke associated with antidepressants. *Neurology* 2006; **67**: 159-160 [PMID: 16832100]

81 **Li N**, Wallén NH, Ladjevardi M, Hjemdahl P. Effects of serotonin on platelet activation in whole blood. *Blood Coagul Fibrinolysis* 1997; **8**: 517-523 [PMID: 9491270]

82 **Hackam DG**, Mrkobrada M. Selective serotonin reuptake inhibitors and brain hemorrhage: a meta-analysis. *Neurology* 2012; **79**: 1862-1865 [PMID: 23077009 DOI: 10.1212/WNL.0b013e318271f848]

83 **McGrath ER**, O'Donnell MJ. Estimating treatment effects in observational studies. *Neurology* 2012; **79**: 1844-1845 [PMID: 23077011 DOI: 10.1212/WNL.0b013e318271f8b6]

84 **Chemerinski E**, Robinson RG. The neuropsychiatry of stroke. *Psychosomatics* 2000; **41**: 5-14 [PMID: 10665263]

85 **Ramasubbu R**, Robinson RG, Flint AJ, Kosier T, Price TR. Functional impairment associated with acute poststroke depression: the Stroke Data Bank Study. *J Neuropsychiatry Clin Neurosci* 1998; **10**: 26-33 [PMID: 9547463]

86 **Singh A**, Black SE, Herrmann N, Leibovitch FS, Ebert PL, Lawrence J, Szalai JP. Functional and neuroanatomic correlations in poststroke depression: the Sunnybrook Stroke Study. *Stroke* 2000; **31**: 637-644 [PMID: 10700497]

87 **Paolucci S**, Antonucci G, Grasso MG, Morelli D, Troisi

E, Coiro P, De Angelis D, Rizzi F, Bragoni M. Post-stroke depression, antidepressant treatment and rehabilitation results. A case-control study. *Cerebrovasc Dis* 2001; **12**: 264-271 [PMID: 11641594]
88 **Naess H**, Lunde L, Brogger J, Waje-Andreassen U. Depression predicts unfavourable functional outcome and higher mortality in stroke patients: the Bergen Stroke Study. *Acta Neurol Scand Suppl* 2010; **190:** 34-38 [PMID: 20586733 DOI: 10.1111/j.1600-0404.2010.01373.x]
89 **House A**, Knapp P, Bamford J, Vail A. Mortality at 12 and 24 months after stroke may be associated with depressive symptoms at 1 month. *Stroke* 2001; **32**: 696-701 [PMID: 11239189]
90 **Meader N**, Moe-Byrne T, Llewellyn A, Mitchell AJ. Screening for poststroke major depression: a meta-analysis of diagnostic validity studies. *J Neurol Neurosurg Psychiatry* 2014; **85**: 198-206 [PMID: 23385849 DOI: 10.1136/jnnp-2012-304194]
91 **Fedoroff JP**, Starkstein SE, Parikh RM, Price TR, Robinson RG. Are depressive symptoms nonspecific in patients with acute stroke? *Am J Psychiatry* 1991; **148**: 1172-1176 [PMID: 1882994]
92 **Williams CL**, Rittman MR, Boylstein C, Faircloth C, Haijing Q. Qualitative and quantitative measurement of depression in veterans recovering from stroke. *J Rehabil Res Dev* 2005; **42**: 277-290 [PMID: 16187241]
93 **Schubert DS**, Burns R, Paras W, Sioson E. Increase of medical hospital length of stay by depression in stroke and amputation patients: a pilot study. *Psychother Psychosom* 1992; **57**: 61-66 [PMID: 1584900]
94 **Lipsey JR**, Spencer WC, Rabins PV, Robinson RG. Phenomenological comparison of poststroke depression and functional depression. *Am J Psychiatry* 1986; **143**: 527-529 [PMID: 3953895]
95 **Paradiso S**, Ohkubo T, Robinson RG. Vegetative and psychological symptoms associated with depressed mood over the first two years after stroke. *Int J Psychiatry Med* 1997; **27**: 137-157 [PMID: 9565720]
96 **Spalletta G**, Ripa A, Caltagirone C. Symptom profile of DSM-IV major and minor depressive disorders in first-ever stroke patients. *Am J Geriatr Psychiatry* 2005; **13**: 108-115 [PMID: 15703319]
97 **Gainotti G**, Azzoni A, Razzano C, Lanzillotta M, Marra C, Gasparini F. The Post-Stroke Depression Rating Scale: a test specifically devised to investigate affective disorders of stroke patients. *J Clin Exp Neuropsychol* 1997; **19**: 340-356 [PMID: 9268809]
98 **Gainotti G**, Azzoni A, Marra C. Frequency, phenomenology and anatomical-clinical correlates of major post-stroke depression. *Br J Psychiatry* 1999; **175**: 163-167 [PMID: 10627800]
99 **Cumming TB**, Churilov L, Skoog I, Blomstrand C, Linden T. Little evidence for different phenomenology in poststroke depression. *Acta Psychiatr Scand* 2010; **121**: 424-430 [PMID: 20384602 DOI: 10.1111/j.1600-0447.2010.01558.x]
100 **Ayerbe L**, Ayis S, Wolfe CD, Rudd AG. Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *Br J Psychiatry* 2013; **202**: 14-21 [PMID: 23284148 DOI: 10.1192/bjp.bp.111.107664ù]
101 **Ayerbe L**, Ayis S, Crichton S, Wolfe CD, Rudd AG. The natural history of depression up to 15 years after stroke: the South London Stroke Register. *Stroke* 2013; **44**: 1105-1110 [PMID: 23404719 DOI: 10.1161/STROKEAHA.111.679340]
102 House A, Dennis M, Mogridge L, Warlow C, Hawton K, Jones L. Mood disorders in the year after first stroke. *Br J Psychiatry* 1991; **158:** 83–92 [PMID: 2015456]
103 **Kase CS**, Wolf PA, Kelly-Hayes M, Kannel WB, Beiser A, D' Agostino RB. Intellectual decline after stroke: the Framingham Study. *Stroke* 1998; **29**: 805-812 [PMID: 9550515]
104 **Townend BS**, Whyte S, Desborough T, Crimmins D, Markus R, Levi C, Sturm JW. Longitudinal prevalence and determinants of early mood disorder post-stroke. *J Clin Neurosci* 2007; **14**: 429-434 [PMID: 17336529]
105 **Aben I**, Verhey F, Strik J, Lousberg R, Lodder J, Honig A. A comparative study into the one year cumulative incidence of depression after stroke and myocardial infarction. *J Neurol Neurosurg Psychiatry* 2003; **74**: 581-585 [PMID: 12700297]
106 **Paolucci S**, Gandolfo C, Provinciali L, Torta R, Toso V; DESTRO Study Group. The Italian multicenter observational study on post-stroke depression (DESTRO). *J Neurol* 2006; **253**: 556-562 [PMID: 16767539]
107 **Aström M**, Adolfsson R, Asplund K. Major depression in stroke patients. A 3-year longitudinal study. *Stroke* 1993; **24**: 976-982 [PMID: 8322398]
108 **Whyte EM**, Mulsant BH, Vanderbilt J, Dodge HH, Ganguli M. Depression after stroke: a prospective epidemiological study. *J Am Geriatr Soc* 2004; **52**: 774-778 [PMID: 15086660]
109 **Carson AJ**. Impact commentaries. Mood disorder as a specific complication of stroke. *J Neurol Neurosurg Psychiatry* 2012; **83**: 859 [PMID: 22291218 DOI: 10.1136/jnnp-2011-301854]
110 **Robinson RG**, Kubos KL, Starr LB, Rao K, Price TR. Mood disorders in stroke patients. Importance of location of lesion. *Brain* 1984; **107** (Pt 1): 81-93 [PMID: 6697163]
111 **Starkstein SE**, Robinson RG, Berthier ML, Parikh RM, Price TR. Differential mood changes following basal ganglia vs thalamic lesions. *Arch Neurol* 1988; **45**: 725-730 [PMID: 3390026]
112 **Starkstein SE**, Robinson RG, Price TR. Comparison of cortical and subcortical lesions in the production of poststroke mood disorders. *Brain* 1987; **110** (Pt 4): 1045-1059 [PMID: 3651794]
113 **Robinson RG**, Kubos KL, Starr LB, Rao K, Price TR. Mood changes in stroke patients: relationship to lesion location. *Compr Psychiatry* 1983; **24**: 555-566 [PMID: 6653097]
114 **Aström M**, Olsson T, Asplund K. Different linkage of depression to hypercortisolism early versus late after stroke. A 3-year longitudinal study. *Stroke* 1993; **24**: 52-57 [PMID: 8418550]
115 **Iacoboni M**, Padovani A, Di Piero V, Lenzi GL. Post-stroke depression: relationships with morphological damage and cognition over time. *Ital J Neurol Sci* 1995; **16**: 209-216 [PMID: 7591672]
116 **MacHale SM**, O'Rourke SJ, Wardlaw JM, Dennis MS. Depression and its relation to lesion location after stroke. *J Neurol Neurosurg Psychiatry* 1998; **64**: 371-374 [PMID: 9527152]
117 **House A**, Dennis M, Warlow C, Hawton K, Molyneux A. Mood disorders after stroke and their relation to lesion location. A CT scan study. *Brain* 1990; **113** (Pt 4): 1113-1129 [PMID: 2397385]
118 **Andersen G**, Vestergaard K, Ingemann-Nielsen M, Lauritzen L. Risk factors for post-stroke depression. *Acta Psychiatr Scand* 1995; **92**: 193-198 [PMID: 7484197]
119 **Carson AJ**, MacHale S, Allen K, Lawrie SM, Dennis M, House A, Sharpe M. Depression after stroke and lesion location: a systematic review. *Lancet* 2000; **356**: 122-126 [PMID: 10963248]
120 **Bhogal SK**, Teasell R, Foley N, Speechley M. Lesion location and poststroke depression: systematic review of the methodological limitations in the literature. *Stroke* 2004; **35**: 794-802 [PMID: 14963278]
121 **Robinson RG**, Bolduc PL, Price TR. Two-year longitudinal study of poststroke mood disorders: diagnosis and outcome at one and two years. *Stroke* 1987; **18**: 837-843 [PMID: 3629640]
122 **Shimoda K**, Robinson RG. The relationship between poststroke depression and lesion location in long-term follow-up. *Biol Psychiatry* 1999; **45**: 187-192 [PMID: 9951566]
123 **Herrmann LL**, Le Masurier M, Ebmeier KP. White matter hyperintensities in late life depression: a systematic review. *J Neurol Neurosurg Psychiatry* 2008; **79**: 619-624 [PMID: 17717021]

124 **Brodaty H**, Withall A, Altendorf A, Sachdev PS. Rates of depression at 3 and 15 months poststroke and their relationship with cognitive decline: the Sydney Stroke Study. *Am J Geriatr Psychiatry* 2007; **15**: 477-486 [PMID: 17545448]

125 **Santos M**, Gold G, Kövari E, Herrmann FR, Bozikas VP, Bouras C, Giannakopoulos P. Differential impact of lacunes and microvascular lesions on poststroke depression. *Stroke* 2009; **40**: 3557-3562 [PMID: 19696424 DOI: 10.1161/STROKEAHA.109.548545]

126 **Choi-Kwon S**, Han K, Choi S, Suh M, Kim YJ, Song H, Cho KH, Nah HW, Kwon SU, Kang DW, Kim JS. Poststroke depression and emotional incontinence: factors related to acute and subacute stages. *Neurology* 2012; **78**: 1130-1137 [PMID: 22459674 DOI: 10.1212/WNL.0b013e31824f8090]

127 **Tang WK**, Chen YK, Lu JY, Chu WC, Mok VC, Ungvari GS, Wong KS. Cerebral microbleeds and depression in lacunar stroke. *Stroke* 2011; **42**: 2443-2446 [PMID: 21757672 DOI: 10.1161/STROKEAHA.111.614586]

128 **Tang WK**, Chen YK, Lu JY, Chu WC, Mok VC, Ungvari GS, Wong KS. Cerebral microbleeds and symptom severity of post-stroke depression: a magnetic resonance imaging study. *J Affect Disord* 2011; **129**: 354-358 [PMID: 20817306 DOI: 10.1016/j.jad.2010.08.007]

129 **Tang WK**, Chen Y, Liang H, Chu WC, Mok VC, Ungvari GS, Wong KS. Cerebral microbleeds as a predictor of 1-year outcome of poststroke depression. *Stroke* 2014; **45**: 77-81 [PMID: 24178917 DOI: 10.1161/STROKEAHA.113.002686]

130 **Werring DJ**, Coward LJ, Losseff NA, Jäger HR, Brown MM. Cerebral microbleeds are common in ischemic stroke but rare in TIA. *Neurology* 2005; **65**: 1914-1918 [PMID: 16380612]

131 **Tennen G**, Herrmann N, Black SE, Levy KS, Cappell J, Li A, Lanctôt KL. Are vascular risk factors associated with post-stroke depressive symptoms? *J Geriatr Psychiatry Neurol* 2011; **24**: 215-221 [PMID: 22228828 DOI: 10.1177/0891988711422526]

132 **Chatterjee K**, Fall S, Barer D. Mood after stroke: a case control study of biochemical, neuro-imaging and socio-economic risk factors for major depression in stroke survivors. *BMC Neurol* 2010; **10**: 125 [PMID: 21192808 DOI: 10.1186/1471-2377-10-125]

133 **Pascoe MC**, Crewther SG, Carey LM, Crewther DP. Inflammation and depression: why poststroke depression may be the norm and not the exception. *Int J Stroke* 2011; **6**: 128-135 [PMID: 21371275 DOI: 10.1111/j.1747-4949.2010.00565.x]

134 **Raison CL**, Miller AH. Is depression an inflammatory disorder? *Curr Psychiatry Rep* 2011; **13**: 467-475 [PMID: 21927805 DOI: 10.1007/s11920-011-0232-0]

135 **Spalletta G**, Bossù P, Ciaramella A, Bria P, Caltagirone C, Robinson RG. The etiology of poststroke depression: a review of the literature and a new hypothesis involving inflammatory cytokines. *Mol Psychiatry* 2006; **11**: 984-991 [PMID: 16894392]

136 **Ormstad H,** Aass HC, Amthor KF, Lund-Sørensen N, Sandvik L. Serum levels of cytokines, glucose, and hemoglobin as possible predictors of poststroke depression, and association with poststroke fatigue. *Int J Neurosci* 2012; **122:** 682-690 [DOI: 10.3109/00207454.2012.709892]

137 **Yang L,** Zhang Z, Sun D, Xu Z, Zhang X, Li L. The serum interleukin-18 is a potential marker for development of post-stroke depression. *Neurol Res* 2010; **32:** 340-346 [DOI: 10.1179/016164110X12656393665080]

138 **Jiménez I,** Sobrino T, Rodríguez-Yáñez M, Pouso M, Cristobo I, Sabucedo M, Blanco M, Castellanos M, Leira R, Castillo J. High serum levels of leptin are associated with post-stroke depression. *Psychol Med* 2009; **39**: 1201-1209 [DOI: 10.1017/S0033291709005637]

139 **Noonan K,** Crewther SG, Carey LM, Pascoe MC, Linden T. Sustained inflammation 1.5 years post-stroke is not associated with depression in elderly stroke survivors. *Clin Interv Aging* 2013; **8:** 69-74 [DOI: 10.2147/CIA.S38547]

140 **Yang L,** Zhang Z, Sun D, Xu Z, Yuan Y, Zhang X, Li L. Low serum BDNF may indicate the development of PSD in patients with acute ischemic stroke. *Int J Geriatr Psychiatry* 2011; **26:** 495-502 [DOI: 10.1002/gps.2552]

141 **Mak KK**, Kong WY, Mak A, Sharma VK, Ho RC. Polymorphisms of the serotonin transporter gene and post-stroke depression: a meta-analysis. *J Neurol Neurosurg Psychiatry* 2013; **84**: 322-328 [PMID: 23236014 DOI: 10.1136/jnnp-2012-303791]

142 **Kohen R**, Cain KC, Buzaitis A, Johnson V, Becker KJ, Teri L, Tirschwell DL, Veith RC, Mitchell PH. Response to psychosocial treatment in poststroke depression is associated with serotonin transporter polymorphisms. *Stroke* 2011; **42**: 2068-2070 [PMID: 21847802]

143 **Kim JM**, Stewart R, Kim SW, Yang SJ, Shin IS, Kim YH, Yoon JS. BDNF genotype potentially modifying the association between incident stroke and depression. *Neurobiol Aging* 2008; **29**: 789-792 [PMID: 17222482]

144 **Poynter B**, Shuman M, Diaz-Granados N, Kapral M, Grace SL, Stewart DE. Sex differences in the prevalence of post-stroke depression: a systematic review. *Psychosomatics* 2009; **50**: 563-569 [PMID: 19996226 DOI: 10.1176/appi.psy.50.6.563]

145 **Aben I**, Denollet J, Lousberg R, Verhey F, Wojciechowski F, Honig A. Personality and vulnerability to depression in stroke patients: a 1-year prospective follow-up study. *Stroke* 2002; **33**: 2391-2395 [PMID: 12364726]

146 **De Ryck A**, Brouns R, Fransen E, Geurden M, Van Gestel G, Wilssens I, De Ceulaer L, Mariën P, De Deyn PP, Engelborghs S. A prospective study on the prevalence and risk factors of poststroke depression. *Cerebrovasc Dis Extra* 2013; **3**: 1-13 [PMID: 23626594 DOI: 10.1159/000345557]

147 **Hackett ML**, Anderson CS. Predictors of depression after stroke: a systematic review of observational studies. *Stroke* 2005; **36**: 2296-2301 [PMID: 16179565]

148 **Parikh RM**, Robinson RG, Lipsey JR, Starkstein SE, Fedoroff JP, Price TR. The impact of poststroke depression on recovery in activities of daily living over a 2-year follow-up. *Arch Neurol* 1990; **47**: 785-789 [PMID: 2357159]

149 **Paolucci S**, Antonucci G, Pratesi L, Traballesi M, Grasso MG, Lubich S. Poststroke depression and its role in rehabilitation of inpatients. *Arch Phys Med Rehabil* 1999; **80**: 985-990 [PMID: 10488996]

150 **Ayerbe L**, Ayis S, Crichton S, Wolfe CD, Rudd AG. The long-term outcomes of depression up to 10 years after stroke; the South London Stroke Register. *J Neurol Neurosurg Psychiatry* 2014; **85**: 514-521 [PMID: 24163430 DOI: 10.1136/jnnp-2013-306448]

151 **Morris PL**, Robinson RG, Andrzejewski P, Samuels J, Price TR. Association of depression with 10-year poststroke mortality. *Am J Psychiatry* 1993; **150**: 124-129 [PMID: 8417554]

P- Reviewer: Sobrino T **S- Editor:** Song XX **L- Editor:**A
E- Editor: Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.64

World J Meta-Anal 2014 August 26; 2(3): 64-70
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

OBSERVATIONAL STUDY

Development of the first Arabic cognitive dental anxiety scale for children and young adults

Abeer Al-Namankany, Paul Ashley, Aviva Petrie

Abeer Al-Namankany, Dental Department, King Fahd Armed Forces Hospital, Jeddah 21159, Saudi Arabia
Paul Ashley, Aviva Petrie, Eastman Dental Institute, University College London, WC1X8LD London, United Kingdom
Author contributions: Al-Namankany A contributed to conducted the research; Ashley P was first supervisor and management support; Petrie A was second supervisor and statistical support.
Correspondence to: Dr. Abeer Al-Namankany, BDS, PGCert Sedation, Dip Hypnosis, MClinDent Paed, MPhil Paed, GCAP, PhD, Paediatric Dentist, Hypnotherapist and Dental Phobia treatment expert, Dental Department, King Fahd Armed Forces Hospital, Prince Mohammed bin Abdul Aziz street, POB 9862, Jeddah 21159, Saudi Arabia. a.alnamankany@yahoo.com
Telephone: +966-26-653000 Fax: +966-26-653000
Received: April 6, 2014 Revised: May 27, 2014
Accepted: June 27, 2014
Published online: August 26, 2014

Abstract

AIM: To validate the Arabic version of abeer children dental anxiety scale.

METHODS: Two ethical approvals for this study were obtained from United Arab Emirates, Ministry of Health and Dubai Health Authority; reference number: 2011/57. The Abeer children dental anxiety scale (ACDAS) was translated from English to Arabic by the native speaker chief investigator, and then back translated by another native speaker in Dubai (AS) to ensure comparability with the original one. Part C of ACDAS was excluded for the schoolchildren because those questions were only applicable for children at the dentist with their parents or legal guardian. A total of 355 children (6 years and over) were involved in this study; 184 in Dubai, 96 from the Religious International Institute for boys and 88 from Al Khansaa Middle School for girls. A sample of 171 children was assessed for external validity (generalizability) from two schools in different areas of London in the United Kingdom.

RESULTS: Receiver operating characteristic curve showed that the cut-off $\geqslant$ 26 for ACDAS gave the optimal results for sensitivity = 90% (95%CI: 81.2%-95.6%), and specificity = 86.6% (95%CI: 78.2%-92.7%), with AUROC = 0.93 (95%CI: 0.90-0.97). Cronbach's Alpha (α) was 0.90 which indicated good internal consistency. Results of the external validity assessing the agreement between ACDAS and dental subscale of the children's fear survey schedule was substantial for the East London school (κ = 0.68, 95%CI: 0.53-0.843); sensitivity = 92.9% (95%CI: 82.7%-98.0%); specificity = 73.5% (95%CI: 55.6%-87.1%) and almost perfect for the Central London school (κ = 0.79; 95%CI: 0.70-0.88); sensitivity = 96.4% (95%CI: 81.7%-99.9%); specificity = 65.9%, (95%CI: 57.4%-73.8%).

CONCLUSION: The Arabic ACDAS is a valid cognitive scale to measure dental anxiety for children age 6 years or over.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Anxiety; Cognition; Children

Core tip: The Abeer children dental anxiety scale (ACDAS) scale is different from existing scales as it is the first dental anxiety scale for children which correlate dental anxiety with cognitive status. It can recognise the stimuli for dental anxiety in a logical order, and has questions concerning the expectation of the child' s legal guardian about the behaviour of the child before the treatment, whether the child has any previous dental treatment experience and the dentist's rating for the child's behaviour at the end of the treatment at the same visit. Finally, when assessing the external validity of the binary ACDAS, it was shown that its results compared favourably with those of the main study (κ = 0.79, sensitivity = 96.4%, specificity = 65.9%) when applied to children in a different London school (κ =

0.68, sensitivity = 92.9%, specificity = 73.5%). Therefore, ACDAS was shown to work well in two different locations with different children, which suggests that it is a generalisable scale. Based on the findings of this study, it is proposed that the ACDAS encompasses the required criteria for the gold standard dental anxiety scale for children.

Al-Namankany A, Ashley P, Petrie A. Development of the first Arabic cognitive dental anxiety scale for children and young adults. *World J Meta-Anal* 2014; 2(3): 64-70 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/64.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.64

INTRODUCTION

Children dental anxiety has been a matter of concern for many years but despite this the etiology is still not entirely understood[1]. Anxiety may occur without cause, or it may be based on a real situation that leads to a reaction that is out of proportion to what would normally be expected. Severe anxiety can have a serious impact on daily life and effect quality of life and its different dimensions, such as speaking, eating, and appearance, and through these also social intercourse[2]. Dental anxiety is cumulative over time, and its development is influenced by multiple variables. It is most likely to start in childhood[3]. There are three general sources of information that have been evaluated as measures of anxiety for children and adults: (1) the behavioral measures, which is what the patient does, such as overt distress, general behavior, or specific motor acts like gripping the chair arms tightly. The results of these measures tend to be more subjective than the objective ones; if two dentists are observing the behavior of the same patient on the same time there is no guarantee that both will score the patient in a similar way. The differences in scoring could depend on the time of the appointment, the experience, the temperament, the age, and the gender of the dentist. Hence, the reliability of these measures will not be strong enough if they are used for research purposes. However, it might be the only method that could be used with preschool children; (2) the physiological measures, which is the measurement of the patient' s responses to the dental anxiety, such as rapid breathing, profuse sweating, muscle tension, pulse rate, or heart rate. These measures were neither reliable nor practical in use with children because the scene, the sound, and the application of the equipment might increase the child's anxiety level[4]. The use of the physiological measures were found to be less appropriate for assessing dental fear in children[5] for several reasons: the standard normal reading for children will vary and depend on age of the child; the results of these measures could be overlapped with current medical problems; the requirement of knowledge and training on how to use these equipments; wrong results by faulty machines; it is not available in all dental clinics; the practicality of using it in terms of cost, time, maintenance and a space in the clinic; last but not least. It is not appropriate to be used for children of all age groups; and (3) the self reported measures, which is what the patient says about his/her fear *via* direct report or scaling, interview, or inventory[6,7]. These measures are the most reliable measures for children who are able to read and have the cognitive ability to understand how to report their anxiety on the scale. Previous studies found that, in adult patients, the self reported anxiety scale can distinguish between high or low dental anxiety in terms of avoidance or distress behaviors. This may not be the case in preschool children; the ability of the young children may not be fully developed and they tend to report more fears regardless of the situation and more likely to show anxiety at separation from the parent. For those young children, the use of the behavioral measures is the best option[8].

Studies of dental anxiety in children rather than in adults may allow us to more reliably investigate the causes and management of dental anxiety. This is due to the limited reliability and validity of adult dental anxiety studies and to the extensive time span between the onset of the anxiety during the childhood and these studies[9]. Although measurement of dental anxiety is important for research and delivery of high quality clinical care, it is the corner stone of dental anxiety management.

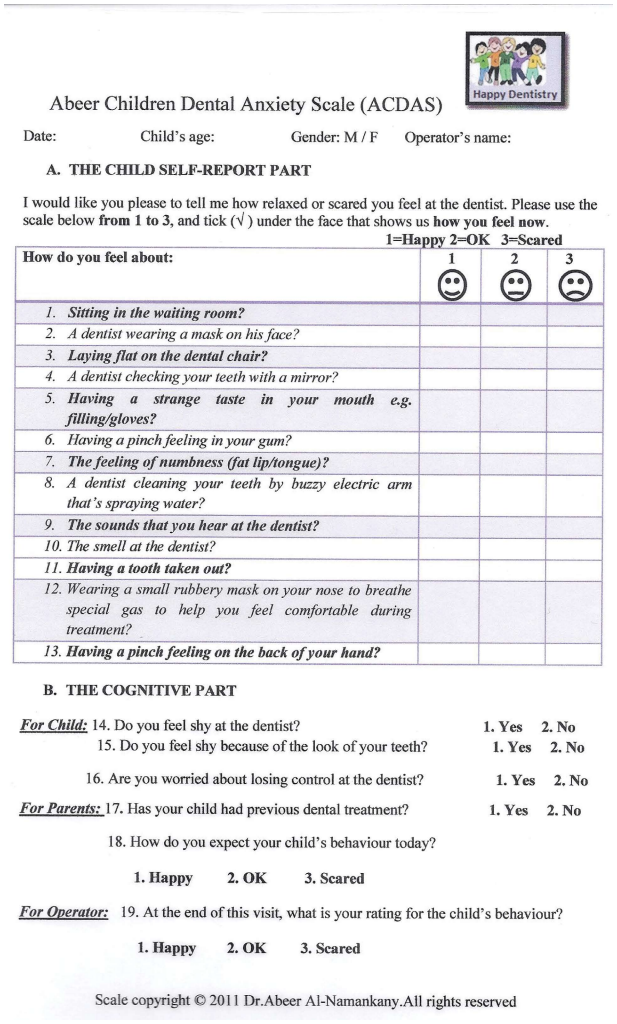
The development of self-reported measures was started in early 1960s and has continued up until the present. Dental anxiety measures have been developed in order to help the dentist detect anxious patients in order to provide better management and treatment. The degree of belief in negative cognition is associated with the severity of DA[10], the negative thinking patterns of the anxious individual is centered on danger and harm. The cognitive measures are widely used as self-report scales that request the patient to respond to list of statements or questions, these measures could be incorporated into pediatric DAM[11]. Abeer children dental anxiety scale (ACDAS) is the first children dental anxiety scale that incorporated the cognitive questions and is a valid cognitive scale to measure DA for children aged $\geqslant$ 6 years[12]. Although there are 14 different dental anxiety scales for children, some of them have been validated in many languages[13]; to date there is no DA scale that validated in Arabic language. Hence, the objective of this study was to validate the Arabic version of ACDAS in order to extend its benefits to more people and to be the first Arabic dental anxiety scale.

MATERIALS AND METHODS

This study was made up of two parts, development of a new scale and then validation of this scale. According to the regulations of the United Arab of Emirates, two ethical approvals for this study were obtained from the Ministry of Health and Dubai Health Authority; Reference number: 2011/57.

Development of the scale

The previously validated (Abeer Children Dental Anxiety



Abeer Children Dental Anxiety Scale (ACDAS)

Happy Dentistry

Date: Child's age: Gender: M / F Operator's name:

A. THE CHILD SELF-REPORT PART

I would like you please to tell me how relaxed or scared you feel at the dentist. Please use the scale below **from 1 to 3**, and tick (√) under the face that shows us **how you feel now**.

1=Happy 2=OK 3=Scared

How do you feel about:	1	2	3
1. *Sitting in the waiting room?*			
2. *A dentist wearing a mask on his face?*			
3. *Laying flat on the dental chair?*			
4. *A dentist checking your teeth with a mirror?*			
5. *Having a strange taste in your mouth e.g. filling/gloves?*			
6. *Having a pinch feeling in your gum?*			
7. *The feeling of numbness (fat lip/tongue)?*			
8. *A dentist cleaning your teeth by buzzy electric arm that's spraying water?*			
9. *The sounds that you hear at the dentist?*			
10. *The smell at the dentist?*			
11. *Having a tooth taken out?*			
12. *Wearing a small rubbery mask on your nose to breathe special gas to help you feel comfortable during treatment?*			
13. *Having a pinch feeling on the back of your hand?*			

B. THE COGNITIVE PART

For Child: 14. Do you feel shy at the dentist? 1. Yes 2. No

15. Do you feel shy because of the look of your teeth? 1. Yes 2. No

16. Are you worried about losing control at the dentist? 1. Yes 2. No

For Parents: 17. Has your child had previous dental treatment? 1. Yes 2. No

18. How do you expect your child's behaviour today?

1. Happy 2. OK 3. Scared

For Operator: 19. At the end of this visit, what is your rating for the child's behaviour?

1. Happy 2. OK 3. Scared

Scale copyright © 2011 Dr.Abeer Al-Namankany.All rights reserved

Figure 1 Abeer children dental anxiety scale-Arabic. Scale copyright © 2011 Al-Namankany. All rights reserved.

Scale "ACDAS")[12] was translated from English to Arabic by the native speaker chief investigator, and then back translated by another native speaker in Dubai to ensure comparability with the original one (Figure 1). ACDAS is a 19 item, cognitive scale which can be used for children from age 6 years and over, we proposed the following name - Abeer Children Dental Anxiety Scale-Arabic (ACDAS-Arabic). It is made up of three parts (Figure 1): (1) this comprises 13 self-reported questions arranged in logical order. Each question uses three faces as a response set. Face "1" represents the feeling of a relaxed not scared "Happy"; face "2" represents a neutral/fair feeling "OK"; and face "3" represents the anxious feeling "Scared". The child is asked to tick under the face that best represents the child's response to the question and a mark (1, 2 or 3) is assigned accordingly. The range of values is therefore from 13 to 39; (2) this comprises three self-reported questions which afford a cognitive assessment, each question uses "Yes" or "No" as a response; and (3) this comprises three questions for further assessment of the child as reported by the legal guardian and the dentist, each question uses"Yes" or "No" as a response.

Validation of the scale

The inclusion criteria for this study were children aged of 6 years or over, with no learning disability, and the ability to read Arabic. The children had to be at least 6 years of age, because younger children do not have the cognitive complexity required to report and react to dental situations accurately and they may not have the experience of dental situations[14]. A convenience sample of 184 students participated in this study; 96 males (The Religious International Institute), and 88 females (Al Khansaa Middle School). The study composed of two parts: assessment of reliability and validity, and assessment of generalizability or external validity.

Assessment of reliability and validity

On the first visit, the local department of school health in Dubai gave permission for the study to be conducted on children in specific schools in their jurisdiction. In addition, permission from each school principal and a verbal consent by the students were also obtained prior to the start of the study. During the class time and in the presence of the teacher for each class, ACDAS-Arabic was completed by each child after being administered twice, once by each of two observers in order to measure the inter-observer reliability, and, in addition, the chief investigator administered ACDAS-Arabic twice, one week apart, to each child in order to measure the intra-observer reliability. Each child on the first visit also completed dental subscale of the children's fear survey schedule (CFSS-DS) after it was administered by the chief investigator in order to assess the validity of ACDAS-Arabic. On the second visit, seven students (4 males/3 females) who participated in the first visit were absent. Therefore there were seven missing from the total sample (n = 184) which resulted in the 177 participants for the analysis.

Statistical methods for numerical anxiety scores: Initially the scores from Part-A of ACDAS-Arabic (the first 13 questions) were summed to provide a numerical anxiety score for each child at each visit. Intra-observer and inter-observer agreement were each assessed by performing a paired t-test to determine if there was a systematic effect, creating a Bland Altman diagram to assess whether the agreement was independent of the magnitude of the score, calculating the British Standards repeatability/reliability coefficient to provide the maximum likely difference between a pair of measurements, and determining Lin's concordance correlation coefficient as a measure of agreement. The Pearson correlation coefficient was determined between the scores of ACDAS-Arabic and CFSS-DS to investigate concurrent validity, and Cronbach's alpha evaluated to assess internal consistency.

Statistical methods for the two anxiety categories (anxious/ not anxious): Creating a categorical outcome facilitates the use of the ACDAS-Arabic scale for clinical and research purposes in terms of translating its numerical score into clinically relevant outcomes (anxious/not anxious). The

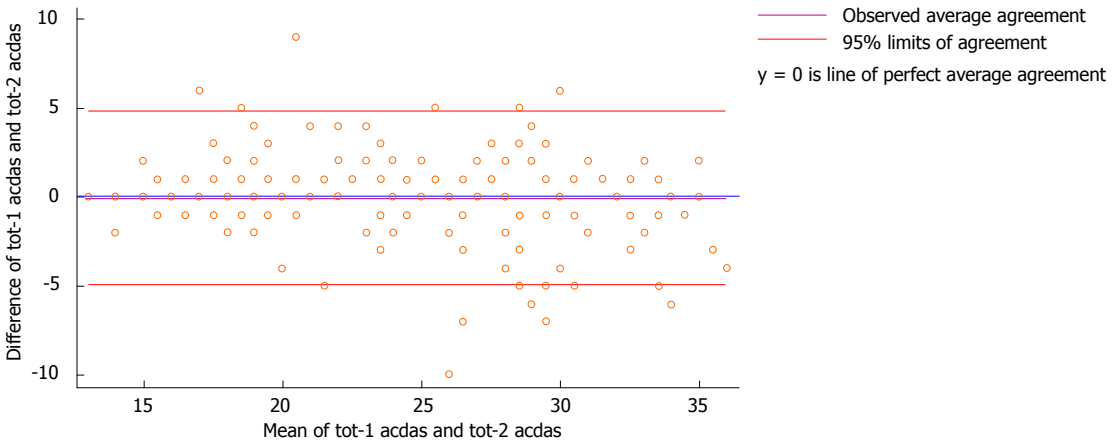


Figure 2 Bland and Altman diagram for the 1st and 2nd scores from one observer.

ACDAS-Arabic questionnaire results from the second visit of the 177 children were used to determine the sensitivity and the specificity for different cut-off values of the total score for Part A to distinguish anxious from not anxious children. The receiver operating characteristic (ROC) curve, plotting the sensitivity against 100, specificity for different cut-offs, was used to select an optimal cut-off value for the new scale. The classification of anxious and not anxious for these 177 children was also determined using the previously published optimal cut-off of ⩾ 36 for the CFSS-DS scale[9].

Cohen's kappa (k) with its confidence interval (CI) was evaluated to assess intra-observer and inter-observer reliability, and the discriminative validity when comparing the binary outcomes of ACDAS-Arabic and CFSS-DS. Convergent validity was assessed by using the Chi-squared test or Fisher's exact test when expected frequencies were small to compare dental anxiety with each of the other variables defined by the questions in Part B of ACDAS-Arabic.

Assessment of eexternal validity (generalizability)

In order to know whether the scale and its dichotomized score will work well in populations that are different from the one from which it was derived, and to assess whether the cut-off "26" of ACDAS produces the similar results in terms of anxiety for different samples of children, a sample of 171 children was assessed for external validity (generalizability) from two schools in different areas of London in the United Kingdom; 81 from St. Alban' s Primary school in Central London and 90 from Cayley primary school in East London. In addition, bootstrapping[15] was used because data had not been collected at other schools on the visit to Dubai and it was not possible to travel to Dubai to collect additional data. Bootstrapping is a simulation process which involves estimating the parameter of interest from each of many random samples of size 177 (in this instance) by sampling with replacement from the original sample of size 177.

RESULTS

Numerical anxiety scores

The analysis of the numerical anxiety scores from ACDAS-Arabic indicated good reliability for both intra- and inter-observer agreement (Lin's concordance correlation coefficient 0.91 (95%CI: 0.89-0.94) and 0.92 (95%CI: 0.90-0.94), respectively; a value of 1 indicates perfect agreement. There was no evidence of a funnel effect in either of the Bland Altman diagrams assessing intra- and inter-observer reliability, and the limits of agreement for them were -4.93 to 4.84 and -3.87 to 5.47, respectively (Figures 2 and 3). The British Standards repeatability/reliability coefficient indicates the maximum likely differences between a pair of measurements were 4.9 and 4.5 for intra-and inter observer reliability, respectively. Using the first set of results for ACDAS-Arabic from the principal observer, the Pearson correlation coefficient between the ACDAS-Arabic and CFSS-DS indicated moderate concurrent validity (r = 0.46, P = 0.007). Cronbach's Alpha (α) for ACDAS-Arabic was 0.90 which indicated a good internal consistency.

Two anxiety categories (anxious/ not anxious)

The sensitivity and the specificity of the ACDAS-Arabic were determined for different cut-off points of the numerical anxiety scores (*i.e.*, the sum of the scores from Questions 1 to 13) as a means of distinguishing anxious from not anxious children. The cut-off point closest to the top left hand corner of the receiver operating characteristic (ROC curve) is circled in red (Figure 4). It gives the optimal results for sensitivity (90.0%, 95%CI: 81.2%-95.6%) and specificity (86.6%, 95%CI: 78.2%-92.7%). The area under the curve was 0.93 (95%CI: 0.90-0.97) as indicated in Figure 3. (A test which is perfect at discriminating between the two outcomes has an area under the curve of one).

There was almost perfect intra-observer agreement for the binary anxiety outcomes, using a cut-off of ⩾ 26 for ACDAS-Arabic to indicate anxiety, when the questionnaire was administered one week apart by the chief investigator (κ = 0.91; 95%CI: 0.85-0.97), and almost perfect inter-observer agreement (κ = 0.89; 95%CI: 0.82-0.96). There was substantial agreement between the two binary anxiety scales (ACDAS-Arabic with a cut-off of ⩾ 26 and CFSS-DS with a cut-off of ⩾ 36) (κ = 0.76; 95%CI: 0.67-0.86),

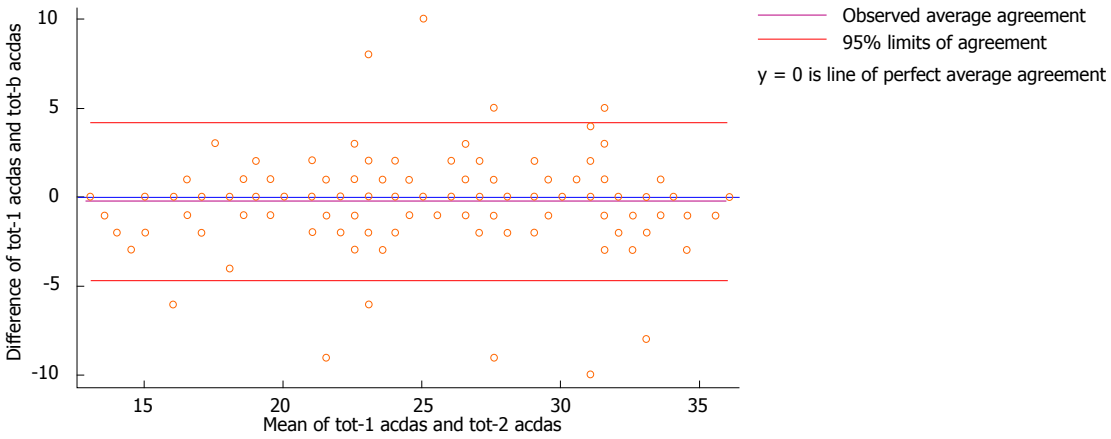


Figure 3 Bland and Altman diagram for the scores by from the two observers (tot-1 and tot-b).

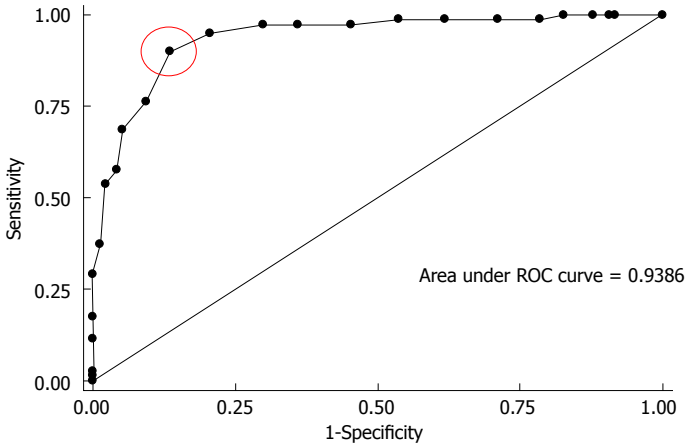


Figure 4 Receiver operating characteristic curve for different cut-offs of Abeer children dental anxiety scale-Arabic.

providing evidence of good discriminative validity.

Convergent validity indicated that there was a strong relationship between DA and cognition. The correlation coefficient is statistically significant ($P = 0.004$) for question 14: "Do you feel shy at the dentist?". There was no evidence of a linear relationship between the score of dental anxiety that was reported by the child and his/her answer to the question 15: "Do you feel shy because of the way your teeth look?" ($P = 0.25$). However, there was a highly significant relationship between child's DA and the cognitive question16: "Are you worried about losing control at the dentist?" ($P < 0.001$).

External validity

One thousand bootstrap replications for 177 observations showed substantial agreement between the two binary scales (ACDAS-Arabic/CFSS-DS). This result compared favorably with the previous result that was obtained from the two schools in London, as shown in Table 1, which suggested that ACDAS-Arabic is working well in another location for another sample and it is a generalizable scale.

DISCUSSION

Given the fact that there is currently no Arabic dental anxiety measure, the idea of the initiator of ACDAS, who is a native Arabic speaker, was to translate ACDAS to Arabic (ACDAS-Arabic) and validate it as the first cognitive and dental anxiety scale in the Arab world. ACDAS was validated as the first cognitive dental anxiety scale for children and adolescents; it included questions about the dental experience in a logical order and not only the most common feared items as the previous scales. Moreover, it included the perception of losing control; embarrassment; self-confidence and the cognitive nature of the child as important factors in anxiety provoking.

This study has shown almost perfect results for both numerical and categorical outcomes; the children had to be at least 6 years of age, because younger children do not have the cognitive complexity required to report and react to dental situations accurately and they may not have the experience of dental situations.

Given the significance of the crucial role of negative cognitive patterns in anxiety evocation that could make the person apprehensive and difficult to treat dentally and who also might not easily comply with anxiety treatment techniques, the present results were in line with the previous similar studies on adults. These demonstrated a strong relation between the negative thoughts and the level of dental anxiety[11,16,17]. The perception of losing control, embarrassment and self-confidence are impor-

Table 1 Comparing Abeer children dental anxiety scale-Arabic (≥ 26) to dental subscale of the children's fear survey schedule (≥ 36)

	Central London	Dubai	East London
Kappa (95%CI)	K = 0.79 (0.70-0.88)	K = 0.76 (0.67 -0.86)	K = 0.68 (0.53-0.84)
Sensitivity (95%CI)	96.40% (81.7%-99.9%)	90% (81.2%-95.6%)	92.90% (82.7%-98.0%)
Specificity (95%CI)	65.90% (57.4%-73.8%)	86.60% (78.2%-92.7%)	73.50% (55.6%-87.1%)

tant factors in anxiety provoking; these results suggest that 91% to 95% of the children who reported negative cognitions on questions 14, 15, and 16 were anxious; 98% were reported in other studies for adults[17]. The cut-off point for anxiety for ACDAS-Arabic (≥ 26) gave the optimal results for sensitivity (86.8%) and specificity (86.2%). These values suggested that ACDAS-Arabic has the ability to identify the anxious and non-anxious individual correctly. In addition, the area under the curve was 0.93: if the score discriminates perfectly, the AUROC equals 1[15].

The strong correlation between the ACDAS-Arabic and the CFSS-DS scores supports the validity of the ACDAS-Arabic in the dental setting, *i.e.*, the ACDAS-Arabic measures what it intends to measure, it includes items that are relevant to the most of children's dental experience and it asks about the child's five sensations (*i.e.*, sight, hearing, taste, smell, and touch). Moreover, it includes items that are relevant to treatment under inhalation and intravenous sedation. Treatment under general anesthesia was not included because the child will be asleep and will not really face the actual dental experience. ACDAS-Arabic is easy to administer and it took a very short time (3 min) to do so.

One of the limitations of this study was the use of CFSS-DS in its English version for validating the ACDAS-Arabic. To date, there is no Arabic DA scale that could be used instead. Therefore the English version of the CFSS-DS had to be translated into Arabic, and read and explained verbally by the chief investigator.

Another limitation was that the order of administration of the ACDAS-Arabic and the CFSS-DS for the school children was not randomized; it was impossible to do this because of the time restriction, as the administration was during the class time. Because of this time constraint, each child could not have a one-to-one interview with the observer in order to complete the ACDAS-Arabic and the CFSS-DS questionnaires. Instead the questionnaires were read to the class as a whole and the all the children in a class completed them at the same time. A third limitation was that the validation of this scale was planned for both a clinical and school setting but, because of the restrictions of cost and the time that the principal investigator would have had to spend in Dubai to obtain the information from a clinical setting, only school children were used.

It is crucial to understand the importance of measuring children dental anxiety and its correlation with the cognitive status of the child. ACDAS helps to highlight the unmet needs of many children who do not go to dentists because of fear of general anesthesia (GA). While some cases may still require GA, with appropriate anxiety management there is a significant number in whom it could be avoided.

Finally, although prevention is better than the treatment, to date there is no study that includes dental anxiety measurement in the list of preventative strategies which usually includes oral hygiene instruction, diet advice, fissure sealant, chlorhexidine and fluoride application[18]. Therefore, the first author suggests the inclusion of dental anxiety measure as a prevention item from the first visit and throughout the dental treatment. Assessing patients' thoughts could be a first step on the development of cognitive treatment strategies for dental anxiety.

The Arabic version of ACDAS is a valid and generalizable cognitive dental anxiety scale for children and adolescents.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Abdul Rahim Almarri, Dr Aisha Sultan, Ms. Fayza Albadri, and the head teachers at all schools in Dubai-UAE. Special thanks to Dr. Paul Ashley who encouraged me to conduct this study in Dubai, and to the children who participated in this study.

COMMENTS

Background

Dental anxiety is still remain as one of the main problems that caused avoiding visiting the dentists. Assessment of dental anxiety is the first step toward a better management.

Research frontiers

The Abeer children dental anxiety scale (ACDAS) scale is different from existing scales as it is the first dental anxiety scale for children which correlate dental anxiety with cognitive status.

Innovations and breakthroughs

It can recognise the stimuli for dental anxiety in a logical order, and has questions concerning the expectation of the child's legal guardian about the behaviour of the child before the treatment, whether the child has any previous dental treatment experience and the dentist's rating for the child's behaviour at the end of the treatment at the same visit.

Applications

Finally, when assessing the external validity of the binary ACDAS, it was shown that its results compared favourably with those of the main study (κ = 0.79, sensitivity = 96.4%, specificity = 65.9%) when applied to children in a different London school (κ = 0.68, sensitivity = 92.9%, specificity = 73.5%).

Terminology

Therefore, ACDAS was shown to work well in two different locations with dif-

ferent children, which suggests that it is a generalisable scale. Based on the findings of this study, it is proposed that the ACDAS encompasses the required criteria for the gold standard dental anxiety scale for children.

Peer review

The work is well done and structured.

REFERENCES

1 **Townend E**, Dimigen G, Fung D. A clinical study of child dental anxiety. *Behav Res Ther* 2000; **38**: 31-46 [PMID: 10645022 DOI: 10.1016/S0005-7967(98)00205-8]

2 **Luoto A**, Lahti S, Nevanperä T, Tolvanen M, Locker D. Oral-health-related quality of life among children with and without dental fear. *Int J Paediatr Dent* 2009; **19**: 115-120 [PMID: 19250394 DOI: 10.1111/j.1365-263X.2008.00943.x]

3 **Tickle M**, Jones C, Buchannan K, Milsom KM, Blinkhorn AS, Humphris GM. A prospective study of dental anxiety in a cohort of children followed from 5 to 9 years of age. *Int J Paediatr Dent* 2009; **19**: 225-232 [PMID: 19486376 DOI: 10.1111/j.1365-263X.2009.00976.x]

4 **Klingberg G**, Berggren U. Dental problem behaviors in children of parents with severe dental fear. *Swed Dent J* 1992; **16**: 27-32 [PMID: 1579885]

5 **Aartman I,** Everdingen T, Hoogstraten J, Schuurs A. Appraisal of Behavioral Measurement Techniques for Assessing Dental Anxiety and Fear in Children: A Review. *J Psychopathol Behav Assess* 1996; **18**: 153-171 [DOI: 10.1007/BF02229114]

6 **Atkins CO**, Farrington FH. Informed consent and behavior management. *Va Dent J* 1994; **71**: 16-20 [PMID: 9540755]

7 **McGrath PA**. Measurement issues in research on dental fears and anxiety. *Anesth Prog* 1986; **33**: 43-46 [PMID: 3458389]

8 **Melamed BG**. Assessment and management strategies for the difficult pediatric dental patient. *Anesth Prog* 1986; **33**: 197-200 [PMID: 3465253]

9 **Boman UW**, Lundgren J, Elfström ML, Berggren U. Common use of a Fear Survey Schedule for assessment of dental fear among children and adults. *Int J Paediatr Dent* 2008; **18**: 70-76 [PMID: 18086029 DOI: 10.1111/j.1365-263X.2007.00863.x]

10 **de Jongh A**, Muris P, ter Horst G, Duyx MP. Acquisition and maintenance of dental anxiety: the role of conditioning experiences and cognitive factors. *Behav Res Ther* 1995; **33**: 205-210 [PMID: 7887880 DOI: 10.1016/0005-7967(94)P4442-W]

11 **Ayer W**. Psychology and dentistry: mental health aspects of patient care. 1st ed. New York: Haworth Press Inc., 2005

12 **Al-Namankany A**, Ashley P, Petrie A. The development of a dental anxiety scale with a cognitive component for children and adolescents. *Pediatr Dent* 2012; **34**: e219-e224 [PMID: 23265158]

13 **Al-Namankany A**, de Souza M, Ashley P. Evidence-based dentistry: analysis of dental anxiety scales for children. *Br Dent J* 2012; **212**: 219-222 [PMID: 22402535 DOI: 10.1038/sj.bdj.2012.174]

14 **Alwin NP**, Murray JJ, Britton PG. An assessment of dental anxiety in children. *Br Dent J* 1991; **171**: 201-207 [PMID: 1910981 DOI: 10.1038/sj.bdj.4807661]

15 **Petrie A**, Sabin C. Medical Statistics at a Glance. 3rd ed. USA: Blackwell Publishing, 2009

16 **Briers S.** Brilliant Cognitive Behavioural Therapy. 1st ed. USA: Prentice Hall, 2009

17 **De Jongh A**. Dental Anxiety: A Cognitive Perspective. Thesis, 1995

18 **Sarmadi R**, Gahnberg L, Gabre P. Clinicians' preventive strategies for children and adolescents identified as at high risk of developing caries. *Int J Paediatr Dent* 2011; **21**: 167-174 [PMID: 20961342]

P- Reviewer: Lopez-Jornet P, Tomofuji T **S- Editor:** Ji FF
L- Editor: A **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.71

World J Meta-Anal 2014 August 26; 2(3): 71-77
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

Interventions to improve treatment adherence among adolescents: A meta-analysis

Tricia A Miller, Brittany L Bannon, M Robin DiMatteo

Tricia A Miller, Brittany L Bannon, M Robin DiMatteo, Department of Psychology, University of California, Riverside, CA 92507, United States

Author contributions: Miller TA and DiMatteo MR were responsible for the conception and design of the study; Miller TA and Bannon BL collected the data, organized, and coded relevant empirical articles; Miller TA and Bannon BL statistically analyzed the data and performed moderator analysis; Miller TA and Bannon BL drafted the manuscript with critical revisions performed by DiMatteo MR.

Correspondence to: Tricia A Miller, MA, Department of Psychology, University of California, 900 University Avenue, Riverside, CA 92507, United States. trishmiller122@gmail.com
Telephone: +1-951-8275243 Fax: +1-951-8273985
Received: November 29, 2013 Revised: April 17, 2014
Accepted: July 17, 2014
Published online: August 26, 2014

Abstract

AIM: To examine the overall effectiveness of interventions designed to improve medical treatment adherence among adolescent patients.

METHODS: PubMed and PsycINFO databases were searched to retrieve and analyze empirical journal articles (from 1948-2013). Only peer-reviewed, English language journals that defined a measure of adherence (or compliance), assessed an intervention aimed at improving adherence among adolescents, and provided information to calculate an *r* effect size were included. Studies were excluded if they lacked assessment of the effectiveness of interventions on improving adherence in adolescents as compared to no interventions or standard care. Case studies or journal articles that examined substance abuse or psychological disorders were also excluded. Analyses were conducted with fixed and random-effects methods, and moderators of intervention efficacy were also examined.

RESULTS: For each study that met the inclusion criteria (n = 45), an effect size r, reflecting the strength and direction of the interventions' relationship to adherence was recorded; a positive r indicated that the intervention increased adolescent adherence, whereas a negative r indicated that the intervention decreased adolescent adherence. The overall effectiveness of adolescent adherence interventions was positive and significant (unweighted mean r = 0.27, 95%CI: 0.21-0.33, P = 0.001). Moderator analyses at the fixed effects level revealed that interventions were less effective when adolescents reported their adherence behaviors, when the type of adherence regimen was a medication regimen, and when the type of intervention was cognitive-modification based.

CONCLUSION: These findings contribute to understanding interventions for enhancing adolescent adherence. Future research should continue to examine the specific challenges faced by adolescents and create targeted interventions.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Adolescent; Adherence (compliance); Intervention; Meta-analysis

Core tip: Estimates of nonadherence among the adolescent population range from 25%-70%, depending on the disease or condition. Intervention components in patient samples vary widely across studies; thus, it is important to systematically identify elements of interventions that are most effective. Meta-analytic techniques were used in this study to provide a comprehensive, quantitative summary of empirical studies evaluating the effectiveness of interventions aimed at improving treatment adherence among adolescents. This meta-analysis showed that interventions were effective, specifically when the type of regimen was behavioral, whereas cognitive-based interventions were less effective.

Miller TA, Bannon BL, DiMatteo MR. Interventions to improve treatment adherence among adolescents: A meta-analysis. *World J Meta-Anal* 2014; 2(3): 71-77 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/71.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.71

INTRODUCTION

Patient nonadherence (or noncompliance) to medical treatment involves the degree to which an individual fails to follow specific disease management activities as directed by his or her healthcare provider[1,2]. Nonadherence can occur in the context of a variety of treatment behaviors, such as medication use, electronic pill counts, diet and exercise, and disease management skills[1]. The prevalence of nonadherence can be close to 25% on average, across a variety of diseases, and specific patient populations[2]. For some chronic diseases such as diabetes, that require complex treatment regimens, rates of nonadherence can exceed 50%[2]. Moreover, it is estimated that nearly 240 million medical visits are wasted annually due to nonadherence[2]. Despite these alarming statistics, patients and healthcare providers remain largely unaware that one major cause of poor health outcomes may be patient nonadherence[3,4].

Adolescent adherence relationship

Adolescence is a unique period of the lifespan in which individuals begin to explore their social identities and seek independence; however, among adolescents coping with illnesses, this autonomy-seeking can bring about confusion, frustration, and conflict with caregivers over prescribed treatment regimens, and consequently, nonadherence. In a study conducted by DiGirolamo *et al*[5] 30% of adolescents with cystic fibrosis reported completing less than half of their daily-prescribed airway clearance regimens that are essential for the prevention of further morbidity and even mortality. In another study of adolescent patients with Type 1 diabetes, 25% reported mismanagement behaviors, such as missing prescribed insulin injections. Similarly, between 50 and 60% of children and adolescents were found to underuse prescribed medications, and less than 10% overused prescribed medications[6,7]. A study by Chappuy *et al*[8] looked at prescription medication adherence in adolescents and found that only 36.2% actually completed their prescribed regimens. Lastly, Guilfoyle *et al*[9] noted that nonadherence to an oral immunosuppressant medication regimen, commonly used to prevent a patient's body from rejecting an organ transplant, is prevalent (approximately 70%) and has been found to significantly compromise the long-term graft survival and life span of adolescents with kidney transplants. Nonadherence is a prevalent and consequential issue for adolescent patients and their families; however, the development of effective interventions to improve adherence behaviors in this age group remains an ongoing challenge. The various types of interventions established to date and their benefits are reviewed below.

Interventions to improve adherence in adolescents

Studies of current interventions to promote adherence among adolescents have consistently shown that educational interventions alone are not sufficient to change adherence behaviors[10]. In fact, a meta-analysis conducted by Dean *et al*[11] revealed that a multifaceted approach to interventions showed the greatest potential efficacy in improving adherence behaviors. The optimal combination of intervention elements remains unclear, however. Dean *et al*[11] suggest that the combination of education and behavioral methods (*e.g.*, reinforcement) for increasing motivation and providing problem-solving strategies can produce the greatest results. Additionally, in an intervention study conducted by Wysocki *et al*[10], behavioral family systems therapy improved both adherence to treatment and family relationship quality among adolescents, providing evidence for the importance of social support and family cohesion in disease management. Moreover, in a separate intervention conducted by Wysocki *et al*[12], researchers looked at adherence to self-monitoring of blood glucose, utilizing a behavioral intervention that compared two groups: a pill count meter-alone group and a pill count meter-plus-behavioral contract group. Results from this study indicated that both groups showed moderate improvement in measures of diabetic control, demonstrating the value of behavioral reminders and patients' commitment to their own care.

In addition to interventions combining educational and behavioral components, those incorporating cognitive-behavioral principles have also been successful. For example, research by van Es *et al*[13] found that adolescents with asthma demonstrated better treatment adherence if they received both education and group therapy for disease-focused issues, including attitudes toward disease and coping skills. In another study by Magyary *et al*[14], a cognitive-behavioral intervention significantly increased therapeutic adherence and self-responsibility for the management of health conditions in children and adolescents. Although the components of adolescent adherence interventions vary widely from study to study, and aspects have yielded results in particular patient samples, it is important that the overall efficacy of this wide range of interventions is subjected to systematic, quantitative review.

Present study

The purpose of the present study is to utilize meta-analytic techniques to review and summarize research findings on the effectiveness of interventions designed to improve medical adherence in adolescents. Additionally, potential moderators of the effectiveness of these interventions will be examined. More specifically, this meta-analysis will test the primary hypothesis that there is a positive relationship between adherence interventions and adolescents' adherence to their medical treatment regimens, such that interventions would, on average,

improve adherence among adolescents. Exploratory moderator analyses will also be used to evaluate whether the following factors explain variability in the intervention studies' effect sizes: type of sample (adolescent only versus children and adolescents combined), type of intervention (*e.g.*, educational, behavioral), type of adherence measure (*e.g.*, self-report, electronic), type of treatment regimen (*e.g.*, medication-based, behavioral), type of illness, parental involvement in the intervention, patient gender, and patient ethnicity.

MATERIALS AND METHODS

Literature search

A "Top-down" literature search was conducted for the retrieval and analysis of empirical journal articles published from 1948 through 2013. PubMed and PsycINFO databases were searched using combinations of the following keywords: adherence (compliance), persistence AND adolescent, youth, teens, children, interventions, disease management, self-management, randomized control trial. In addition, the reference sections of obtained journal articles were examined for relevant studies.

Inclusion and exclusion criteria

Studies were included if published in a peer-reviewed, English language journal, if they defined and explained a method of measuring adherence (or compliance), and if they assessed an intervention that aimed to improve adherence to medical treatment among adolescents. Studies were included only if they provided an effect size *r* or statistical information to calculate an effect size representing the magnitude and direction of the interventions' effect on adherence. Meta-analytic techniques were used to extract average *r* effect sizes and assess their significance. Furthermore, relevant journal articles were coded for moderators of the interventions' effectiveness, including: age or age range, total *N*, location of the study (United States or non-United States), type of patient illness, type of intervention, how adherence was measured, type of regimen, whether or not there was parental involvement in the management of care, patient gender, and patient ethnicity. Studies were excluded if they: measured the effectiveness of interventions on adherence in adult patients; did not ASSESS patient adherence to treatment; lacked an intervention to improve adherence; lacked assessment of the effectiveness of an intervention on improving adherence in adolescents; and/or provided no information to calculate an *r* effect size. Finally, case studies or journal articles that examined substance abuse or psychological disorders (*i.e.*, depression, bipolar disorder) were excluded, because treatment adherence in mental health and substance use is beyond the scope of the present research, although it is an important issue for future examination.

Effect size calculation

An *r* effect size was calculated from Cohen's *d*, *P* or descriptive statistics. If a study reported a significant result but did not report a *P*-value, then the one-tailed *P*-value was assumed to be 0.025. If a study reported results that were non-significant and no exact *P*-value was provided, then the study was conservatively assigned $P = 0.5$ one-tailed, and $r = 0$[15]. An effect size of $r = 0$ indicated that the intervention did not have an effect on adolescent adherence. For studies in which there were multiple measures of adherence, the *r* for each measure was converted to a Fisher's Z_r and they were averaged.

Statistical analysis

The effect size *r* was used because *r* most clearly illustrates both the *strength* (from 0.00 to 1.00) and *direction* (positive or negative) of the relationship between variables[15,16]. In this meta-analysis, a positive *r* indicates an improvement in adherence as a result of the intervention, whereas a negative *r* indicates a decrease in adolescent adherence to medical treatment as a result of the intervention (relative to control or standard care groups). The effect size *r* was obtained for each of the studies. If statistics were provided that could be transformed into an *r* (*e.g.*, *t*, Z and *P*-value, chi-square, or 1 degree of freedom in the numerator *F*; or means and standard deviations), the appropriate statistical analyses were conducted to yield a Phi, Pearson Product-Moment, or point-biserial correlation coefficient[17]. All calculations involving *r* were performed by transforming *r* to the Fisher's *Z* transformation of *r* and then returning the results back to the scale of *r*.

The random effects model was used to combine effect size statistics using the unweighted mean *r* based on *k* (the total number of studies included). This method allows for the generalization of findings to other studies beyond those that were included in this meta-analysis[15,18]. The fixed effects model was also used to carry out weighted mean analyses and tests of heterogeneity based on *N* (the total number of participants across all studies). All analyses of moderators were first performed using the random effects model; if results were not significant at the random effects level, results from the fixed effects model were provided. Random effects tests of methodological and substantive moderators were conducted to examine the heterogeneity of the study effects. These include: sample type (adolescent and children versus adolescents only), total *N*, location of the study (United States or non-United States), patient illness, type of intervention, how adherence was measured, type of regimen, whether or not there was a parent involved in the management of care, patient gender and patient ethnicity. In addition, for the effects that were significant, the fail safe *N* was calculated (to address the file drawer problem) that indicated the number of studies, new, unpublished or unretrieved with no effect that would be needed in order for significant results to be declared non-significant at $P < 0.05$[17]. The standardized odds ratio and standardized relative risk (including 95%CI) were calculated from the unweighted mean *r* using the binomial effect size display

(BESD). The BESD is a useful tool for effect size estimation that can be used to display changes in success rates (*i.e.*, survival or improvement rates) that are attributable to specific treatment procedures[17]. Preliminary statistical analyses were conducted using SPSS 12.0 (*i.e.*, calculation of means, medians, standard deviations, correlations and *t*-tests). A T1-84 Plus graphing calculator and Excel 2008 v.12.2.3 were used for essential calculation verification.

Results

Meta-analytic calculations were performed on 45 independent studies to examine the overall effectiveness of interventions aimed at improving adherence to medical treatment among adolescents. Twenty-four of the 45 studies combined the results of the interventions' effectiveness on adherence among both adolescents and pediatric patients. The remaining 21 studies looked at the effectiveness of interventions aimed at improving adherence among adolescents only. Therefore, meta-analytic computations were done for the total sample (k = 45) and also separately for these two groups, constituting a moderator analysis for "type of sample." In addition, for each sample (k = 45, and the subgroups of k = 24 adolescents plus children, k = 21 adolescents only) the following statistics were computed: the total number of subjects (N), the median r and range, the fixed effects weighted mean r with a 95%CI, the random effects model unweighted mean r with a 95% confidence interval, the fail safe N, the standardized odds ratio with a 95%CI, and the standardized relative risk with a 95% confidence interval (Table 1).

Interventions to improve adherence

Across 45 independent studies, with a total of 3890 participants, the average relationship between an adherence intervention and improvement in adolescent adherence (as compared to a control group or to a group receiving standard care) was positive and significant under the random effects model (unweighted mean r = 0.27, P < 0.001). This demonstrated that interventions aimed at improving adolescent adherence were effective. The median r of 0.23 was close in magnitude to the unweighted mean and to the weighted (by sample size) mean r of 0.18. Effect sizes in the positive direction indicated that interventions aimed at increasing adherence were effective; conversely, effect sizes in the negative direction indicated that these interventions reduced adherence. Within this sample of studies, there were only two negative r effect sizes: -0.24 and -0.05.

Both the fixed (weighted) and random (unweighted) effects models indicated a positive and significant effect of interventions on improving adherence to medical treatment in adolescents (Table 1). The random effects model indicated that adherence interventions were effective [r = 0.27, $t_{(44)}$ = 7.55, P < 0.001]. Therefore, the effectiveness of interventions to improve adherence among adolescents can be generalized to studies outside the present sample. The fixed effects model also indicated that interventions aimed at improving adherence were effective (r = 0.18, 95%CI: 0.15-0.21; P < 0.001). The fail safe N demonstrated that more than 2189 studies with non-significant results would have to be included in order for these results to be rendered non-significant; this number exceeds the tolerance level of 235 unpublished (or otherwise non-retrievable) studies with null results that might possibly exist. The standardized odds ratio, using the BESD, indicated that the odds of being adherent to medical treatment were 3.03 times higher if adolescents participated in an adherence intervention compared with the odds if he/she did not participate in an intervention (95%CI: 2.35-3.94; P < 0.001). The standardized relative risk (also calculated using the BESD) indicated that the risk of poor adherence to medical treatment was 1.74 times higher if adolescents did not participate in an adherence intervention (95%CI: 1.53-1.99; P < 0.001). Furthermore, the effect sizes were significantly heterogeneous (χ^2 = 122.82, P = 2.24 × 10^{-9}), which indicated that moderators might account for this variation in effect sizes.

Moderator analyses

Moderator analyses using the random effects model revealed no significant results in studies that combined children and adolescents and in studies of adolescents only. Meta-analytic calculations were applied to each sample type separately and are detailed below.

In the 24 studies that included a combined sample of adolescents and children, there was a moderate, significant and positive effect of interventions using both the fixed (weighted) and random (unweighted) effects models (Table 1). The random effects model indicated that interventions to improve adherence to medical treatment in adolescents were effective [r = 0.32, $t_{(23)}$ = 5.58, P < 0.001]. Additionally, the weighted mean (fixed effects model) yielded a similarly positive result (r = 0.27, 95%CI: 0.22-0.32; P < 0.001). The fail safe N demonstrated that more than 800 studies with non-significant results would have to be included in order for these results to be rendered non-significant. However, the tolerance level suggested that 130 unpublished null studies might possibly exist. The BESD-based, standardized odds ratio indicated that the odds of adhering to medical treatment were 3.77 times higher if patients participated in an adherence intervention as compared to the odds if they had not (95%CI: 2.44-5.99; P < 0.001). The standardized relative risk indicated that the risk of nonadherence to medical treatment was 1.94 times higher if the patient did not participate in an adherence intervention (95%CI: 1.56-2.45; P < 0 001). Furthermore, the 24 effect sizes were heterogeneous, (χ^2 = 53.36, P = 3.28 × 10^{-4}). Inspection of the distribution of r's revealed a range from r = -0.24 to r = 0.71.

In the comparison subgroup of 21 studies that assessed adolescent patients only, analyses revealed a moderate, yet positive and significant, effect of interventions on improvements in adherence to medical treatment,

Table 1 Summary of overall meta-analysis results

Effect of adolescent adherence interventions	K[4]	Total *n*[5]	Unweighted median *r* (range)	Weighted mean *r*[6] (95%CI)	Unweighted mean *r*[7] (95%CI)	Fail safe *n*	Standardized odds ratio[8] (95%CI)	Standardized relative risk[9] (95%CI)
Interventions[1]	45	3890	0.23 (-0.24-0.71)	0.18 (0.15-0.21)[b]	0.27 (0.21-0.33)[b]	2189 (tolerance level 235)	3.03 (2.35, 3.94)[b]	1.74 (1.53-1.99)[b]
Adolescent and Children Interventions[2]	24	1476	0.325 (-0.24-0.71)	0.27 (0.22-0.32)[b]	0.32 (0.22-0.42)[b]	800 (tolerance level 130)	3.77 (2.44, 5.99)[b]	1.94 (1.56, 2.45)[b]
Adolescent Only Interventions[3]	21	2414	0.18 (-0.05-0.51)	0.12 (0.08-0.16)[b]	0.20 (0.13-0.27)[b]	324 (tolerance level 115)	2.25 (1.69-3.03)[b]	1.5 (1.30-1.74)[b]

[b]$P < 0.01$. [1]Analyses for the entire sample of intervention studies; [2]Analyses for the subgroup of studies that combined adolescent with pediatric samples; [3]Analyses for the subgroup of studies that included adolescent-only samples; [4]Number of samples; [5]Total *n* across all samples; [6]Effect size obtained from the fixed effects model, or weighted by the total number of participants across studies. [7]Effect size obtained from the random effects model, or based on the total number of studies included. [8]The standardized odds ratio depicts the odds of being adherent in the intervention group relative to the control group. Across all analyses, the intervention group had a higher likelihood of improved adherence than the control group (receiving no interventions or standard care). [9]The standardized relative risk can be interpreted as the control group's risk for nonadherence as compared to the intervention group's risk. In each set of analyses, the control group was at significantly greater risk for nonadherence. The Fail Safe *n* exceeds the level of tolerance for future null results making it unlikely that the "file drawer problem" is a source of bias. The binomial effect size display from the unweighted mean effects (random effects model) was used to obtain the standardized odds ratio and relative risk. The heterogeneity test ($k = 45$) for the overall adolescent adherence interventions was significant ($\chi^2 = 122.82$, $P < 0.001$). The heterogeneity test ($k = 24$; adolescent and children samples) for adolescent adherence interventions was significant ($\chi^2 = 53.36$, $P < 0.001$). The heterogeneity test ($k = 21$; adolescent only samples) for the second group of adolescent adherence interventions was also significant ($\chi^2 = 44.53$, $P < 0.001$).

using both the fixed (weighted) and random (unweighted) effects models (Table 1). First, the random effects model indicated that adolescent-only adherence interventions were effective [$r = 0.20$, $t_{(20)} = 5.18$, $P < 0.001$]. The fixed effects model revealed the same significant intervention efficacy, albeit with a slightly smaller effect ($r = 0.12$, 95%CI: 0.08-0.16; $P < 0.001$). In addition, the fail safe *N* demonstrated that more than 324 studies with non-significant results would have to be included in order for these results to be rendered non-significant. The tolerance level suggested that 115 unpublished null studies possibly exist. The standardized odds ratio (BESD-based) indicated that the odds of being adherent to medical treatment were 2.25 times higher if the adolescent participated in an adherence intervention compared with the odds if he/she had not participated in an intervention (95%CI: 1.69-3.03; $P < 0.001$). The standardized relative risk indicated that the risk of nonadherence to medical treatment was 1.50 times higher if the adolescent did not participate in an adherence intervention (95%CI: 1.30-1.74; $P < 0.001$). Furthermore, this set of 21 effect sizes was heterogeneous ($\chi^2 = 44.53$, $P = 1.28 \times 10^{-3}$).

Analysis of other potential moderators of the relationship between interventions and improvements in adherence was conducted. Moderator analysis at the fixed effects level revealed three significant moderators: type of regimen, intervention type, and self-report by adolescents of their own adherence behaviors. For type of regimen, the effectiveness of interventions aimed at improving adherence was moderated by whether or not the intervention was a medication regimen. Specifically, adherence interventions had a greater positive effect on adherence to health behaviors such as diet, exercise, appointment keeping or screening regimens than to medication regimens ($z = -1.77$, $P = 0.039$). In addition, the fixed effects analyses revealed that interventions were less successful at improving adherence in studies where adolescents reported their own adherence behaviors, as compared to having a parent or guardian report adolescent adherence behaviors ($z = -1.91$, $P = 0.038$). Finally, interventions that involved only cognitive modification were less effective in improving adherence than were approaches based on educational intervention, behavioral intervention or a combination of both ($z = -2.14$, $P = 0.16$).

RESULTS

The present meta-analysis provided a comprehensive, quantitative summary of empirical studies evaluating the effectiveness of interventions aimed at improving medical treatment adherence among adolescents. The main hypothesis, that interventions focused on improving treatment adherence are indeed effective, was supported. Although moderator analyses using the random effects model revealed no significant overall moderators of studies' effect sizes, there were several significant results from the fixed effects approach. In studies where adolescent patients reported their own adherence behaviors, interventions demonstrated reduced efficacy, suggesting the possibility of measurement challenges in this work. Interventions improved non-medication regimen adherence (*e.g.*, diet, exercise, appointment keeping) more than medication adherence, perhaps due to the greater challenges of health behavior change, or because medication regimens may be more difficult to target. Finally, educational and behavioral interventions (both combined and in isolation) were more effective than cognitive approaches. These cognitive approaches often attempt to change adolescents' attitudes and beliefs about risk, and they may be less effective because adolescence is a time of heightened vulnerability to risk taking behaviors[19]. Research suggests that because of the temporal gap between puberty and the slower maturation of the cognitive-control system,

changing the context in which the risky behavior occurs maybe more successful than changing the way adolescents think about risk[19].

DISCUSSION

These results provided compelling evidence for the success of efforts to address adolescent nonadherence through interventions designed to assist with the complexities of treating chronic illnesses in this age group. It should be noted that although there were no significant differences between studies that combined adolescents and children samples and those that included adolescents alone, future research should continue to look at these age groups separately, as barriers to adherence can differ between children and adolescents[20]. In doing so, interventions aimed at improving adherence can be targeted to better address the specific needs of each group.

Strengths, limitations and future directions

This systematic quantitative review of adolescent adherence interventions sought to explain what aspects of these interventions were the most successful. With regards to comprehensiveness, several search strategies were utilized and all references were carefully cross-checked. Furthermore, although the mean effect sizes were moderate in size, they may be important in clinical application, and therefore should not be underestimated. Research in the medical field commonly reports small, but highly significant findings with major implications for health. For example, the relationship between consumption of aspirin and the occurrence of heart attacks is in the range of r = 0.03-0.04[21,22]. In other words, there is a 3% to 4% risk difference in prevention of a serious health outcome due to consumption of a simple medication such as aspirin, making the application of this finding very important clinically. In the present research, an unweighted mean r of 0.27 reflects a 27 percent difference in the risk of nonadherence between patients who receive an adherence enhancing intervention and those who do not. Adherence interventions, thus, can have a profound impact on improving adherence among adolescents.

Limitations of this research include the possibility that some empirical studies were missed unintentionally. For example, it is possible that statistically significant findings had greater likelihood of publication, but the large fail safe *N*'s in this review made it unlikely that the current results exhibited the "file drawer bias." Additionally, several studies in the meta-analysis combined the results of both adolescents and children in their reports of the interventions' effectiveness on improving adherence. Therefore, results from the present meta-analysis should be interpreted with caution. Future studies should assess adherence behaviors in children and adolescents separately to allow for exploration of potential age-specific factors that may influence adolescents' adherence and interventions' efficacy. Findings from this study (*i.e.*, the positive effects of multi-faceted, educational/behavioral interventions as compared to cognitive approaches, and of parental assessments of adherence) also underscore the importance of shared decision-making and the role that adolescents, healthcare providers, and parents or caregivers play in the management of disease.

In sum, future research should identify the psychological and behavioral aspects and determinants of adolescent adherence. The present review could assist in the development of specific interventions to enhance adolescent adherence to various medical treatments and types of treatment regimens. Future adherence interventions should also measure and seek to determine both mediators and moderators of adherence interventions' effectiveness in order to fine-tune their development and eliminate the less successful elements. Lastly, future studies should recruit adolescent-only samples, thereby providing meta-analytic opportunities to better understand the challenges (or facilitators) of treatment adherence specific to adolescents.

ACKNOWLEDGEMENTS

The authors would like to thank Mia Montoya for helping to analyze effect sizes and moderators, in addition to, managing the data table for this meta-analysis. Thanks are also due to research assistants: Sang Nguyen, Shiromi Ratnayake, Amberly Navalta, Kim Pham, Vanessa Gutierrez, Shriraj Shah, and Punam Patel for their assistance.

COMMENTS

Background

For adolescents coping with illness, autonomy seeking and other aspects of this developmental transition can lead to confusion, frustration, and conflict with caregivers over prescribed treatment regimens, and consequently, nonadherence. Components of adolescent adherence interventions have been found to vary widely from study to study; thus it is important that the overall efficacy of this wide range of interventions is subjected to systematic, quantitative review.

Research frontiers

Interventions are effective at improving adherence among adolescents. However, previous adolescent studies have treated adolescent and child patients as a homogenous group and reported findings based on the combination of adolescents and children. This limits the interpretability of findings; more quantitative reviews are needed that focus on the unique challenges of adolescent nonadherence.

Innovations and breakthroughs

Previous studies have reported inconsistent results on the efficacy of interventions aimed at improving adherence among adolescents. Some research suggests that multifaceted interventions are more effective than single-approach interventions. The present meta-analysis found that cognitive-based interventions were less effective at improving adherence in adolescents than educational and behavioral approaches.

Applications

The study results suggested that interventions aimed at improving treatment adherence in adolescents are effective, specifically when the type of regimen was behavioral. In addition, cognition-based interventions were less effective. Given the unique challenges adolescents face in coping with illness, future research should consider developmentally appropriate intervention aspects.

Terminology

Patient adherence (or compliance) is the degree to which patients follow treatment directives given by clinicians or other health care providers.

Peer review

Patient nonadherence is a worldwide public health concern, linked to poor clinical outcomes and increased medical costs. The paper was well written. It is interesting and novel topic for present social medical questions.

REFERENCES

1 **DiMatteo MR**, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. *Med Care* 2002; **40**: 794-811 [PMID: 12218770 DOI: 10.1097/01.MLR.0000024612.61915.2D]

2 **DiMatteo MR**. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. *Med Care* 2004; **42**: 200-209 [PMID: 15076819 DOI: 10.1097/01.mlr.0000114908.90348]

3 **Zolnierek KB**, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Med Care* 2009; **47**: 826-834 [PMID: 19584762 DOI: 10.1097/MLR.0b013e31819a5acc]

4 **DiMatteo MR**, Haskard-Zolnierek KB, Martin LR. Improving patient adherence: A three-factor model to guide practice. *Health Psychol Rev* 2011; 1-18 [DOI: 10.1080/17437199.2010.537592]

5 **DiGirolamo AM**, Quittner AL, Ackerman V, Stevens J. Identification and assessment of ongoing stressors in adolescents with a chronic illness: an application of the behavior-analytic model. *J Clin Child Psychol* 1997; **26**: 53-66 [PMID: 9118176]

6 **Chmelik F**, Doughty A. Objective measurements of compliance in asthma treatment. *Ann Allergy* 1994; **73**: 527-532 [PMID: 7998669]

7 **Coutts JA**, Gibson NA, Paton JY. Measuring compliance with inhaled medication in asthma. *Arch Dis Child* 1992; **67**: 332-333 [PMID: 1575560 DOI: 10.1136/adc.67.3.332]

8 **Chappuy H**, Tréluyer JM, Faesch S, Giraud C, Chéron G. Length of the treatment and number of doses per day as major determinants of child adherence to acute treatment. *Acta Paediatr* 2010; **99**: 433-437 [PMID: 19912146 DOI: 10.1111/j.1651]

9 **Guilfoyle SM**, Goebel JW, Pai AL. Efficacy and flexibility impact perceived adherence barriers in pediatric kidney post-transplantation. *Fam Syst Health* 2011; **29**: 44-54 [PMID: 21417523 DOI: 10.1037/a0023024]

10 **Wysocki T**, Harris MA, Buckloh LM, Mertlich D, Lochrie AS, Taylor A, Sadler M, Mauras N, White NH. Effects of behavioral family systems therapy for diabetes on adolescents' family relationships, treatment adherence, and metabolic control. *J Pediatr Psychol* 2006; **31**: 928-938 [PMID: 16401678 DOI: 10.1093/jpepsy/jsj098]

11 **Dean AJ**, Walters J, Hall A. A systematic review of interventions to enhance medication adherence in children and adolescents with chronic illness. *Arch Dis Child* 2010; **95**: 717-723 [PMID: 20522463 DOI: 10.1136/adc.2009.175125]

12 **Wysocki T**, Green L, Huxtable K. Blood glucose monitoring by diabetic adolescents: compliance and metabolic control. *Health Psychol* 1989; **8**: 267-284 [PMID: 2767019 DOI: 10.1037/0278-6133.8.3.267]

13 **van Es SM**, Nagelkerke AF, Colland VT, Scholten RJ, Bouter LM. An intervention programme using the ASE-model aimed at enhancing adherence in adolescents with asthma. *Patient Educ Couns* 2001; **44**: 193-203 [PMID: 11553420 DOI: 10.1016/S0738-3991(00)00195-6]

14 **Magyary D**, Brandt P. A school-based self-management program for youth with chronic health conditions and their parents. *Can J Nurs Res* 1996; **28**: 57-77 [PMID: 9128476]

15 **Rosenthal R**, DiMatteo MR. Meta-analysis: recent developments in quantitative methods for literature reviews. *Annu Rev Psychol* 2001; **52**: 59-82 [PMID: 11148299 DOI: 10.1146/annurev.psych.52.1.59]

16 **DiMatteo MR**, Haskard KB. Further challenges in adherence research: measurements, methodologies, and mental health care. *Med Care* 2006; **44**: 297-299 [PMID: 16565628]

17 **Rosenthal R,** Rosnow RL. Essentials of behavioral research: Methods and data analysis. 3rd ed. McGraw-Hill, 2008: 663-690

18 **Frattaroli J**. Experimental disclosure and its moderators: a meta-analysis. *Psychol Bull* 2006; **132**: 823-865 [PMID: 17073523 DOI: 10.1037/0033-2909.132.6.823]

19 **Steinberg L**. Adolescence. 8th ed. New York: McGraw-Hill, 2007

20 **DiMatteo MR**, Miller TA. Treatment adherence in adolescence. In: O'Donohue L, Benuto L, Tolle, editors. Handbook of adolescent health psychology. New York: Springer, 2013: 373-386

21 **Rosenthal R**, Rosnow RL. Essentials of behavioral research: Methods and data analysis. 2nd ed. New York: McGraw-Hill, 1991

22 **Rosenthal R,** DiMatteo MR. Meta-Analysis: Stevens' handbook of experimental psychology. 3rd ed. In: Pashler H, Wixted J, editors. Hoboken (NJ): John Wiley & Sons Inc, 2002

P- Reviewer: Deng Bo, Shen XC **S- Editor:** Wen LL
L- Editor: A **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.78

World J Meta-Anal 2014 August 26; 2(3): 78-90
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

META-ANALYSIS

Efficacy of therapeutic ultrasound *vs* sham ultrasound on pain and physical function in people with knee osteoarthritis: A meta-analysis of randomized controlled trials

Norma J MacIntyre, Ahmed Negm, Adalberto Loyola-Sánchez, Mohit Bhandari

Norma J MacIntyre, Ahmed Negm, Adalberto Loyola-Sánchez, School of Rehabilitation Science, McMaster University, Hamilton, Ontario L8S 1C7, Canada
Mohit Bhandari, Department of Surgery, McMaster University, Hamilton, Ontario L8L 8E7, Canada
Author contributions: All authors contributed to the design of the research study and interpretation of the data; MacIntyre NJ, Negm A and Loyola-Sánchez A contributed to the acquisition of data as described in the methods; MacIntyre NJ performed the search and data analyses and wrote the paper; Negm A, Loyola-Sánchez A and Bhandari M reviewed the paper critically for important intellectual content and approved the version to be published.
Supported by The Canadian Institutes of Health Research Randomized Controlled Trials Mentorship Program (NJM, MB), No. MTP 108229
Correspondence to: Norma J MacIntyre, PT, PhD, Associate Professor, School of Rehabilitation Science, McMaster University, IAHS 403-1400 Main St W., Hamilton, Ontario L8S 1C7, Canada. macint@mcmaster.ca
Telephone: +1-905-5259140-21166 Fax: +1-905-5240069
Received: November 27, 2013 Revised: May 5, 2014
Accepted: May 28, 2014
Published online: August 26, 2014

Abstract

AIM: To determine the efficacy of therapeutic ultrasound vs sham for improving pain and physical function immediately post-intervention in people with knee osteoarthritis (OA).

METHODS: We hand searched meta-analyses on the topic published in 2010 and updated the search in three electronic databases (MEDLINE, EMBASE, CINAHL) January 1, 2009 to September 5, 2013 to identify relevant studies. The inclusion criteria were human randomized controlled trials published in the English language in which active therapeutic ultrasound was compared to sham ultrasound, data for people with knee OA were reported separately, participants were blinded to treatment allocation and outcomes assessed before and after treatment included pain, self-reported physical function and performance-based physical function. Two reviewers independently screened titles and abstracts retrieved in the search to identify trials suitable for full text review. Data extraction and risk of bias assessment of the identified trials were completed independently by two reviewers. Pooled analyses were conducted using inverse-variance random effects models.

RESULTS: We screened 1013 titles and abstracts. Meta-analysis of pain outcomes from 5 small trials (281 participants/OA knees) showed that, compared to sham ultrasound, therapeutic ultrasound improves pain [standardized mean difference (SMD) (95%CI) = -0.39 (-0.70--0.08); P = 0.01] but not physical function [self-reported in 3 trials (130 participants/OA knees): SMD (95%CI) = -0.21 (-0.55-0.14), P = 0.24; walking performance in 4 trials (130 participants/OA knees): SMD (95%CI) = -0.11 (-0.59-0.37), P = 0.65). For the walking performance outcome, the dispersion of the estimated effects exceeded that expected due to sampling error (χ^2 = 8.37, P = 0.04, I^2 = 64%). Subgroup analyses of three trials that administered high dose ultrasound improved the consistency (I^2 = 28%) but the treatment effect remained insignificant.

CONCLUSION: Meta-analyzed double-blind placebo-controlled randomized trials provide low-strength evidence that therapeutic ultrasound decreases knee OA pain and very low-strength evidence that it does not improve physical function.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Ultrasonic therapy; Physical therapy modalities; Osteoarthritis; Double-blind method; Evidence-

based medicine

Core tip: Controversy exists regarding the efficacy of therapeutic ultrasound in the management of knee osteoarthritis (OA). Lack of participant blinding in effectiveness trials introduces bias known to exaggerate treatment effect estimates particularly for outcomes such as pain and self-reported physical function. We meta-analyzed data from double- and triple-blind trials only and high level evidence shows that therapeutic ultrasound decreases knee OA pain but does not increase physical function immediately following treatment. Due to the methodological quality of the included trials, we conclude that a large well-designed trial is required before this clinical question can be answered definitively.

MacIntyre NJ, Negm A, Loyola-Sánchez A, Bhandari M. Efficacy of therapeutic ultrasound *vs* sham ultrasound on pain and physical function in people with knee osteoarthritis: A meta-analysis of randomized controlled trials. *World J Meta-Anal* 2014; 2(3): 78-90 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/78.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.78

INTRODUCTION

Knee osteoarthritis (OA) is a highly prevalent chronic condition and a leading cause of lower extremity disability in community-dwelling older adults in North America[1]. To date, no treatment exists which modifies the disease and, despite symptom management, pain and functional limitations may progress to the point where total joint replacement is required[2]. As the population ages and the prevalence of obesity increases, the associated economic and personal burden associated with knee OA is expected to rise[2]. Current clinical practice guidelines for managing knee OA recommend a combination of pharmacologic and nonpharmacologic treatment in order to decrease the need for surgical replacement of the damaged joint[3,4].

Therapeutic ultrasound is widely used in clinical settings for various musculoskeletal conditions[5,6]. In a provincial survey of 123 Canadian physical therapists treating clients with knee OA, 81% reported at least some use of ultrasound therapy in their multicomponent management of knee OA[7]. Recent meta-analyses demonstrate that 10 to 24 sessions of continuous or pulsed ultrasound reduces knee OA pain[8-10]. Using two different methods for investigating the effect of dose, two meta-analyses reported that low dose ultrasound achieved using the pulsed mode to administer low intensity (0.375-0.625 W/cm^2) sound waves produced greater pain relief than high dose ultrasound[8,9]. The authors of the most recent meta-analysis back transformed the effect estimate for pain to a visual analogue scale (VAS) score and found a difference of -16.3 [95%CI: -20.9-(-11.7)] mm which was judged to reflect a clinically important change[10]. The effect of therapeutic ultrasound on physical function (self-reported and walking performance) was not significant in the two meta-analyses published in 2010[8,9]. However, the meta-analysis published in 2012 (6 trials, 387 participants) found clinical and statistically significant effects on composite physical function and gait function outcomes[10]. Although these systematic reviews provide evidence for the efficacy of therapeutic ultrasound in the management of knee OA pain and physical disability, all reported that the few, small trials with low methodological quality eligible for inclusion limit the confidence in the effect estimates[8-10].

Meta-analyses of randomized trials provide the highest level of evidence for evaluating the effectiveness of clinical interventions such as therapeutic ultrasound. However the quality of this evidence depends on study design characteristics that yield comparable intervention and control groups. A meta-epidemiologic study of 1973 trials found that intervention effect estimates in trials using subjective outcomes such as self-reported measures were inflated and heterogeneity between trials was increased when double blinding was absent or unclear[11]. It appears that bias due to lack of double blinding exaggerates the effect estimates and heterogeneity for subjectively assessed outcomes more than other study design flaws such as inadequate/unclear random sequence generation and inadequate/unclear allocation concealment[11]. In at least half of the trials included in previous systematic reviews investigating the efficacy of ultrasound therapy on pain and physical function in people with knee OA, no attempt was made to administer sham ultrasound[8-10]. As a result, between-trial heterogeneity and the estimates of the effect of ultrasound on self-reported pain and physical function outcomes published in the highest level evidence available to date are likely to be inflated due to a lack of or unclear blinding of participants with knee OA.

Therefore the objective of this systematic review and meta-analysis was to determine if therapeutic ultrasound *vs* sham ultrasound is effective in improving pain and physical function immediately following the intervention period in people with knee OA blinded to treatment allocation. As a secondary objective, we determined treatment safety based on reported side effects and adverse events.

MATERIALS AND METHODS

Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) recommendations were followed to ensure full and transparent reporting of this review[12]. This systematic review was conducted following a pre-determined protocol which is available from the authors upon request. This protocol was not registered and is not publicly available.

Search strategy

We updated the search for relevant papers completed for previous systematic reviews on this topic (to February 2009)[8,9], by searching OVID MEDLINE, OVID EMBASE, and CINAHL (through EBSCOhost) electronic

Table 1 Example search strategy: Strategy used to search the OVID Medline electronic database (January 1, 2009 to September 5 2013)

1	Exp Osteoarthritis/
2	osteoarthr$.ti,ab,sh.
3	gonarthr$.ti,ab,sh.
4	coxarthr$.ti,ab,sh.
5	arthr$.ti,ab.
6	[(knee$ or hip$ or joint$) adj3 (pain$ or ach$ or discomfort$)].ti,ab.
7	[(knee$ or hip$ or joint$) adj3 stiff$].ti,ab.
8	Exp Ultrasonic Therapy/
9	Exp Ultrasonography/
10	us.fs.
11	(ultrasound$ or ultrasonic$).tw.
12	short wave therapy.tw.
13	ultrasonograph$.tw.
14	randomized controlled trial.pt.
15	controlled clinical trial.pt.
16	randomized controlled trial.sh.
17	random allocation.sh.
18	double blind method.sh.
19	single blind method.sh.
20	clinical trial.pt.
21	Exp Clinical Trial/
22	(clin$ adj25 trial$).ti,ab.
23	[(singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)].ti,ab.
24	placebos.sh.
25	placebo$.ti,ab.
26	random$.ti,ab.
27	research design.sh.
28	comparative study.sh.
29	exp evaluation studies/
30	follow up studies.sh.
31	prospective studies.sh.
32	(control$ or prospectiv$ or volunteer$).ti,ab.
33	1 or 2 or 3 or 4 or 5 or 6 or 7
34	8 or 9 or 10 or 11 or 12 or 13
35	14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
36	33 and 34 and 35
37	animal/
38	animal/ and human/
39	37 not 38
40	36 not 39
41	Limit 40 to yr = "2009-Current"

databases from January 1, 2009 to September 5, 2013. The search strategy combined Medical Subject Headings (MeSH) terms and text words related to design (randomized controlled trial), condition (knee osteoarthritis), and intervention (therapeutic ultrasound). We reproduced the specific search strategy used for each database published in the previous systematic review focusing on clinical outcomes[9]. (Although previously published, the replicated strategy used to search the Medline electronic database is shown in Table 1 to comply with the PRISMA checklist.) No language limit was placed on the search in order to increase sensitivity. Duplicates were removed after all databases were searched. We searched for published trials only and attempted to contact primary authors to request additional information if necessary.

Study eligibility criteria and selection

Two reviewers (NM, AL) independently screened the titles and abstracts for all citations identified by the search and retrieved all parallel-group randomized sham-controlled trials assessing the effect of ultrasound on pain or physical function in people with knee OA published in English. Studies examining other joints were only included if the data for people with knee OA were reported separately. Trials that included other interventions in addition to active ultrasound were included as long as the sham ultrasound group received the same combination of interventions. Lack of blinding of study participants was an exclusion criterion because we considered this to be a major source of biased results regarding ultrasound efficacy on self-reported pain and physical function outcomes[11]. Abstracts published as conference proceedings and not as full trials were excluded due to insufficient reporting of data for extraction. Cohen's unweighted kappa (κ) was used to measure agreement between reviewers. Disagreement was solved by consensus including a third reviewer (AN). A final list of eligible studies was prepared for full text review after title and abstract screening.

Data extraction and management

The data collection form used for our previous systematic review[8] was modified for this updated review to reflect the narrowed research question focusing on patient-centred outcomes of pain and physical function. The form was independently pilot-tested on two randomly-selected studies by two reviewers (NM, AN) to ensure consistency in coding instructions. A double extraction method was followed using the refined data collection form. A physical therapist with expertise in research methodology and OA rehabilitation (NM) and an orthopaedic surgeon with expertise in OA (AN) reviewed the papers and extracted the data in a standardized manner. As recommended in the Cochrane Handbook[13], the reviewers were not blinded to any aspect of the trials during data extraction. Any disagreement was resolved through consensus.

The information extracted from each study included: (1) characteristics of the study participants (age, gender, diagnostic criteria, joint involvement, and knee OA severity); (2) characteristics of the therapeutic ultrasound intervention (device, frequency, mode, intensity, effective radiating area, surface area treated, application protocol, number and length of treatment sessions); (3) description of co-interventions; (4) description of pain and physical function outcomes and corresponding data; (5) reported adverse events and reasons for loss to follow up; and (6) general characteristics of the studies (location, clinical setting and funding source). The ultrasound dose was calculated as for our previous review[8] using the following formula:

Energy (J/cm^2) = [(average temporal intensity) (time) (effective radiating area)]/treated surface area

Group means for outcomes at baseline, post treatment, change from baseline and standard deviations (SDs), or the information from which SDs could be derived, were extracted. For trials that included two groups receiving active ultrasound (continuous and pulsed

mode), the formulae for combining means and SDs for two groups published in the Cochrane Handbook were used for estimating the effect of active ultrasound on the clinical outcomes of interest. (See Table 7.7a in the Cochrane Handbook 2005[13]). When a trial presented outcomes at time points other than pre- and post-intervention, we extracted the data for all time points, however, the mean values at the end of the treatment period were used in the meta-analysis. Heterogeneity in trial outcomes was minimized by pooling the data for the same outcome measure where possible, or pooling those that were most similar in terms of constructs assessed (*e.g.*, pain with movement measured by VAS or numeric rating scale, NRS).

Risk of bias and quality assessment

Two reviewers (NM, AN) independently assessed risk of bias for each study and the level of agreement was determined using linear weighted kappa (κ). Any disagreement regarding risk of bias was resolved by consensus. The methodological domains recommended by The Cochrane Collaboration were assessed: randomization, treatment allocation concealment, blinding of participants, care givers and outcome assessors, completeness of outcome data, selective outcome reporting and other potential threats to validity[14]. Table 2 illustrates the criteria for assessing risk of bias in these domains. An overall risk of bias was considered for each study and across all studies based on the criteria outlined in the Cochrane Handbook[14]. We assigned studies as having high risk of bias if at least one criterion was not met. The quality of the body of evidence for each outcome was determined considering within-trial risk of bias, directness of evidence, heterogeneity, risk of publication bias and precision of effect estimates as recommended by the Grading of Recommendations, Assessment Development and Evaluation (GRADE) working group[15]. We defined treatment effects as precise when pooled estimates had reasonably narrow 95%CIs and the pooled sample size was greater than 400[15].

Statistical analysis

We focused on patient-centred outcomes of pain and physical function. Follow up duration was to the end of the ultrasound therapy treatment. Review Manager Version 5 was used for data analyses. Effect size for each outcome (by study and overall) was estimated using the standardized mean difference (SMD) calculated as the raw mean difference divided by the pooled variance for each outcome to allow comparison of estimates of effect. Where the same outcome measure was used, the mean difference was also calculated. We used inverse-variance random-effects models to pool results to account for the inevitable variation in patient populations, concomitant treatments, and specific components of the physical therapy intervention as recommended when the number of studies is small and the reasons for heterogeneity across the studies cannot be reliably evaluated[16]. We evaluated the consistency of findings by comparing the direction and strength of effect along with the degree of statistical heterogeneity (based on the χ^2 and I^2 statistical tests) in effects across studies. χ^2 values with $P \geqslant 0.1$ and $I^2 < 60\%$ were considered to be acceptable homogeneity for pooling the data[16]. We used subgroup analyses and planned to conduct meta-regression to evaluate the effects of *a priori*-defined clinical characteristics (OA severity) and study characteristics (ultrasound dose) on pain and physical function outcomes. These subgroups were of interest because our previous meta-analysis demonstrated that the effect estimates for pain reduction were increased for those receiving low dose therapeutic ultrasound and heterogeneity observed for physical function and walking performance was reduced in the subgroup of trials including participants with mild knee OA severity[8]. We hypothesize that therapeutic ultrasound will be more effective in people with less severe knee OA in whom tissue damage is less advanced[17]. Confidence intervals (CI) at the 95% level were calculated for pooled estimates for each outcome and the *Z* test was used for determining the treatment effect. Statistical significance was considered at $P \leqslant 0.05$.

For each study reporting each of the three clinical outcomes of interest, we evaluated funnel plots for asymmetry in the standard error of the SMD as a function of the SMD. We did not use statistical tests for publication bias given that we did not include unpublished data sources in our search strategy.

Role of the funding source

The study was funded, in part, by CIHR (NM, MB). CIHR had no role in the development of the question or the review protocol, literature search, data extraction and analysis, interpretation of the results, or preparation of the manuscript.

RESULTS

Figure 1 is a flowchart of the results of search strategy. The initial database and hand searches retrieved 1226 citations (n = 213 duplicates). After title and abstract screening, 10 randomized controlled trials were identified as eligible and retrieved for full text review. Of these, six were eligible for inclusion in the systematic review and five provided published data that permitted pooling. Reasons for exclusion are shown in Figure 1. The agreement between the reviewers in identifying the studies eligible for full text review and for inclusion in this review was κ = 1.0.

Study characteristics

Characteristics of the six parallel-group, randomized, placebo-controlled trials eligible for inclusion in the systematic review are summarized in Table 1[18-23]. The RCTs were published between 1992 and 2012. Five trials included two arms (active ultrasound and sham ultrasound)[18,19,21-23], whereas the trial by Tascioglu *et al*[20]

Table 2 Criteria used to classify risk of bias in trials according to components of methodological quality

Random sequence generation	
Low risk	Referring to a random number table; Using a computer random number generator; Coin tossing; Shuffling cards or envelopes; Throwing dice; Drawing of lots
High risk	Sequence generated by odd or even date of birth; Sequence generated by some rule based on date (or day) of admission; Sequence generated by some rule based on hospital or clinic record number; Allocation by judgment of the clinician; Allocation by preference of the participant; Allocation based on the results of a laboratory test or a series of tests; Allocation by availability of the intervention
Unclear	Insufficient information
Allocation concealment	
Low risk	Central allocation (including telephone, web-based); Sequentially numbered, opaque, sealed envelopes; An equivalent method was used to conceal allocation
High risk	Using an open random allocation schedule (*e.g.*, a list of random numbers); Assignment envelopes were used without appropriate safeguards (*e.g.*, if envelopes were unsealed or nonopaque or not sequentially numbered); Alternation or rotation; Date of birth; Case record number; Any other explicitly unconcealed procedure
Unclear	Insufficient information
Blinding of each of: participant, care provider, outcome assessor	
Low risk	Unlikely that the blinding could have been broken
High risk	Likely that the blinding could have been broken
Unclear	Insufficient information
Completeness of outcome data collection	
Low risk	No missing outcome data; Reasons for missing outcome data unlikely to be related to true outcome; Missing outcome data balanced in numbers across all groups, with similar reasons for missing data across groups; Missing data have been imputed using appropriate methods; For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size
High risk	Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; "As-treated" analysis done with substantial departure of the intervention received from that assigned at randomization. Potentially inappropriate application of simple imputation. For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size
Unclear	Insufficient information
Completeness of outcome reporting	
Low risk	The study protocol is available and all of the study's pre-specified outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified
High risk	Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (*e.g.*, subscales) that were not pre-specified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study
Unclear	Insufficient information
Other potential sources	
Low risk	The study appears to be free of other sources of bias
High risk	There is at least one important risk of bias. For example, the study had a potential source of bias related to the specific study design used or has been claimed to be fraudulent, or has some other problem
Unclear	Insufficient information

included three arms in which one arm received active pulsed ultrasound (low dose), one arm received active continuous ultrasound (high dose), and one arm received sham ultrasound. Three trials[19,20,23] were conducted in Turkey and one trial was conducted in each of the United States of America[18], Canada[22], and China[21]. The duration of the intervention varied from 5 d to 8 wk and the dosage varied from 26 J/cm^2 to 196.3 J/cm^2 (Table 3). The control groups in all the studies received sham ultrasound and the participants were blinded to the allocated treatment. Only one trial confirmed that treatment providers were blinded to the intervention allocation[22]. In the trial conducted by Yang *et al*[21] it was unclear how the sham ultrasound was delivered. Furthermore, the trial by Yang *et al*[21] reported a "curative effect score" for the 87 participants (100 knees) randomized to active or sham ultrasound groups rather than reporting group means (SD) for pain and physical function outcomes and attempts to contact the authors to secure the raw data were unsuccessful. This calculated efficacy index was significantly improved for pain ($P < 0.001$) and physical function ($P < 0.001$)[21]. Nevertheless, the data from this trial[21] could not be included in the meta-analysis. The five trials eligible for meta-analysis reported data for a total of 281 participants and OA knees[18-20,22,23]. Table 3 shows that the number of participants providing data for analyses was lower than the number of participants randomized (varying from 2 to 8 participants per trial) in all trials meta-analyzed. Reasons given for loss to follow up included protocol violation (used analgesics)[19,20], illness[18], dissatisfaction with treatment[18], lack of time to attend/lack of regular attendance at sessions[20,23], transportation problems[18], or incomplete baseline assessment before withdrawing from the study[22]. The average age across the studies was approximately 61 years with the majority of participants being female (Table 3). Four

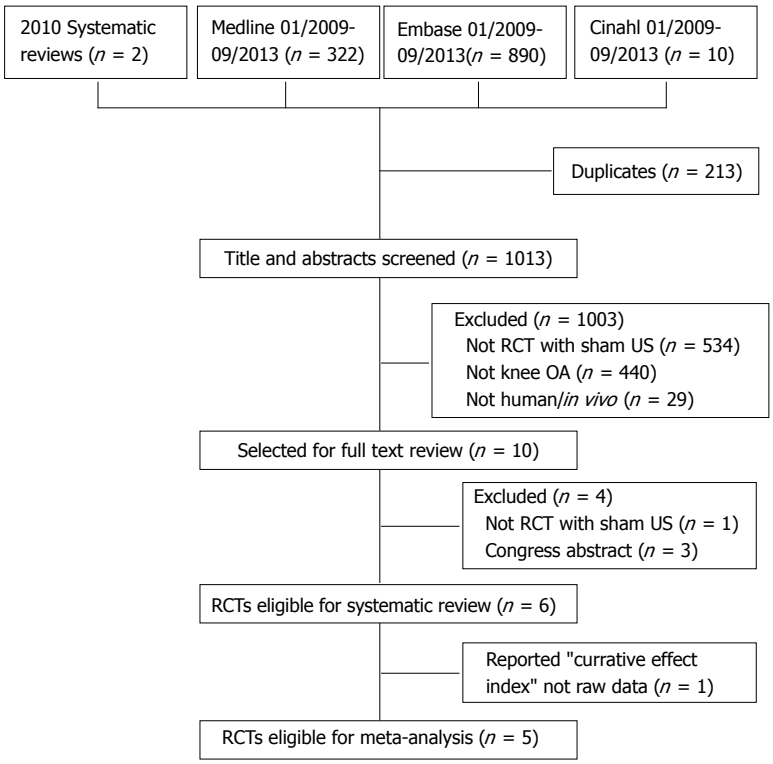


Figure 1 Flow diagram for identification of eligible trials evaluating the effect of ultrasound vs sham on pain and physical function.

trials[19,20,22,23] confirmed that participants had mild to moderate radiographic knee OA whereas one included trial[18] reported results for 69 participants/knees with unknown radiographic severity. (The latter trial included eight participants with a history of total knee replacement and all 69 had restricted knee range of motion for more than 6 mo[18].) Three trials reported that the majority of participants had bilateral knee OA[18,22,23] with one of these trials including only participants with bilateral knee OA[23]. Two of the five meta-analyzed trials applied active and sham ultrasound in combination with concurrent treatments (including isometric quadriceps exercises) that were identical for both study groups[18,23]. In contrast, two trials[19,20] did not allow any concurrent treatments including the use of analgesics during the intervention period. Three trials[18,20,22] administered low dose ultrasound (< 150 J/cm^2) and three trials[19,20,23] administered high dose ultrasound ($\geqslant 150$ J/cm^2). Three trials[19,22,23] reported scores on more than 1 pain outcome measure and two of these trials[22,23] reported scores on more than 1 self-reported physical function outcome measure. Table 3 specifies the pain and self-reported physical function outcomes selected for inclusion in the meta-analysis in order to reduce heterogeneity arising from differences in outcomes used across studies. Three trials[19,22,23] administered the WOMAC LK 3.1 questionnaire and reported scores on the physical function subscale. One trial[20], administering the WOMAC LK3.1 to assess the effect of ultrasound therapy on disability, reported only the total score which was not included in the meta-analysis for the self-reported physical function outcome. Another trial[21] administered the Lequesne Severity Index to assess the effect of ultrasound therapy on self-reported physical function but did not report the group means in a format that permitted pooling as mentioned previously. Walking performance was measured as walking speed (time to walk a specified distance) in four trials[18-20,23], and as distance walked (metres walked in six minutes) in one trial[22]. One trial[18] reported means for walking velocity (metres/s) pre and post treatment for the total sample only. We were unsuccessful in obtaining group means from the primary author and therefore these walking performance data could not be included in the meta-analysis. The units of measurement for the meta-analysed walking performance values were time (in seconds), where lower values indicate better scores, and distance (in metres), where higher values indicate better scores. Therefore, the group means for distance walked in six minutes were entered into the meta-analysis as negative values to adjust for the directional difference in interpreting better scores for this outcome.

Trial risk of bias and quality of the evidence

The agreement between reviewers in determining risk of bias was $\kappa = 0.81$. Table 4 summarizes the methodological quality assessment of the six trials retrieved for full text review. Two trials[22,23] reported adequate random sequence generation and allocation concealment. Only one trial[21] did not report the blinding of participants adequately. Only one trial[22] confirmed that care providers were blinded (triple blind) and all five of the meta-analyzed trials confirmed that assessors were blinded to treatment allocation (double blind)[18-20,22,23]. In two included trials[19,20], it was unclear if complete data were collected. In one trial[18] included in the meta-analysis, data reporting was incomplete. Only one trial[21] (not included

Table 3 Characteristics of the studies selected for full text review

	Falconer *et al*[18]	Ozgönenel *et al*[19]	Tascioglu *et al*[20]	Yang *et al*[21]	Loyola-Sánchez *et al*[22]	Ulus *et al*[23]
Trial registry number	Not available	Not available	Not available	Not available	NTC00931749	Not available
Trial duration	4-6 wk (12 sessions 2-3 x/wk)	2 wk (10 sessions 5 x/wk)	2 wk (10 sessions 5x/wk)	5 d (5 sessions); + 1 mo follow up	8 wk (24 sessions 3 x/wk)	3 wk (15 sessions 5 x/wk)
Sample size	Randomized: 74; analyzed: 69 (35 CG)	Randomized: 67; analyzed: 65 (31 CG)	Randomized: 90; analyzed: 82 (27 CG)	Randomized and analyzed: 87 (100 knees; 50 CG)	Randomized: 27; analyzed: 25 (13 CG)	Randomized: 42; analyzed: 40 (20 CG)
Sample characteristics (Mean SD/*n* reported)	Age approximately 67.5 (11) yr; 50 F; All restricted knee ROM ⩾ 6 mo; 8 knee joint replacement; 51 bilateral OA	Age approximately 55 (7.5) yr; 54 F; Newly diagnosed; 31 mild OA, 36 moderate OA	Age approximately 60 (3) yr; 56 F; Disease duration 6.5 yr; 48 mild OA, 34 moderate OA	Age 58.3 yr; 72 F; Disease duration 2.8 yr	Age approximately 61.8 (10) yr; 21 F; 8 had mild OA, 19 had moderate OA; 24 bilateral OA	Age approximately 60.5 (9.5) yr; 34 F; disease duration 8.9(8.7) yr; 17 mild OA, 23 moderate OA; all bilateral OA
Ultrasound device	Chattanooga Intellect 200	Peterson .250	Sonopuls 434	NERCUM	Chattanooga Intellect Mobile	Sonopuls 434
Application protocol	12 min; 1 MHz; intensity: 1.7 W/cm^2, continuous mode (*n* = 34)	5 min; 1 MHz; intensity 1 W/cm^2; continuous mode (*n* = 34)	5 min; 1 MHz; intensity 1 W/cm^2, continuous mode (*n* = 27); pulsed (duty cycle 20%, *n* = 28)	No details: 15min treatment model then 20min rehabilitation model (*n* = 50 knees)	9.5 min; 1MHz; intensity 1 W/cm2; pulsed mode (duty cycle 20%, *n* = 12)	10 min; 1 MHz; intensity 1 W/cm^2; continuous mode (*n* = 20)
Sham Application	Start button not pushed	Applicator disconnected from back of device	No output delivered	No output delivered	Ceramic crystal removed from soundhead	No output delivered and applicator disconnected from back of device
Application site	Knee flexed or extended per most restricted motion; treated surface area 100 cm^2	Patellofemoral and tibiofemoral borders; treated surface area 25 cm^2	Antero-medial and lateral parts of extended knee; treated surface area 60 cm^2	Knee extended; 4 soundheads fixed on joint line; treated surface area not reported	Knee flexed to 90°; Soundhead fixed at antero-medial joint line. treated surface area 5 cm^2	Antero-medial and lateral parts of extended knee; treated surface area 60 cm^2
Dosage	26 J/cm^2	150.7 J/cm^2	196.3 J/cm^2 39.3 J/cm^2	Unable to calculate	114 J/cm^2	196.3 J/cm^2
Concurrent treatment	Stretching, joint mobilizations, exercises (ROM, bridging, isometric quads, home program)	none	none	none	None reported; use of analgesics not reported	Hotpacks (20 min); IFC (10 min); exercises (isometric quads); analgesics except during physio
Outcomes included in meta-analysis	Pain – 10 cm VAS	Pain on movement in past wk – 10 cm VAS WOMAC LK 3.1 Physical Function subscale Walking speed [time (s) to walk 50 m]	Pain on movement in past wk – 10 cm VAS Walking speed [time (s) to walk 20 m]	none	Pain following walking test – 11 point NRS WOMAC LK 3.1 Physical Function subscale Distance (m) walked in 6 min	Pain on activity – 10 cm VAS WOMAC LK 3.1 Physical Function subscale Walking speed (time (s) to walk 50 m)
Funding source	Non-profit	Not reported	Not reported	NERCUM, institutional	Government	Unfunded
Comments	Trial author confirmed mode was continuous; pain data extracted from graphs; request for pain and walking data was unsuccessful		Treated surface area not reported; estimated to be 3x the sound head size (based on parts of knee treated)	Attempts to contact authors for pain and physical function data that could be pooled were unsuccessful	1 of the 2 reviewers co-authored the trial; outcomes pooled in this review were secondary outcomes in trial	Treated surface area not reported; estimated to be 3x the sound head size (based on parts of knee treated)

OA: Knee osteoarthritis; CG: Control group; F: Female; NERCUM: National Engineering Research Centre of Ultrasound Medicine; ROM: Range of motion; IFC: Interferential current electrotherapy; VAS: Horizontal visual analogue scale (10 = worst pain); NRS: Numeric rating scale (10 = worst pain); WOMAC LK 3.1: Western Ontario and McMaster Universities Osteoarthritis Index (Likert scale version) self-report questionnaire physical function subscale (68 = worst limitation).

Table 4 Methodological quality assessment of the randomized sham-controlled trials

Ref.	Random Sequence Generation	Allocation Concealment	Blinding of Participant	Blinding of Care Provider	Blinding of Assessor	Data Collection Complete	Complete Outcome Reporting	Free of Other Potential Bias	Risk of Bias[1]
Falconer *et al*[18]	Unclear	Unclear	Yes	No	Yes	Yes	No	Yes	High
Ozgönenel *et al*[19]	Unclear	Unclear	Yes	No	Yes	Unclear	Yes	Yes	High
Tascioglu *et al*[20]	Unclear	Unclear	Yes	No	Yes	Unclear	Yes	Yes	High
Yang *et al*[21]	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	No	No	High
Loyola-Sánchez *et al*[22]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low
Ulus *et al*[23]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	High

[1]Overall risk of bias for each study assessed as Low if has a Yes answer for ALL the items or as High if has NO or Unclear answer for one or more items.

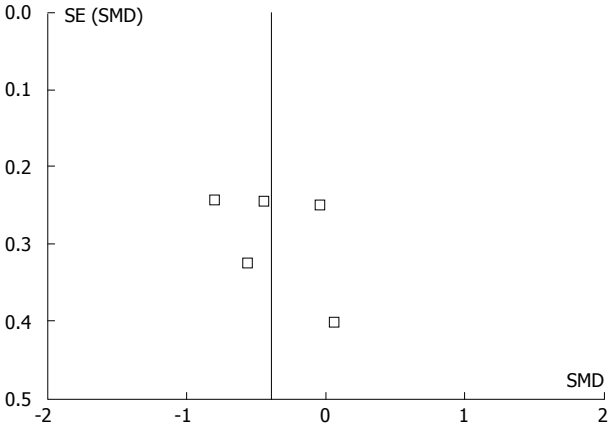


Figure 2 Funnel plot illustrating the statistical precision plotted as a function of the standardized mean difference for the effect of therapeutic ultrasound on patient-reported knee pain. Statistical inferences are not possible when fewer than 10 trials are available; therefore the pseudo 95% confidence limits around the summary treatment effect are not shown. SMD: Standardized mean difference.

in the meta-analysis) was deemed to have other potential sources of bias due to randomization of more than one knee from the same participant. Based on our criteria for judging the overall methodological quality assessment for each trial, risk of bias was low in one trial and high in five trials (Table 4). Therefore, risk of bias across the studies is high. Slight asymmetry is noted in the funnel plots for each of the outcomes as illustrated in Figure 2 for the pain outcome.

Effects of the intervention

In the five trials included in the analysis assessing efficacy of therapeutic ultrasound on knee OA pain, a total of 281 participants were randomised: 155 to an active ultrasound group and 126 to a sham ultrasound group[18-20,22,23]. Figure 3A illustrates that therapeutic ultrasound (pulsed or continuous) was effective in reducing pain compared to sham ultrasound [SMD = -0.39 (-0.70, -0.08); $P = 0.01$]. The proportion of the dispersion in effect estimates across studies fell within that expected due to sampling error ($I^2 = 36\%$) and this low heterogeneity is unlikely to represent a true difference in effects in the studies ($\chi^2 = 6.26$, $P = 0.18$). Subgroup analyses based on dosage planned *a priori* demonstrated that low dose therapeutic ultrasound resulted in a significant decrease in pain [3 trials; SMD = -0.45 (-0.85, -0.05); $P = 0.03$; Figure 3B] and reduced the heterogeneity further ($\chi^2 = 2.82$, $P = 0.24$; $I^2 = 29\%$). In contrast, pooling the three studies administering high dose ultrasound *vs* sham ultrasound yielded a statistically insignificant decrease in pain and, although still acceptable, the heterogeneity increased [3 trials, SMD = -0.46 (-0.94, 0.03); $P = 0.06$; $\chi^2 = 4.49$, $P = 0.11$; $I^2 = 55\%$]. We were unable to perform subgroup analyses based on OA severity because all five either included participants with mild to moderate radiographic knee OA or did not report radiographic OA severity. Meta-regression analyses were not performed due to the small number of trials available for subgroup comparisons.

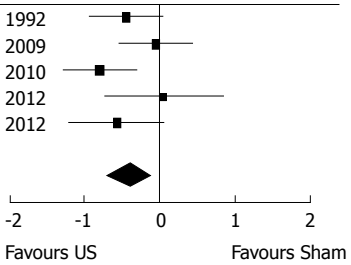
Three trials (130 participants and OA knees) assessed the effect of therapeutic ultrasound on self-reported physical function. Figure 3C illustrates that the point estimate for self-reported physical function in people with knee OA favours therapeutic ultrasound, however, the 95%CI crosses the point of no difference (SMD= -0.21, 95%CI: -0.55-0.14, $P = 0.24$). Homogeneity for pooling the data across studies was acceptable ($\chi^2 = 1.92$, $P = 0.24$, $I^2 = 0\%$). Due to the small numbers of studies that reported self-reported physical function outcomes, we could not perform meta-regression analyses.

Four trials assessed the effect of therapeutic ultrasound on physical function with respect to walking performance in people with knee OA. Figure 3D illustrates that walking performance in the group who received active ultrasound was not significantly different from that in the group who received sham ultrasound (SMD= -0.11, 95%CI: -0.59-0.37, $P = 0.65$). In this analyses, the dispersion of the estimated effects exceeded that expected due to sampling error ($\chi^2 = 8.37$, $P = 0.04$, $I^2 = 64\%$). Figure 3E illustrates that only the subgroup analysis of three trials administering high dose therapeutic ultrasound *vs* sham reduced the heterogeneity ($I^2 = 28\%$). The sample size was very small and the small effect estimate was not statistically significant (SMD = 0.04, 95%CI: -0.33-0.41, $P = 0.82$). Of the four trials available for meta-analysis, all included participants with mild to moderate knee OA. Meta-regression analyses planned *a priori* were not performed due to the small number of trials available for subgroup comparisons.

A

Study or Subgroup	US Therapy Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean difference IV, Random, 95%CI	Year	Std. Mean difference IV, Random, 95%CI
Falconer	2.6	2.91	34	3.9	2.95	35	23.9%	-0.44 [-0.92, 0.44]	1992	
Ozgonenel	3.9	2	34	4	2.6	31	23.4%	-0.04 [-0.53, 0.44]	2009	
Tascioglu	5.24	1.79	55	6.67	1.78	27	23.9%	-0.79 [-1.27, -0.32]	2010	
Loyola Sanchez	1.58	2.02	12	1.46	2.02	13	12.1%	0.06 [-0.73, 0.84]	2012	
Ulus	3.3	1.83	20	4.45	2.16	20	16.7%	-0.56 [-1.20, 0.07]	2012	
Total (95%CI)			155			126	100.0%	-0.39[-0.70, -0.08]		

Hetergeneity: Tau^2 = 0.04; χ^2 = 6.26, df = 4 (P = 0.18); I^2 = 36%
Test for overall effect: Z = 2.49 (P = 0.01)



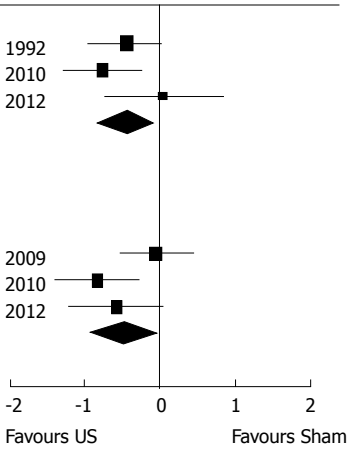
B

Study or Subgroup	US Therapy Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean difference IV, Random, 95%CI	Year	Std. Mean difference IV, Random, 95%CI
1.2.1 US dose < 150 J/cm^2										
Falconer	2.6	2.91	34	3.9	2.95	35	43.0%	-0.44 [-0.92, 0.04]	1992	
Tascioglu	5.25	1.9	28	6.67	1.78	27	36.0%	-0.76 [-1.31, -0.21]	2010	
Loyola Sanchez	1.58	2.02	12	1.46	2.02	13	21.0%	0.06 [-0.73, 0.84]	2012	
Subtotal (95%CI)			74			75	100.0%	-0.45[-0.85, -0.05]		

Hetergeneity: Tau^2 = 0.04; χ^2 = 2.82, df = 2 (P = 0.24); I^2 = 29%
Test for overall effect: Z = 2.21 (P = 0.03)

1.2.2 US dose ⩾ 150 J/cm^2										
Ozgonenel	3.9	2	34	4	2.6	31	37.2%	-0.04 [-0.53, 0.44]	2009	
Tascioglu	5.22	1.7	27	6.67	1.78	27	33.3%	-0.82 [-1.38, -0.26]	2010	
Ulus	3.3	1.83	20	4.45	2.16	20	29.5%	0.56 [-1.20, 0.07]	2012	
Subtotal (95%CI)			81			78	100.0%	-0.46 [-0.94, 0.03]		

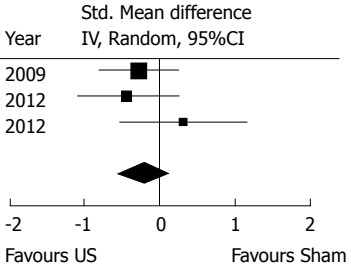
Hetergeneity: Tau^2 = 0.10; χ^2 = 4.49, df = 2 (P = 0.11); I^2 = 55%
Test for overall effect: Z = 1.85 (P = 0.06)



C

Study or Subgroup	US Therapy Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean difference IV, Random, 95%CI	Year	Std. Mean difference IV, Random, 95%CI
Ozgonenel	23.6	11.6	34	27.1	14	31	50.2%	-0.27 [-0.76, 0.22]	2009	
Tascioglu	28.65	10.62	20	32.45	7.33	20	30.5%	-0.41 [-1.04, 0.22]	2012	
Loyola Sanchez	23.92	11.3	12	20.38	13	13	19.3%	0.28 [-0.51, 1.07]	2012	
Total (95%CI)			66			64	100.0%	-0.21[-0.55, 0.14]		

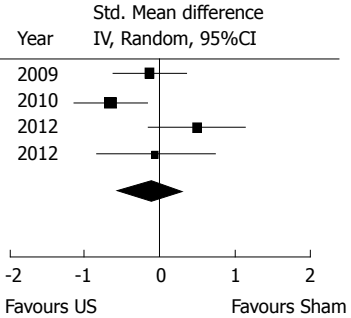
Hetergeneity: Tau^2 = 0.00; χ^2 = 1.92, df = 2 (P = 0.38); I^2 = 0%
Test for overall effect: Z = 1.17 (P = 0.24)



D

Study or Subgroup	US Therapy Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean difference IV, Random, 95%CI	Year	Std. Mean difference IV, Random, 95%CI
Ozgonenel	35.5	6.7	34	36.4	7.6	31	28.3%	-0.12 [-0.61, 0.36]	2009	
Tascioglu	21.4	2.87	55	23.19	2.54	27	28.8%	-0.64 [-1.11, -0.17]	2010	
Ulus	68.05	44.53	20	51.25	15.74	20	23.6%	0.49 [-0.14, 1.12]	2012	
Loyola Sanchez	-414.55	68.53	12	-411.85	89.8	12	19.3%	-0.03 [-0.82, 0.75]	2012	
Total (95%CI)			121			91	100.0%	-0.11 [-0.59, 0.37]		

Hetergeneity: Tau^2 = 0.15; χ^2 = 8.24, df = 3 (P = 0.04); I^2 = 64%
Test for overall effect: Z = 0.45 (P = 0.65)

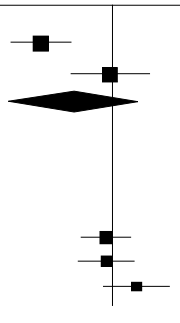


E

Study or Subgroup	US Therapy Mean	SD	Total	Sham Mean	SD	Total	Weight	Std. Mean difference IV, Random, 95%CI	Year	Std. Mean difference IV, Random, 95%CI
1.5.1 US dose < 150 J/cm^2										
Tascioglu	20	1.94	28	23.19	2.54	27	51.8%	-1.39 [-1.99, -0.80]	2010	
Loyola Sanchez	-414.55	68.53	12	-411.85	89.8	13	48.2%	-0.03 [-0.82, 0.75]	2012	
Subtotal (95%CI)			40			40	100.0%	-0.74 [-2.07, 0.60]		

Hetergeneity: Tau^2 = 0.80; χ^2 = 7.36, df = 1 (P = 0.007); I^2 = 86%
Test for overall effect: Z = 1.09 (P = 0.28)

1.5.2 US dose ⩾ 150 J/cm^2										
Ozgonenel	35.5	6.7	34	36.4	7.6	31	38.9%	-0.12 [-0.61, 0.36]	2009	
Tascioglu	22.85	2.99	27	23.19	2.54	27	34.3%	-0.12 [-0.65, 0.41]	2010	
Ulus	68.05	44.53	20	51.25	15.74	20	26.8%	0.49 [-0.14, 1.12]	2012	



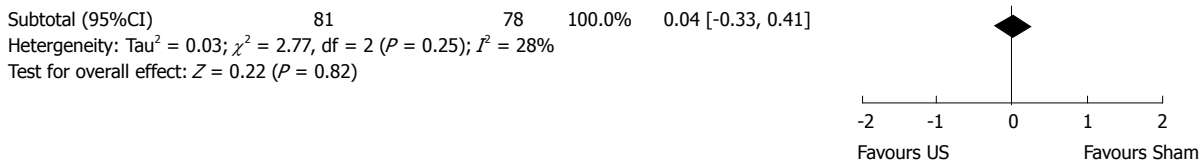


Figure 3 Forest plot. A: For meta-analysis of the effect of ultrasound (US) compared to sham on pain; B: For trials reporting the pain outcome subgrouped according to US dose; C: For meta-analysis of the effect of US compared to sham treatment on self-reported physical function; D: For meta-analysis of the effect of US compared to sham treatment on walking performance; E: For trials reporting a walking performance outcome subgrouped according to US dose.

Adverse events

Two trials reported that no major complaints were reported[18,23]. Two trials reported no side effects although in one trial[19] two participants in the sham ultrasound group had to use analgesics and were dropped from the study and in another trial[20] six out of 90 participants (one in the active pulsed ultrasound group, two in the active continuous ultrasound group and three in the sham ultrasound group) had to use nonsteroidal anti-inflammatory drugs or analgeiscs because of intolerable pain and were withdrawn from the study. In one trial[21], adverse effects included dizziness (n = 3), mental stress, palpitation or fatigue however it was not stated if these participants received active or sham ultrasound. In the triple blind trial[22], two participants reported a "stinging" sensation during active ultrasound treatment and two participants receiving sham ultrasound reported a similar 'stinging' sensation during treatment.

Strength of the body of evidence

The strength of the body of evidence for pain reduction and improvement in self-reported physical function and walking performance is summarized in Table 5. The finding that therapeutic ultrasound reduces pain is based on low-strength evidence (Table 5). For the self-reported physical function outcome, the mean difference in score is shown in Table 5 since all three of the trials provided data for the WOMAC LK3.1 physical function subscale that permitted pooling. The finding that therapeutic ultrasound does not improve self-reported physical function is based on very low strength evidence (Table 5). We report the findings for walking performance obtained when we pooled all four trials that assessed this outcome although our criteria for pooling data were not met (P = 0.04 for χ^2 and I^2 = 64%). Acceptable homogeneity was achieved when we meta-analyzed the 3 trials (159 participants and knees) applying high dose ultrasound (χ^2 = 2.77, P = 0.25, I^2 = 28%). We did not report the results of this subgroup analysis in the Summary of Findings Table because we felt the very small sample size contributed to the imprecision of the effect estimate for walking performance to a similar degree as the inconsistency thus the strength of the body of evidence could not be upgraded.

DISCUSSION

This systematic review provides a meta-analysis of the efficacy of therapeutic ultrasound *vs* sham ultrasound for decreasing pain and improving physical function in people with knee OA who were blinded to treatment allocation. The main finding is that therapeutic ultrasound treatment decreases pain but does not improve physical function significantly in this patient population. Our results confirm those of recent meta-analyses[8-10]. Confidence in the treatment effect estimates is increased by including only those studies in which participants were blinded to the treatment they received[11]. Whereas we did remove an important source of subjective assessment bias, the strength of the evidence for pain and physical function outcomes is low and very low, respectively. Few trial participants from either the active or sham ultrasound group reported adverse events or side effects. These findings demonstrate that therapeutic ultrasound is a safe and effective treatment for pain in people with knee OA and further research is needed to determine if physical function improves.

In our systematic review, the treatment effect estimate is lower compared to those reported previously [SMD = -0.39 (-0.70, -0.08); P = 0.01] demonstrating the importance of study design characteristics of the trials included in the meta-analyses. Of the three new trails added to our meta-analysis, two had low risk of bias due to randomized sequence generation, allocation concealment and blinding of participants to this meta-analysis (Table 4). In one trial[22], the person administering the intervention was also blinded and the overall risk of bias was low. In contrast, trials published prior to 2011 had unclear risks of bias due to random sequence generation and allocation concealment and the care providers were not blinded. Given that the two outcomes of interest in our meta-analyses are self-reported, the estimates of treatment effect and between-trial heterogeneity may still be inflated[11] since these studies contribute 216 participants/OA knees to the total sample of 281 participants and OA knees. Our sample was too small to determine whether ultrasound dose influences the treatment effect estimates and all participants had mild to moderate OA so the influence of radiographic OA severity could not be investigated. Nevertheless, our results confirm the findings of previous meta-analyses[8-10] reporting that therapeutic ultrasound reduces knee OA pain and further research will clarify if there is a beneficial treatment effect with respect to physical function outcomes.

While recently published trials have ensured adequate

Table 5 Summary of findings

Outcomes	Difference[1] in ultrasound group mean relative to the control group mean (95%CI)	No of Participants and knees (studies)	Strength of the body of evidence[2]	Inconsistency (I^2)	Outcome specific risk of bias
Pain VAS; NRS Follow-up: 2-8 wk	0.39 standard deviations lower [-0.70-(-0.08)]	281 (5 studies)	Low	36%	High risk of bias of the included studies, imprecision due to small sample size and wide CI
Self-reported physical function WOMAC® LK 3.1 Physical function; Follow-up: 2-8 wk	2.49 points lower (-0.55-0.14)	130 (3 studies)	Very low	0%	High risk of bias of the included studies, imprecision due to very small sample size and wide CI
Walking performance 50 m walk speed (s); 20 m walk speed (s); 6MWT (m) Follow-up: 2-8 wks	0.11 standard deviations lower (-0.59-0.37)	212 (4 studies)	Very low	64%	High risk of bias in the included studies, imprecision due to small sample size and wide CI, inconsistent

[1]Standardized mean difference reported for pain and walking performance; Mean difference reported for self-reported physical function outcomes; [2]GRADE Working Group grades of evidence, High quality: Further research is very unlikely to change our confidence in the estimate of effect. Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Very low quality: We are very uncertain about the estimate. VAS: Horizontal 10 cm visual analogue scale for pain (maximum score = 10, lower score is better); NRS: Pain with movement numeric rating scale (maximum score = 10, lower score is better); WOMAC LK 3.1 Physical function: Likert Scale version of the Western Ontario and McMaster Osteoarthritis Index physical function subscale (maximum score = 80, lower score is better); 6MWT: 6 min walk test (distance measured in metres, higher score is better).

blinding of the participants, the trials included in the current meta-analysis lack methodological rigour in terms of registering the protocol, assuring that risk of bias is low, and recruiting a sample size large enough to ensure precise estimates of efficacy. One trial[22] addressed the feasibility of recruitment and describes the burden of attending a centre for a passive treatment unproven in terms of efficacy as a barrier to participation and retention. Given the growing prevalence of knee OA, strategies to increase sample size and minimize loss to follow up is required in order to conduct a definitive high quality trial addressing this clinical question. If high quality evidence confirms that ultrasound is an effective conservative treatment for knee OA, the wide availability of the treatment ensures rapid translation of this evidence into clinical practice[6,7].

Apart from the walking performance outcome, between-trial homogeneity suggests that the interventions provided in the five trials were comparable. However, subgroup analyses pooling only the trials administering low dose ultrasound reduced the heterogeneity in the pain outcome from 36% to 29%. Table 3 highlights other differences in the interventions administered. Data from two trials that included co-interventions were pooled with data from three trials that did not describe co-interventions. Co-interventions included exercise in both trials[18,23] and the use of analgesics (except during physiotherapy) in one[23] of these trials. These treatments have proven efficacy in the treatment of knee OA and, clinical practice guidelines recommend multicomponent physical therapy for this population[3,4]. However, the interactions between these co-interventions and ultrasound therapy are not known. Another source of variability in the intervention was the application site. Of the five trials included in the meta-analyses, three trials[19,20,23] administered therapeutic ultrasound using a moving sound head applied to the anteromedial and lateral parts of the knee; one trial[18] used a moving sound head applied to the soft tissue around the knee which limited joint range of motion the most; one trial[22] used a stationary sound head applied to the anteromedial tibiofemoral joint line. Despite these differences, the data for pain and self-reported physical function outcomes had acceptable homogeneity for pooling. Taken together, these interventions were effective in reducing knee OA pain. However, further research is required to identify the critical components before a specific protocol for administering therapeutic ultrasound can be recommended.

We chose to focus this review on published trials that compared active and sham ultrasound in order to identify and synthesize the highest quality evidence available in the English-language literature. The funnel plots may be interpreted as evidence of publication bias however asymmetry is attenuated due to the small number of trials and the similar small sample size in each trial. We expected publication bias and therefore we did not plan to perform statistical tests to confirm this. Given that the Cochrane Handbook recommends a minimum of 10 studies in order to run the inferential analysis[24], we could not have conducted statistical testing anyway.

Whereas our systematic review provides high level evidence that therapeutic ultrasound reduces knee OA pain and may improve physical function based on trials in which participants did not know if they received active or sham ultrasound, this review was not without limitations. Random sequence generation and allocation concealment were inadequately described in three of the five meta-analyzed trials. However, it has been suggested that the selection bias which may be present as a result of these study design characteristics are most problematic when the participants' prognosis is easy to assess at the time of

randomization[11]. Since the progression of clinical symptoms in chronic knee OA is difficult to predict, we believe that the influence of selection bias on the effect estimates in the current review is minimal. Inadequate design and/or reporting of trial methods in the trials included in our systematic review limit the strength of the body of evidence which increases the uncertainty in the results. Finally, we chose to focus this review on double blind trials published in English; we did not request data for the three trials published only as abstracts in conference proceedings nor did we search trial registries to identify unpublished data. Furthermore, we were unsuccessful in our attempts to contact two of the primary authors to clarify trial methodology and secure study data suitable for pooling. The inclusion of these unpublished data may have yielded different results.

In conclusion, our meta-analysis suggests that therapeutic ultrasound decreases pain in people with knee OA; however, it is very likely that this conclusion will change when more research is conducted. Therapeutic ultrasound appears to be no better than sham ultrasound for improving self-reported physical function or walking performance. However, the very low strength of the body of evidence for the physical function outcomes leaves us very uncertain about these estimates and further research is necessary to answer this question.

ACKNOWLEDGEMENTS

We thank Heather N deBoer for her assistance with updating the literature search. This review protocol was designed and initiated while NJ MacIntyre was completing the Canadian Institutes of Health Research Randomized Controlled Trials Mentorship Program under the mentorship of M. Bhandari.

COMMENTS

Background

Knee osteoarthritis (OA) is a common problem resulting in pain and mobility limitations in people around the globe. Often symptoms progress to the point where joint replacement is required. Interventions are required that will delay or alleviate the need for surgical interventions. Multiple non-surgical interventions are used to treat patient-reported symptoms of pain and activity limitations and the efficacy of a single treatment component is difficult to evaluate.

Research frontiers

Past systematic reviews evaluating the efficacy of therapeutic ultrasound which pooled data from studies in which participants did not know if there received active ultrasound therapy (blinded) or received the same interventions with the exception of ultrasound therapy (not blinded) suggest ultrasound is effective in reducing knee OA pain whereas reductions in physical function are not statistically significant. Efficacy of ultrasound with respect to patient-reported pain and physical function outcomes should be evaluated in randomized controlled trials in which participants are blinded.

Innovations and breakthroughs

Based on this meta-analysis, including only participants who were blinded to treatment allocation, knee OA pain decreases but physical function is not significantly improved following a course of therapeutic ultrasound treatment. These findings strengthen our confidence in similar findings reported in previous systematic reviews.

Applications

Therapeutic ultrasound appears to be an effective treatment to include in the multi-component non-surgical management of the important clinical problem of knee OA pain. Physical function was not statistically improved immediately following a course of therapeutic ultrasound treatment. Large, methodologically rigorous randomized trials in which the participants, health care providers, and assessors are blinded to treatment allocation and followed up over a longer period are needed to provide a clear answer to this clinical question.

Terminology

Knee osteoarthritis is characterized by degenerative changes to cartilage and other tissues in and around the joint. Therapeutic ultrasound delivers high frequency sound waves and is a physical agent commonly applied by physical therapists to reduce inflammation and/or enhance tissue healing. These effects are postulated to result in improvements in patient-reported pain and physical function (self-reported or performance-based).

Peer review

This meta-analysis was used to determine the efficacy of therapeutic ultrasound versus sham for improving pain and physical function immediately post-intervention in people with knee OA. This is an interesting meta-analysis.

REFERENCES

1 **Guccione AA**, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, Kelly-Hayes M, Wolf PA, Kreger BE, Kannel WB. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *Am J Public Health* 1994; **84**: 351-358 [PMID: 8129049 DOI: 10.2105/AJPH.84.3.351]

2 **Felson DT**. An update on the pathogenesis and epidemiology of osteoarthritis. *Radiol Clin North Am* 2004; **42**: 1-9, v [PMID: 15049520 DOI: 10.1016/S0033-8389(03)00161-1]

3 **Hochberg MC**, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, Towheed T, Welch V, Wells G, Tugwell P. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res* (Hoboken) 2012; **64**: 465-474 [PMID: 22563589 DOI: 10.1186/1471-2474-14-320]

4 **Richmond J**, Hunter D, Irrgang J, Jones MH, Snyder-Mackler L, Van Durme D, Rubin C, Matzkin EG, Marx RG, Levy BA, Watters WC, Goldberg MJ, Keith M, Haralson RH, Turkelson CM, Wies JL, Anderson S, Boyer K, Sluka P, St Andre J, McGowan R. American Academy of Orthopaedic Surgeons clinical practice guideline on the treatment of osteoarthritis (OA) of the knee. *J Bone Joint Surg Am* 2010; **92**: 990-993 [PMID: 20360527 DOI: 10.2106/JBJS.I.00982]

5 **Bélanger AY.** Ultrasound. Evidence-based guide to therapeutic physical agents. In: Philadelphia PA, editor. USA: Lippincott Williams & Wilkins, 2003: 223-261

6 **Nussbaum EL**, Burke S, Johnstone L, Lahiffe G, Robitaille E, Yoshida K. Use of electrophysical agents: Findings and implications of a survey of practice in metro Toronto. *Physiother Can* 2007; **59:** 118-131 [DOI: 10.3138/ptc.59.2.118]

7 **MacIntyre NJ**, Busse JW, Bhandari M. Physical therapists in primary care are interested in high quality evidence regarding efficacy of therapeutic ultrasound for knee osteoarthritis: a provincial survey. *ScientificWorldJournal* 2013; **2013**: 348014 [PMID: 23844391 DOI: 10.1155.2013/348014]

8 **Loyola-Sánchez A**, Richardson J, MacIntyre NJ. Efficacy of ultrasound therapy for the management of knee osteoarthritis: a systematic review with meta-analysis. *Osteoarthritis Cartilage* 2010; **18**: 1117-1126 [PMID: 20637297 DOI: 10.1016/j.joca.2010.06.010]

9 **Rutjes AW**, Nüesch E, Sterchi R, Jüni P. Therapeutic ultrasound for osteoarthritis of the knee or hip. *Cochrane Database Syst Rev* 2010; **(1)**: CD003132 [PMID: 20091539 DOI: 10.1002/14651858.CD003132.pub2]

10 **Wang SY**, Olson-Kellogg B, Shamliyan TA, Choi JY, Ramakrishnan R, Kane RL. Physical therapy interventions for knee pain secondary to osteoarthritis: a systematic review. *Ann Intern Med* 2012; **157**: 632-644 [PMID: 23128863 DOI: 10.7326/0003-4819-157-9-201211060-00007]

11 **Savović J**, Jones H, Altman D, Harris R, Jűni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess* 2012; **16**: 1-82 [PMID: 22989478 DOI: 10.3310/hta16350]

12 **Liberati A**, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; **62**: e1-34 [PMID: 19631507 DOI: 10.1016/j.jclinepi.2009.06.006]

13 **Higgins JPT**, Deeks JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: URL: http://www.cochrane-handbook.org

14 **Higgins JPT**, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: URL: http://www.cochrane-handbook.org

15 **Guyatt GH**, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**: 924-926 [PMID: 18436948 DOI: 10.1136/bmj.39489.470347.AD]

16 **Deeks JJ,** Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: URL: http://www.cochrane-handbook.org

17 **Gurkan I**, Ranganathan A, Yang X, Horton WE, Todman M, Huckle J, Pleshko N, Spencer RG. Modification of osteoarthritis in the guinea pig with pulsed low-intensity ultrasound treatment. *Osteoarthritis Cartilage* 2010; **18**: 724-733 [PMID: 20175971 DOI: 10.1016/j.joca.2010.01.006]

18 **Falconer J**, Hayes KW, Chang RW. Effect of ultrasound on mobility in osteoarthritis of the knee. A randomized clinical trial. *Arthritis Care Res* 1992; **5**: 29-35 [PMID: 1581369]

19 **Ozgönenel L**, Aytekin E, Durmuşoglu G. A double-blind trial of clinical effects of therapeutic ultrasound in knee osteoarthritis. *Ultrasound Med Biol* 2009; **35**: 44-49 [PMID: 18829151 DOI: 10.1016/j.ultrasmedbio.2008.07.009]

20 **Tascioglu F**, Kuzgun S, Armagan O, Ogutler G. Short-term effectiveness of ultrasound therapy in knee osteoarthritis. *J Int Med Res* 2010; **38**: 1233-1242 [PMID: 20925995 DOI: 10.1177/147323001003800404]

21 **Yang PF**, Li D, Zhang SM, Wu Q, Tang J, Huang LK, Liu W, Xu XD, Chen SR. Efficacy of ultrasound in the treatment of osteoarthritis of the knee. *Orthop Surg* 2011; **3**: 181-187 [PMID: 22009649]

22 **Loyola-Sánchez A**, Richardson J, Beattie KA, Otero-Fuentes C, Adachi JD, MacIntyre NJ. Effect of low-intensity pulsed ultrasound on the cartilage repair in people with mild to moderate knee osteoarthritis: a double-blinded, randomized, placebo-controlled pilot study. *Arch Phys Med Rehabil* 2012; **93**: 35-42 [PMID: 22200383 DOI: 10.1155/2013/348014]

23 **Ulus Y**, Tander B, Akyol Y, Durmus D, Buyukakıncak O, Gul U, Canturk F, Bilgici A, Kuru O. Therapeutic ultrasound versus sham ultrasound for the management of patients with knee osteoarthritis: a randomized double-blind controlled clinical study. *Int J Rheum Dis* 2012; **15**: 197-206 [PMID: 22462424 DOI: 10.1111/j.1756-185X.2012.01709.x]

24 **Sterne JAC**, Egger M, Moher, D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from: URL: http://www.cochrane-handbook.org

P- Reviewer: Dai SM, Guo X **S- Editor:** Wen LL **L- Editor**: A **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.91

World J Meta-Anal 2014 August 26; 2(3): 91-97
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

META-ANALYSIS

Role of statin on mortality outcome in pneumonia patients: A meta-analysis

Lekha Saha, Narinder Kumar, Pratibha Khosla, Sharonjeet Kaur

Lekha Saha, Pratibha Khosla, Sharonjeet Kaur, Department of Pharmacology, Post Graduate Institute of Medical Education and Research, Chandigarh 160012, India
Narinder Kumar, Department of Statistics, Panjab University, Chandigarh 160014, India
Author contributions: Saha L, Khosla P and Kaur S were equally involved in planning, data extraction, calculating and writing the manuscript; Kumar N is involved in data extraction, statistical analysis and writing of the manuscript.
Correspondence to: Dr. Lekha Saha, Assistant Professor, Department of Pharmacology, Post Graduate Institute of Medical Education and Research, Sector 12, Chandigarh 160012, India. lekhasaha@rediffmail.com
Telephone: +91-172-2645262 Fax: +91-172-2744401
Received: March 11, 2014 Revised: May 23, 2014
Accepted: June 18, 2014
Published online: August 26, 2014

Abstract

AIM: To determine the role of statin on mortality outcome in patient with pneumonia.

METHODS: For the present meta-analysis, we search the published literatures online through PubMed, Embase, Scopus and the Cochrane Library databases and the search words used were "statins'", "bacteraemia", "pneumonia", and "ICU infections". During the online search our focus was on full text articles, peer-reviewed, observational cohort or case control studies and randomized controlled trials. Those studies were selected whose outcome was hospital mortality among patients with pneumonia whether or not on statins. In this meta-analysis, 30 d mortality was used as the primary outcome as it has been demonstrated in the previous research that 30 d mortality is primarily because of community acquired pneumonia. As all studies were observational, where statin users were compared with historical rather than randomized controls, odds ratio for in-hospital or all-cause 30 d mortality was used as the primary effect measure used in the meta-analysis.

RESULTS: We came across the total 25 studies comprising 35355 patients (2734 statin users and 32621 statin non-users) during the electronic search. Four studies out of 25 were included in the final analysis. In this meta-analysis, when data regarding the use of statin in pneumonia patients on mortality was pooled, its results showed the non-significant effect of the statin on mortality outcome.

CONCLUSION: Although statins seems to be useful in the treatment of pneumonia patients but for statistical conclusion, further randomized controlled trials needs to be done or their results still waited to be published of ongoing trials, with the conclusion that presently statins showing no clinical benefit in the pneumonia patients.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Mortality; Pneumonia; Statin; Meta-analysis

Core tip: The relation between statin treatment and improvement in outcomes in patients with pneumonia have been reported in published literatures. This is believed to be due to anti-inflammatory and immunoregulatory effects rather than an effect on cholesterol metabolism. Recently, however, the potential benefit of statins has been called into question and the study has suggested that there may be confounding factor responsible for the observed benefit. All the studies were observational. In this present meta-analysis of randomized controlled trials, results showed the non-significant effect of the statin on mortality outcome.

Saha L, Kumar N, Khosla P, Kaur S. Role of statin on mortality outcome in pneumonia patients: A meta-analysis. *World J Meta-Anal* 2014; 2(3): 91-97 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/91.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.91

INTRODUCTION

Statins, an inhibitor of 3-hydroxy-3 methylglutaryl coenzyme A reductase (HMG CoA reductase), was used to reduce total cholesterol, low-density lipoprotein cholesterol, apolipoprotein B and triglyceride levels[1-5]. Statins are effective in reducing the cardiovascular mortality and morbidity in patients with or without coronary artery disease. Apart from its lipid lowering effect, statins also possess multiple pleiotropic effects like anti-inflammatory/immunomodulatory effects. These immunomodulatory effects appear to be mediated through the blockade of mevalonate synthesis[6]. The positive impact observed in the cardiovascular disorders is due to their antioxidant effect, plaque stability, favorable coagulation profile, normalization of sympathetic outflow and immunomodulatory effect[7]. By virtue of their pleiotropic effect they have got a role to play in various other disorders like sepsis, dementia, osteoporosis, bacteremia, venous thromboembolism[6-9].

There is an increased rate of morbidity and mortality in hospitalized Patients with pneumonia[10]. Antimicrobial therapy is the mainstay of treatment for pneumonia. Irrespective of the efforts towards more timely treatment of pneumonia and advances in antimicrobial therapy, mortality rates for patients hospitalized with bacterial pneumonia remains relatively unchanged[11]. Studies published in the literatures have demonstrated a link between statin treatment and improvements in patients with bacterial pneumonia. Several large retrospective observational studies have shown a reduced incidence of pneumonia and improvement in outcomes in pneumonia patients taking statins[12-14]. This is believed to be due to anti-inflammatory and immunoregulatory effects rather than an effect on cholesterol metabolism. Statin drugs are known to reduce cytokine levels in stable patients with coronary artery disease and hypercholesterolemia[15-17]. Excess pro-inflammatory cytokines are found to be associated with severity of pneumonia and the development of complications such as acute respiratory distress syndrome[18-22]. Agents with anti-inflammatory properties such as statins might therefore have the potential to improve pneumonia outcomes. Recently, however, the potential benefit of statins has been called into question and the study has suggested that the confounding factors might be the reason for the observed benefit[23]. A series of studies with contradictory results have been published in the literature that has made clinicians hopeful but confused. With this background, we planned to do a meta-analysis of studies using statins in patients with pneumonia to study the role of statin use in the management of pneumonia.

MATERIALS AND METHODS

Literature search

To get the studies published in the literature, we search the various databases online like PubMed, Embase, Scopus and the Cochrane Library. The search words used were "statins", "bacteraemia", "pneumonia"', and "ICU infections". The type of studies looked for were peer-reviewed, full text, observational cohort or case controls studies and randomized controlled trials. Those studies whose outcome were hospital mortality in patients with pneumonia, whether or not received statins. Those studies which included patients with pneumonia due to bacterial infection were included in the present meta-analysis. Those studies were excluded from the meta-analysis where the effects of statins on viral (*i.e.*, influenza and human immunodeficiency virus infection), fungal and protozoan infections were studied. Experimental and laboratory studies were also excluded from the meta-analysis. In our search, there were no time or language limits. For additional relevant information, all references from the identified articles were scanned and search. The following data were extracted and tabulated from the selected articles: study design, patient settings, type of infection, the number of patients included, primary and secondary outcomes. For randomized controlled trials (RCTs), we search online registry of RCTs through http://www.controlled-trial.com for relevant studies. The last date of our review process was July 31, 2013.

Study selection

Those studies which were related to use of statin in pneumonia were first identified. Observational cohort studies whether prospective or retrospective or case control studies which compared 30 d mortality in pneumonia patients among statin users and statin non-users were included in this meta-analysis. We could not identify any randomized control trials.

Statistical analysis

Patients who are taking statin at time of hospital admission were defined as current statin users. The primary outcome in the meta-analysis was 30 d mortality in patients with bacterial pneumonia because studies in the literatures have shown that 30 d mortality is primarily due to community acquired pneumonia rather than other coexisting co morbid conditions. Odds ratio (OR) for in-hospital or all-cause 30 d mortality was measured effect in this meta-analysis as all studies used for meta-analysis were observational, where statin users were compared with historical rather than randomized controls. The heterogeneity between reports was assessed by using χ^2 test of heterogeneity. As significant heterogeneity was found between the reports ($P < 0.05$), we used a random effects model to calculate weighted summary (total) odds ratio and their 95% confidence intervals (CIs) to take the decision. The Mantel-Haenszel method was used to compute the total ORs and CIs. ORs with 95% confidence intervals of all the four studies and the overall odds ratio with 95%CI are depicted in the forest plot (Figure 1) and Peto plot (Figure 2).

RESULTS

The total number of studies searched for literature was

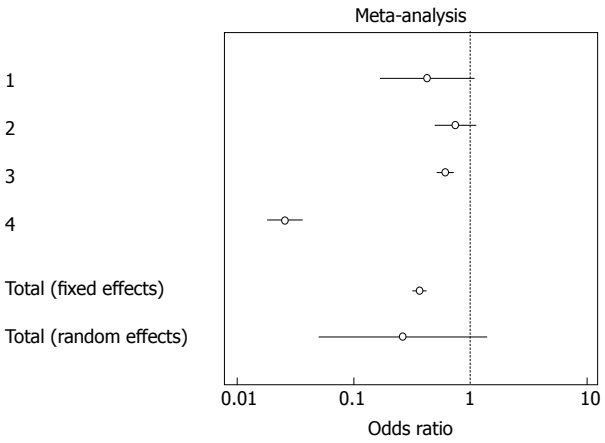


Figure 1 Forest plot with fixed and random effects model comparing mortality among statin users and nonusers.

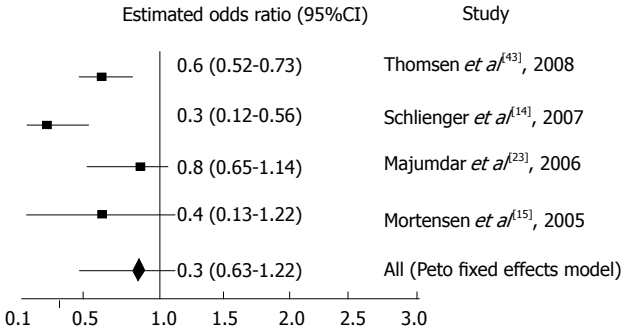


Figure 2 Peto plot with fixed effects model comparing mortality among statin users and nonusers.

twenty five. Among all these, two were review articles[7,24], another fifteen commented upon the infections like sepsis, bacteremia and other post operative settings[10,25-38], one trial is complete, but the results are still unpublished[39] and in an another trial, recruitment is still going on[40]. There were two studies which have evaluated the impact of statin use along with other drugs[41,42]. Four studies were actually evaluated to document the effect of statin in pneumonia[14,15,23,43] (Figure 3).

Four studies out of the total 25 studies comprising 35355 patients (2734 statin users and 32621 statin nonusers) were included in this meta-analysis and in the final analysis. The details of these 4 studies have been mentioned in Table 1. It included both the prospective and retrospective observational studies. All studies were published from 2005 to 2008 (Table 2). In this meta-analysis, when data regarding the use of statin in pneumonia patients on mortality was pooled, its results showed the non-significant effect of the statin on mortality outcome (Figure 1).

DISCUSSION

Studies regarding the statins use have favourably focused on their cardiovascular outcome. Regarding their pleiotropic effects, studies have been conducted which

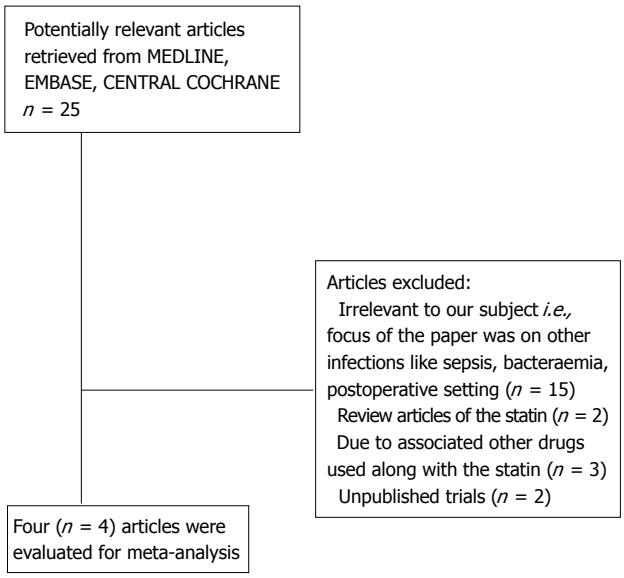


Figure 3 Flow chart of literature search.

have shown varying/mixed results. Different observational studies regarding statin use in pneumonia have too shown invariably mixed results. The calculated weighted summary (total) OR obtained in this meta-analysis suggests that there was no significant effect of the statin use on mortality outcome in patients with pneumonia.

The important known facts regarding the value of any meta-analysis is that it depends upon the quality of reporting of information in individual trials. If we individually comment on the results of these studies regarding their effect on mortality in the pneumonia patients, it was found that three studies reported lower mortality on pneumonia outcome[13,42,43] and one study reported the nonsignificant effect with its use[23]. The studies which we omitted are the ones in which other cardiovascular drugs have been used like aspirin, beta blockers and angiotensin converting enzyme (ACE) inhibitors, which can bias the results of the meta-analysis[42], in another study by Mortensen 2008[41], again along with stations, ACE inhibitors were added and results of starting a group and station along with ACE inhibitors were compared. The third conducted by Van de Garde[13] was excluded due to the reasons that it evaluated the risk outcome associated with the statin use on pneumonia patients rather than the mortality in them.

Chopra and Flanders[44] in their evidence based study demonstrated results favoring and against the use of statins in mortality outcome, with the ultimate conclusions that further randomized, controlled studies are needed to reach a valid conclusion[44]. Kopterides *et al*[45] also looked at the outcomes in a different infections, including the sepsis, bacteraemia and pneumonia, but without reaching the firm conclusions of clinical benefit of the statin therapy. A study conducted by Falagas *et al*[46] has also reported similar results, but no definitive conclusions could be drawn from the pooled data.

The major limitation of this meta-analysis is the in-

Table 1 Summary of the different studies evaluating the effect of statins on mortality in pneumonia

Ref.	Study design	Patient's setting and study groups	Outcomes and odds ration (95%CI)	Conclusion
Thomsen *et al*[43], 2008 Preadmission Use of Statins and outcomes After Hospitalization With Pneumonia: Population-Based Cohort Study of 29900 Patients	Retrospective population based cohort study	29900 adults hospitalized with pneumonia for the first time between January 1, 1997, and December 31, 2004 in northern Denmark. Data on statin and other medication use, comorbidities, socioeconomic markers, laboratory findings,bacteremia, pulmonary complications and death were obtained from medical databases. SU: 1371 NSU: 28529	Of patients with pneumonia, 1371 (4.6%) were current statin users. Mortality among statin users was lower than among nonusers: 10.3% *vs* 15.7% after 30 d and 16.8% *vs* 22.4% after 90 d, corresponding to adjusted 30- and 90-d mortality rate ratios of 0.69 (95%CI: 0.58-0.82) and 0.75 (0.65-0.86).	The use of statins is associated with de creased mortality after hospitalization with pneumonia
Schlienger *et al*[14], 2007 Statins and the Risk of Pneumonia: A population-Based, Nested Case-Control Study	Population-based, retrospective, nested case-control analysis	The study population (134262 patients aged > 30 yr) consisted of 55118 patients who took statins and/or fibrates, 29144 patients with hyperlipidemia not taking lipid-lowering agents, and 50000 randomly selected patients without hyperlipidemia and without lipid-lowering treatment. Authors identified 1253 patients with pneumonia and matched them with 4838 control subjects based on age, sex, general practice, and index date.After adjusting for comorbidity and frequency of visits to general practitioners, we calculated the risks (OR with 95% confidence intervals) of uncomplicated pneumonia, hospitalization for pneumonia with survival, and fatal pneumonia in participants who used statins compared with those who did not. SU: 927 NSU: 326	30 d Mortality ratio 0.262 (0.182-0.377)	Significant lower mortality among statin users as compared to non users
Majumdar *et al*[23], 2006 Statins and outcomes in patients admitted to hospital with community acquired pneumonia: population based prospective cohort study	Prospective population based cohort study	3415 adult patients admitted with CAP and categorised according to use of statins for at least one week before admission and during hospital stay. SU: 325 NSU: 3090	Of 3415 patients with pneumonia admitted to hospital, 624 (18%) died or were admitted to an intensive care unit. Statin users were less likely to die or be admitted to an intensive care unit than non-users [50/325 (15%) *vs* 574/3090 (19%), OR = 0.80, *P* = 0.15]. After more complete adjustment for confounding, however, the OR changed from potential benefit (0.78, adjusted for age and sex) to potential harm (1.10, fully adjusted including propensity scores, 95%CI: 0.76-1.60)	Statins are not associated with reduced mortality or need for admission to an intensive care unit in patients with pneumonia; reports of benefit in the setting of sepsis may be a result of confounding.
Mortensen *et al*[15], 2005 The effect of prior statin use on 30-d mortality for patients hospitalized with community-acquired pneumonia	Retrospective cohort study	A retrospective cohort study conducted at two tertiary teaching hospitals. Eligible subjects were admitted with a diagnosis of, had a chest x-ray consistent with, and had a discharge ICD-9 diagnosis of pneumonia. Subjects were excluded if they were "comfort measures only" or transferred from another acute care hospital. Subjects were considered to be on a medication if they were taking it at the time of presentation. Data was abstracted on 787 subjects at the two hospitals. SU: 110 NSU: 677	In the multivariable regression analysis, after adjusting for the propensity score and processes of care, the use of statins at presentation (OR = 0.36, 95%CI: 0.14-0.92) was significantly associated with decreased 30-d mortality	Prior outpatient statin use was associated with decreased mortality in patients hospitalized with community-acquired pneumonia despite their use being associated with comorbid illnesses likely to contribute to increased mortality

SU: Statin users; NSU: Nonstatin users.

Table 2 Individual and total evaluation number of statin users and nonusers on the mortality in pneumonia in various studies

Ref.	Mortality in SU	Mortality in NSU	OR	95%CI
Thomsen *et al*[43]	141/1372	4489/28528	0.61	0.514-0.732
Schlienger *et al*[14]	54/927	229/326	0.262	0.182-0.377
Majumdar *et al*[23]	25/325	309/3090	0.75	0.491-1.147
Mortensen *et al*[15]	5/110	67/677	0.434	0.171-1.101
Total (fixed effects)	225/2734	5094/32621	0.373	0.322 - 0.431
Total (Random effects)	225/2734	5094/32621	0.267	0.0499 - 1.428
Test for heterogeneity	246.1742			
Computed value				
Degrees of freedom	3			
Significance level	$P < 0.05$			

SU: Statin users; NSU: Nonstatin users.

clusion of only observational studies and marked heterogeneity seen in these clinical studies. The heterogeneity was related to the design of the studies along with the sample size, current statin use, associated Co morbidities, and identification of the clinical infection seen in these patients. Another important issue regarding statin use is a healthy user effect, which refers to those patients receiving statins also belongs to higher socioeconomic classes than patients who do not. These patients have the advantage of higher education, better awareness regarding their health which increases the complaint regarding their visits to doctors and treatment schedule, with the results for a better outcome in case of infection.

Based on currently available clinical evidence it can be concluded that statins use shows no clinical benefit in terms of mortality outcome in the pneumonia patients. This conclusion is despite the fact the three studies reported the positive outcome regarding the statin use in the pneumonia patients as compared to the only single one reporting the negative results. As the studies included were heterogeneous, therefore the results of ongoing trials and further randomized controlled trials could only provide a definitive evidence regarding the effect of statins on mortality in pneumonia.

COMMENTS

Background

Statins, an 3-hydroxy-3 methylglutaryl coenzyme A reductase (HMG CoA) reductase inhibitor, used to reduce total cholesterol, low-density lipoprotein cholesterol, apolipoprotein B and triglyceride levels. Statins are also effective in reducing the cardiovascular mortality and morbidity in patients with or without coronary artery disease. Apart from its lipid lowering effect, statins also possess multiple pleiotropic effects like anti-inflammatory/immunomodulatory effects. The association between statin use and improvement in outcomes in patients with pneumonia has been reported in published literatures. This is believed to be due to anti-inflammatory and immunoregulatory effects rather than an effect on cholesterol metabolism.

Research frontiers

Several large retrospective observational studies have shown a reduced incidence of pneumonia and improvement in outcomes in pneumonia patients taking statins. Recently, however, the potential benefit of statins has been called into question and study has suggested that the confounding factors might be the reason for the observed benefit of those studies. A series of studies with contradictory results have been published in the literature that has made clinicians hopeful but confused. With this background, the authors planned to do a meta-analysis of studies using statins in patients with pneumonia to study the role of statin use in the management of pneumonia.

Innovations and breakthroughs

Four studies out of the total 25 studies comprising 35355 patients (2734 statin users and 32621 statin non-users) were included in this meta-analysis and in the final analysis. It included both the prospective and retrospective observational studies. All studies were published from 2005 to 2008. Odds ratio (OR) for in-hospital or all-cause 30 d mortality was measured effect in this meta-analysis as all studies used for meta-analysis were observational, where statin users were compared with historical rather than randomized controls. The heterogeneity between reports was assessed by using χ^2 test of heterogeneity. As significant heterogeneity was found between the reports ($P < 0.05$), the authors used a random effects model to calculate weighted summary (total) odds ratio and their 95%CIs to take the decision. The Mantel-Haenszel method was used to compute the total OR and CIs. OR with 95%CIs of all the four studies and the overall OR with 95%CI are depicted in the forest plot and Peto plot.

Applications

In this meta-analysis, when data regarding the use of statin in pneumonia patients on mortality was pooled, its results showed the non-significant effect of the statin on mortality outcome. As the studies included were heterogeneous, therefore the results of ongoing trials and further randomized controlled trials could only provide a definitive evidence regarding the effect of statins on mortality in pneumonia.

Terminology

HMG CoA reductase is 3-hydroxy-3 methylglutaryl coenzyme A reductase is the rate-controlling enzyme of the mevalonate pathway, the metabolic pathway that produces cholesterol and other isoprenoids. Pleiotropic effects of statin are those properties of statins other than cholesterol lowering effects like the effect on osteoporosis, dementia and so on.

Peer review

Authors done a meta-analysis to identify the effectiveness of statin use on mortality due to pneumonia. Authors conducted study well.

REFERENCES

1 **Simvastatin Survival Study Group.** Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; **344**: 1383-1389 [PMID: 7968073]

2 **Sacks FM**, Pfeffer MA, Moye LA, Rouleau JL, Rutherford JD, Cole TG, Brown L, Warnica JW, Arnold JM, Wun CC, Davis BR, Braunwald E. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. *N Engl J Med* 1996; **335**: 1001-1009 [PMID: 8801446]

3 **The Long-term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group.** Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cho-

lesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. *N Engl J Med* 1998; **339**: 1349-1357 [PMID: 9841303 DOI: 10.1056/NEJM199811053391902]
4 **West of Scotland Coronary Prevention Group.** West of Scotland Coronary Prevention Study: identification of High-risk Groups and Comparison with Other Cardiovascular Intervention Trials. *Lancet* 1996; **348**: 1339-1342 [DOI: 10.1016/S0140-6736(96)04292-4]
5 **Downs JR**, Clearfield M, Weis S, Whitney E, Shapiro DR, Beere PA, Langendorfer A, Stein EA, Kruyer W, Gotto AM. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *JAMA* 1998; **279**: 1615-1622 [PMID: 9613910]
6 **Terblanche M**, Almog Y, Rosenson RS, Smith TS, Hackam DG. Statins and sepsis: multiple modifications at multiple levels. *Lancet Infect Dis* 2007; **7**: 358-368 [PMID: 17448939]
7 **Tandon V,** Bano G, Khajuria V, Parihar A, Gupta S. Pleiotropic effects of statins. *Indian J Pharmacol* 2005; **37**: 77-85 [DOI: 10.4103/0253-7613.15106]
8 **Waldman A**, Kritharides L. The pleiotropic effects of HMG-CoA reductase inhibitors: their role in osteoporosis and dementia. *Drugs* 2003; **63**: 139-152 [PMID: 12515562 DOI: 10.2165/00003495-200363020-00002]
9 **Chopra V**, Choksi PU, Cavusoglu E. Beyond lipid lowering: the anti-hypertensive role of statins. *Cardiovasc Drugs Ther* 2007; **21**: 161-169 [PMID: 17468937]
10 **Kruger P**, Fitzsimmons K, Cook D, Jones M, Nimmo G. Statin therapy is associated with fewer deaths in patients with bacteraemia. *Intensive Care Med* 2006; **32**: 75-79 [PMID: 16283159]
11 **Thomsen RW**, Riis A, Nørgaard M, Jacobsen J, Christensen S, McDonald CJ, Sørensen HT. Rising incidence and persistently high mortality of hospitalized pneumonia: a 10-year population-based study in Denmark. *J Intern Med* 2006; **259**: 410-417 [PMID: 16594909 DOI: 10.1111/j.1365-2796.2006.01629.x]
12 **Fry AM**, Shay DK, Holman RC, Curns AT, Anderson LJ. Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988-2002. *JAMA* 2005; **294**: 2712-2719 [PMID: 16333006 DOI: 10.1001/jama.294.21.2712.]
13 **van de Garde EM**, Hak E, Souverein PC, Hoes AW, van den Bosch JM, Leufkens HG. Statin treatment and reduced risk of pneumonia in patients with diabetes. *Thorax* 2006; **61**: 957-961 [PMID: 16809409 DOI: 10.1136/thx.2006.062885]
14 **Schlienger RG**, Fedson DS, Jick SS, Jick H, Meier CR. Statins and the risk of pneumonia: a population-based, nested case-control study. *Pharmacotherapy* 2007; **27**: 325-332 [PMID: 17316144 DOI: 10.1592/phco.27.3.325]
15 **Mortensen EM**, Restrepo MI, Anzueto A, Pugh J. The effect of prior statin use on 30-day mortality for patients hospitalized with community-acquired pneumonia. *Respir Res* 2005; **6**: 82 [PMID: 16042797 DOI: 10.1186/1465-9921-6-82]
16 **Jialal I**, Stein D, Balis D, Grundy SM, Adams-Huet B, Devaraj S. Effect of hydroxymethyl glutaryl coenzyme a reductase inhibitor therapy on high sensitive C-reactive protein levels. *Circulation* 2001; **103**: 1933-1935 [PMID: 11306519 DOI: 10.1161/01.CIR.103.15.1933]
17 **Musial J**, Undas A, Gajewski P, Jankowski M, Sydor W, Szczeklik A. Anti-inflammatory effects of simvastatin in subjects with hypercholesterolemia. *Int J Cardiol* 2001; **77**: 247-253 [PMID: 11182189]
18 **Strandberg TE**, Vanhanen H, Tikkanen MJ. Effect of statins on C-reactive protein in patients with coronary artery disease. *Lancet* 1999; **353**: 118-119 [PMID: 10023901]
19 **Almirall J**, Bolíbar I, Toran P, Pera G, Boquet X, Balanzó X, Sauca G. Contribution of C-reactive protein to the diagnosis and assessment of severity of community-acquired pneumonia. *Chest* 2004; **125**: 1335-1342 [PMID: 15078743]
20 **Fernandez-Serrano S**, Dorca J, Coromines M, Carratala J, Gudiol F, Manresa F. Molecular inflammatory responses measured in blood of patients with severe community acquired pneumonia. *Clin Diagn Lab Immunol* 2003; **10**: 813-820 [PMID: 12965910 DOI: 10.1128/CDLI.10.5.813–820.2003]
21 **Glynn P**, Coakley R, Kilgallen I, Murphy N, O'Neill S. Circulating interleukin 6 and interleukin 10 in community acquired pneumonia. *Thorax* 1999; **54**: 51-55 [PMID: 10343632]
22 **Puren AJ**, Feldman C, Savage N, Becker PJ, Smith C. Patterns of cytokine expression in community-acquired pneumonia. *Chest* 1995; **107**: 1342-1349 [PMID: 7750329 DOI: 10.1378/chest.107.5.1342]
23 **Majumdar SR**, McAlister FA, Eurich DT, Padwal RS, Marrie TJ. Statins and outcomes in patients admitted to hospital with community acquired pneumonia: population based prospective cohort study. *BMJ* 2006; **333**: 999 [PMID: 17060337 DOI: 10.1136/bmj.38992.565972.7C]
24 **Daniel Pella**, Rafael Rybar and Viola Mechirova. Pleoitropic effect of statins. *Acta Cardiol Sin* 2005; **21**: 190-198
25 **Mortensen EM**, Restrepo MI, Copeland LA, Pugh JA, Anzueto A, Cornell JE, Pugh MJ. Impact of previous statin and angiotensin II receptor blocker use on mortality in patients hospitalized with sepsis. *Pharmacotherapy* 2007; **27**: 1619-1626 [PMID: 18041882 DOI: 10.1592/phco.27.12.1619]
26 **Yang KC**, Chien JY, Tseng WK, Hsueh PR, Yu CJ, Wu CC. Statins do not improve short-term survival in an oriental population with sepsis. *Am J Emerg Med* 2007; **25**: 494-501 [PMID: 17543651]
27 **Tseng MY**, Hutchinson PJ, Czosnyka M, Richards Hugh, Pickard JD, Kirkpatrick PJ. Effects of acute pravastatin treatment on intensity of rescue therapy, length of inpatient stay, and 6-month outcome in patients after aneurysmal subarachnoid hemorrhage. *Stroke* 2007; **38**: 1545–1550 [DOI: 10.1161/STROKEAHA.109.556332]
28 **Martin CP**, Talbert RL, Burgess DS, Peters JI. Effectiveness of statins in reducing the rate of severe sepsis: a retrospective evaluation. *Pharmacotherapy* 2007; **27**: 20-26 [PMID: 17192158 DOI: 10.1592/phco.27.1.20]
29 **Gupta R**, Plantinga LC, Fink NE, Melamed ML, Coresh J, Fox CS, Levin NW, Powe NR. Statin use and sepsis events [corrected] in patients with chronic kidney disease. *JAMA* 2007; **297**: 1455-1464 [PMID: 17405971 DOI: 10.1001/jama.297.13.1455]
30 **Almog Y**, Novack V, Eisinger M, Porath A, Novack L, Gilutz H. The effect of statin therapy on infection-related mortality in patients with atherosclerotic diseases. *Crit Care Med* 2007; **35**: 372-378 [PMID: 17205009]
31 **Fernandez R**, De Pedro VJ, Artigas A. Statin therapy prior to ICU admission: protection against infection or a severity marker? *Intensive Care Med* 2006; **32**: 160-164 [PMID: 16086178]
32 **Hackam DG**, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet* 2006; **367**: 413-418 [PMID: 16458766]
33 **Almog Y**, Shefer A, Novack V, Maimon N, Barski L, Eizinger M, Friger M, Zeller L, Danon A. Prior statin therapy is associated with a decreased rate of severe sepsis. *Circulation* 2004; **110**: 880-885 [PMID: 15289367 DOI: 10.1161/01.CIR.0000138932.17956.F1]
34 **Thomsen RW**, Hundborg HH, Johnsen SP, Pedersen L, Sørensen HT, Schønheyder HC, Lervang HH. Statin use and mortality within 180 days after bacteremia: a population-based cohort study. *Crit Care Med* 2006; **34**: 1080-1086 [PMID: 16484926]
35 **Liappis AP**, Kan VL, Rochester CG, Simon GL. The effect of statins on mortality in patients with bacteremia. *Clin Infect Dis* 2001; **33**: 1352-1357 [PMID: 11565076 DOI:

10.1086/323334]
36 **Subramaniam K**, Koch CG, Bashour A, O'Connor M, Xu M, Gillinov AM, Starr NJ. Preoperative statin intake and morbid events after isolated coronary artery bypass grafting. *J Clin Anesth* 2008; **20**: 4-11 [PMID: 18346602 DOI: 10.1016/j.jclinane.2007.09.003]
37 **Coleman CI**, Lucek DM, Hammond J, White CM. Preoperative statins and infectious complications following cardiac surgery. *Curr Med Res Opin* 2007; **23**: 1783-1790 [PMID: 17597556]
38 **Hauer-Jensen M**, Fort C, Mehta JL, Fink LM. Influence of statins on postoperative wound complications after inguinal or ventral herniorrhaphy. *Hernia* 2006; **10**: 48-52 [PMID: 16151608]
39 Pravastatin and ventilatory associated pneumonia. Available from: URL: http://clinicaltrials.gov
40 Statin-Vap statin-Vap - Statins and Ventilator-Associated Pneumonia. Available from: URL: http: //clinicaltrials.gov
41 **Mortensen EM**, Pugh MJ, Copeland LA, Restrepo MI, Cornell JE, Anzueto A, Pugh JA. Impact of statins and angiotensin-converting enzyme inhibitors on mortality of subjects hospitalised with pneumonia. *Eur Respir J* 2008; **31**: 611-617 [PMID: 17959631 DOI: 10.1183/09031936.00162006]
42 **Chalmers JD**, Singanayagam A, Murray MP, Hill AT. Prior statin use is associated with improved outcomes in community-acquired pneumonia. *Am J Med* 2008; **121**: 1002-1007.e1 [PMID: 18954848]
43 **Thomsen RW**, Riis A, Kornum JB, Christensen S, Johnsen SP, Sørensen HT. Preadmission use of statins and outcomes after hospitalization with pneumonia: population-based cohort study of 29,900 patients. *Arch Intern Med* 2008; **168**: 2081-2087 [PMID: 18955636 DOI: 10.1001/archinte.168.19.2081]
44 **Chopra V**, Flanders SA. Does statin use improve pneumonia outcomes? *Chest* 2009; **136**: 1381-1388 [PMID: 19892677 DOI: 10.1378/chest.09-0941]
45 **Kopterides P**, Falagas ME. Statins for sepsis: a critical and updated review. *Clin Microbiol Infect* 2009; **15**: 325-334 [PMID: 19416304 DOI: 10.1111/j.1469-0691.2009.02750.x]
46 **Falagas ME**, Makris GC, Matthaiou DK, Rafailidis PI. Statins for infection and sepsis: a systematic review of the clinical evidence. *J Antimicrob Chemother* 2008; **61**: 774-785 [PMID: 18263570 DOI: 10.1093/jac/dkn019]

P- Reviewer: Surani S, Undela K **S- Editor:** Ji FF **L- Editor:** A **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.98

World J Meta-Anal 2014 August 26; 2(3): 98-106
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

META-ANALYSIS

Robot-assisted *vs* laparoscopy-assisted gastrectomy for gastric cancer: A meta-analysis based on 3518 subjects

Zhi-Dong Lin, Mao Liu, Dan Tang, Huan Li, Bai-Meng Zhang

Zhi-Dong Lin, Dan Tang, Bai-Meng Zhang, Department of General Surgery, the Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai 519000, Guangdong Province, China
Mao Liu, Department of Cardiology, the Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai 519000, Guangdong Province, China
Huan Li, Sun Yat-Sen University Cancer Center, Guangzhou 510060, Guangdong Province, China
Author contributions: Lin ZD and Liu M contributed equally to this work; Lin ZD, Tang D and Zhang BM designed the research; Tang D, Li H and Liu M performed the research; Lin ZD and Liu M contributed analytic tools and analyzed the data; Lin ZD, Tang D, Liu M and Zhang BM wrote the paper; Lin ZD and Liu M contributed equally to this work.
Correspondence to: Dr. Zhi-Dong Lin, Department of General Surgery, the Fifth Affiliated Hospital of Sun Yat-Sen University, No.52 Meihua East Rd, Zhuhai 519000, Guangdong Province, China. linzhidongtd@21cn.com
Telephone: +86-756-2528811 Fax: +86-756-2528333
Received: January 8, 2014 Revised: May 16, 2014
Accepted: June 10, 2014
Published online: August 26, 2014

Abstract

AIM: To compare the short-term clinical outcomes of robot-assisted gastrectomy (RAG) with laparoscopy-assisted gastrectomy (LAG) in gastric cancer patients.

METHODS: Articles were identified through a literature search of Pubmed, EMBASE, Scopus, Web of Science, Chinese National Knowledge Infrastructure and the Cochrane Library. Weighted mean differences (WMDs) and odds ratios (ORs) were selected as effect sizes for quantitative variables and qualitative variables, respectively. And 95%CIs were also calculated.

RESULTS: A total of 13 studies with 3518 patients were included. RAG was associated with longer operative time (WMD = 46.26 min, 95%CI: 31.89-60.63, P < 0.00001), less blood loss [WMD = -37.19 mL, 95%CI: -60.16-(-14.23), P = 0.002] and shorter postoperative hospital stay [WMD = -0.65 d, 95%CI: -1.24-(-0.05), P = 0.03] than LAG. No significant difference in the numbers of retrieved lymph nodes was found between the two groups (WMD = 1.46, 95%CI: -0.19-3.10, P = 0.08). There was no significant difference in mortality (OR = 1.55, 95%CI: 0.49-4.94, P = 0.45), overall complications (OR = 1.00, 95%CI: 0.80-1.26, P = 0.98), anastomosis leakage (OR = 1.02, 95%CI: 0.62-1.65, P = 0.95) and anastomosis stenosis rates (OR = 0.54, 95%CI: 0.18-1.57, P = 0.25).

CONCLUSION: RAG is effective and safe in the treatment of gastric cancer. RAG is a promising alternative to laparoscopic surgery. Long-term randomized controlled studies with large scale and improved designs are needed to further evaluate the long-term outcomes.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Gastric cancer; Robot; Laparoscopy; Gastrectomy; Meta-analysis

Core tip: A total of 13 studies with 3518 patients were included in this meta-analysis. The results indicated that robot-assisted gastrectomy was associated with longer operative time (WMD = 46.26 min, 95%CI: 31.89, 60.63, P < 0.00001), less blood loss [WMD = -37.19 mL, 95%CI: -60.16-(-14.23), P = 0.002] and shorter postoperative hospital stay [WMD = -0.65 d, 95%CI: -1.24-(-0.05), P = 0.03] than laparoscopy-assisted gastrectomy. Robot-assisted gastrectomy is effective and safe in the treatment of gastric cancer and will be a promising alternative to laparoscopic surgery. Long-term randomized controlled studies with large scale and improved designs are needed to further evaluate the long-term outcomes.

Lin ZD, Liu M, Tang D, Li H, Zhang BM. Robot-assisted *vs* laparoscopy-assisted gastrectomy for gastric cancer: A meta-analysis

based on 3518 subjects. *World J Meta-Anal* 2014; 2(3): 98-106 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/98.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.98

INTRODUCTION

Gastric cancer is the fourth leading cancer and second leading cause of cancer death in the world[1,2]. At present, radical gastrectomy with lymph node (LN) dissection is still the mainstay of treatment for gastric cancer[3]. Since 1994, the laparoscopy-assisted gastrectomy (LAG) has become widely accepted in Asian countries because it offers less invasiveness and pain, speedier recovery, milder morbidity and shorter hospital stay[4-7]. According to the report of Japanese Society of Endoscopic Surgery, the total number of patients who had undergone LAG for gastric cancer was 34645 until 2013[8]. However, the instruments of LAG have a limited range of motion and are usually associated with a long learning curve, especially in LN dissection[9].

Another minimally invasive approach for gastric cancer seems to be more promising. Hashizume *et al*[10] had performed distal gastrectomy successfully with the assistance of the da Vinci computer-enhanced surgical system in 2002. They found that the robotic system enhanced visualization of both the operative field and precision of the necessary techniques. It may therefore help surgeons overcome many of the difficulties associated with the endoscopic approach. Since then, several studies[11-15] have been conducted to evaluate the safety and efficacy of robot-assisted gastrectomy (RAG) for gastric cancer. However, most of them were case control studies and their sample sizes were rather small. Therefore, in this study, we conducted a meta-analysis to compare the short-term clinical outcomes of RAG with LAG in gastric cancer patients.

MATERIALS AND METHODS

Search strategy

We performed an electronic search of Pubmed, EMBASE, Scopus, Web of Science, Cochrane Library and Chinese National Knowledge Infrastructure from the inception to December 13th, 2013. The following search terms were used: gastric cancer, gastric carcinoma, gastrectomy, robotic, robot, laparoscopy and laparoscopic. Only English and Chinese articles were considered. We also searched additional articles through the reference lists of related papers. Two investigators screened the articles independently.

Study selection

Two investigators identified appropriate articles and conducted data extraction independently. Eligible studies should match all of the following: (1) study design: prospective or retrospective cohort studies, randomized or nonrandomized controlled studies, case-control studies; (2) study population: gastric cancer patients who received RAG or LAG; (3) grouping: RAG group *vs* LAG group; and (4) outcomes: intraoperative outcome (operative time, blood loss, number of retrieved LNs, conversion to open gastrectomy) and postoperative outcome (overall complications, anastomosis leakage, anastomosis stenosis, bleeding, intestinal obstruction, mortality and postoperative hospital stay). Meeting abstracts, case reports, editorials and reviews were excluded.

Data extraction and quality assessment

We extracted the study type, country, patient characteristics, age, clinical outcomes, operating cost and the number of cases for each article. The quality of the included studies was evaluated by Newcastle-Ottawa quality assessment (NOS) scale[16]. A study can be awarded a highest score of nine. Data extraction was completed independently by two investigators.

Statistical analysis

All statistical tests were performed with Review Manager 5.1 software (Cochrane Collaboration, Copenhagen, Denmark). In this study, I^2 was used to investigate the heterogeneity. In the analysis process, if $I^2 \geqslant 50\%$, we ran a random-effect model. On the other hand, a fixed-effect model was chosen if $I^2 < 50\%$. In the analysis of quantitative variables (operative time, blood loss, number of retrieved LNs and postoperative hospital stay), we chose weighted mean difference (WMD) with 95%CI as summary statistics. As for qualitative variables (overall complications, anastomosis leakage, anastomosis stenosis, bleedings, intestinal obstruction, conversion to open gastrectomy and mortality), odds ratios (ORs) with 95% CIs were used accordingly. A value of $P < 0.05$ (two-tailed test) was considered statistically significant.

RESULTS

Study selection

At the beginning of search process, 535 publications were reviewed. After a screening process, 11 retrospective studies[1,3,6,9,11-15,17,18], one nonrandomized prospective study[19] and one randomized control trial[20] were included (Figure 1). Twelve studies[1,3,6,9,11-13,15,17-20] were from Asia and one[14] from Europe. Totally, 3518 patients with gastric cancer were included in this meta-analysis. Among them, 1143 cases were in RAG group, the other 2375 patients received LAG (Table 1). Ten studies[1,3,9,12-15,17-19] were published in English, and three[6,11,20] published in Chinese. No significant publication bias was found (Figure 2).

Comparison of intraoperative outcomes between RAG and LAG group

In this pooled analysis, operative time, blood loss and number of retrieved LNs were included. In total, there were 13 studies[1,3,6,9,11-15,17-20] which reported of the operative time, 12 studies[1,3,6,9,11-14,17-20] reported of blood loss

Table 1 Main characteristic of the included studies

Ref.	Country	Group	No. of patients	Age (yr)	Males (%)	BMI (kg/m^2)	TNM stage (Ⅰ/Ⅱ/Ⅲ/Ⅳ)	NOS score
Eom *et al*[17]	South Korea	RAG	30	52.8	70	24.2	25/3/2/0	6
		LAG	62	57.9	66.1	24.1	56/6/0/0	
Huang *et al*[9]	Taiwan China	RAG	39	65.1	48.7	24.2	29/7/3/0	6
		LAG	64	65.6	67.2	24.7	55/9/0/0	
Hyun *et al*[18]	South Korea	RAG	38	54.2	65.8	23.8	30/5/3/0	7
		LAG	83	60.3	66.3	23.8	67/9/7/0	
Kang *et al*[12]	South Korea	RAG	100	53.2	63	23.7	82/11/7/0	6
		LAG	282	58.8	67.7	23.6	NR	
Kim *et al*[3]	South Korea	RAG	436	54.2	60.8	23.6	350/51/32/0	7
		LAG	861	58.8	63.9	23.5	714/96/43/0	
Kim *et al*[13]	South Korea	RAG	16	53.8	62.5	21.3	NR	6
		LAG	11	57.9	90.9	25.3	NR	
Liu *et al*[6]	China	RAG	48	51.8	85.4	21.2	14/5/27/2	7
		LAG	48	52.1	83.3	21	16/6/23/3	
Noshiro *et al*[19]	Japan	RAG	21	66	66.7	22.8	18/-/-/-	6
		LAG	160	69	63.8	21.8	113/-/-/-	
Pugliese *et al*[14]	Italy	RAG	16	71	NR	28.8	NR	7
		LAG	48	71	NR	28.8	NR	
Woo *et al*[1]	South Korea	RAG	236	54	57.6	23.5	236/0/0/0	6
		LAG	591	58.3	61.6	23.5	591/0/0/0	
Yoon *et al*[15]	South Korea	RAG	36	53.9	50	23.2	29/7/0/0	7
		LAG	65	56.9	47.7	23.6	55/7/3/0	
Zhang *et al*[11]	China	RAG	97	56.1	68	22.5	23/22/52/0	7
		LAG	70	54.8	70	21.7	8/17/45/0	
Zhao *et al*[20]	China	RAG	30	71.8	73.3	23.6	2/18/9/1	8
		LAG	30	72.4	76.7	23.9	1/25/3/1	

BMI: Body mass index; TNM: Tumor node metastases; NOS: Newcastle-ottawa quality scale; RAG: Robot-assisted gastrectomy; LAG: Laparoscopy-assisted gastrectomy; NR: Not Reported.

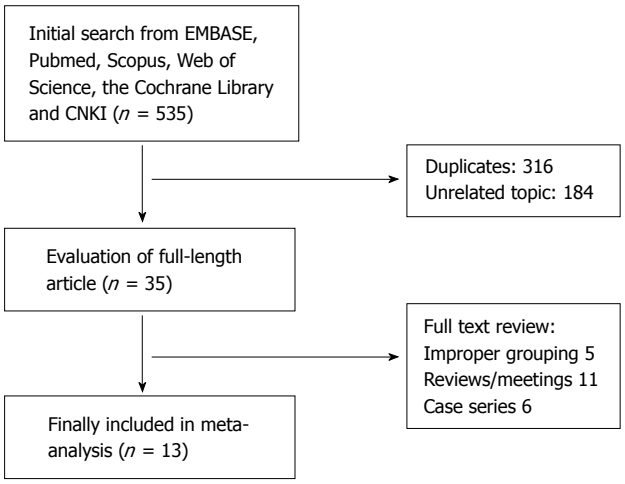


Figure 1 Flow chart of literature search for meta-analysis.

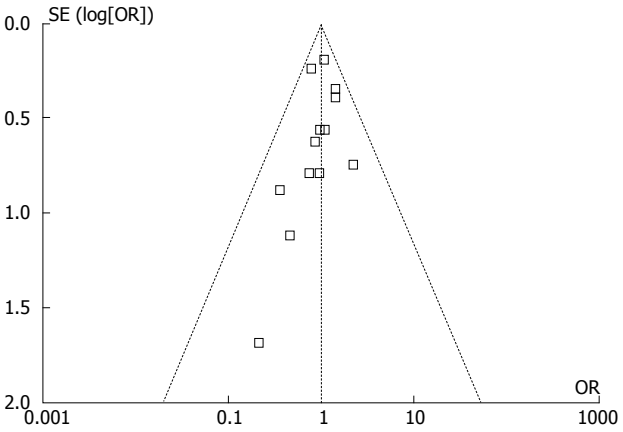


Figure 2 Funnel plot of included studies in this meta-analysis.

and 12 studies[1,3,6,9,11,13-15,17-20] reported of the number of retrieved LNs. In the heterogeneity tests of operative time, blood loss and number of retrieved LNs, I^2 were 91%, 94% and 70%, respectively. Accordingly, we chose the random-effect model.

As shown in Figure 3A, the mean operative time for the RAG group was on average 46 min longer than the LAG group (WMD = 46.26 min, 95%CI: 31.89-60.63, $P < 0.00001$), while mean blood loss was significantly less in the RAG group [WMD = -37.19 mL, 95%CI: -60.16-(-14.23), $P = 0.002$]. The pooled results also indicated that there was no significant difference in the number of retrieved LNs between the two groups (WMD = 1.46, 95%CI: -0.19-3.10, $P = 0.08$).

Comparison of postoperative outcomes between RAG and LAG group

Overall complications, anastomosis leakage and anastomosis stenosis were included for analysis. Information in detail is shown in Figure 3B. Thirteen studies[1,3,6,9,11-15,17-20] which reported of the overall complications were included. No statistical heterogeneity was found in this analysis

A

Study or Subgroup	RAG Mean	RAG SD	RAG Total	LAG Mean	LAG SD	LAG Total	Weight	Mean difference IV, Random, 95%CI	Year
1.1.1 Operative time									
Kim	259.2	38.9	16	203.9	36.4	11	6.9%	55.30 [26.56, 84.04]	2010
Pugliese	344	62	16	235	23	48	6.6%	109.00 [77.93, 140.07]	2010
Woo	219.5	46.8	236	170.7	55.8	591	9.1%	48.80 [41.32, 56.28]	2011
Eom	229.1	52.2	30	189.4	52.2	62	7.7%	39.70 [16.95, 62.45]	2012
Huang	415.9	101.2	39	362.2	72.8	64	5.9%	53.70 [17.27, 90.13]	2012
Zhang	272.3	46.1	97	240.3	89.1	70	7.7%	32.00 [9.20, 54.80]	2012
Kang	202.1	52.3	100	173.5	51.2	282	9.0%	28.60 [16.73, 40.47]	2012
Kim	226	54	436	176	63	861	9.4%	50.00 [43.41, 56.59]	2012
Yoon	305.8	115.8	36	210.2	57.7	65	5.5%	95.60 [55.26, 135.94]	2013
Hyun	234.4	48	38	220	60.6	83	8.1%	14.40 [-5.67, 34.47]	2013
Zhao	156.7	14.9	30	159	27.1	30	9.1%	-2.30 [-13.37, 8.77]	2013
Liu	238	29.6	48	221.5	33.1	48	9.0%	16.50 [3.94, 29.06]	2013
Noshiro	439	86	21	315	90	160	5.6%	124.00 [84.66, 163.34]	2013
Subtotal (95%CI)			1143			2375	100.0%	46.26 [31.89, 60.63]	
Heterogeneity: Tau^2 = 559.65; χ^2 = 139.08, df = 12 (P < 0.00001); I^2 = 91%									
Test for overall effect: Z = 6.31 (P < 0.00001)									
1.1.2 Blood loss									
Pugliese	90	48	16	148	53	48	8.6%	-58.00 [-85.89, -30.11]	2010
Kim	30.3	15.1	16	44.7	37.1	11	9.0%	-14.40 [-37.54, 8.74]	2010
Woo	91.6	152.6	236	147.9	269	591	8.5%	-56.30 [-85.44, -27.16]	2011
Eom	152.8	173.5	30	88.3	173.5	62	4.8%	64.50 [-11.13, 140.13]	2012
Huang	93.9	89	39	163.7	154.7	64	7.0%	-69.80 [-116.88, -22.72]	2012
Kang	93.3	84.6	100	173.5	145.2	282	8.9%	-80.20 [-103.91, -56.49]	2012
Kim	85	160	436	112	229	861	9.0%	-27.00 [-48.44, -5.56]	2012
Zhang	80.8	53.1	97	153.7	26.4	70	9.6%	-72.90 [-85.14, -60.66]	2012
Hyun	131.3	10.1	38	130.5	17.8	83	9.8%	0.80 [-4.20, 5.80]	2013
Noshiro	91.4	108.6	21	109.5	165.7	160	6.5%	-18.10 [-71.17, 34.97]	2013
Liu	107.8	55.9	48	132.7	51.5	48	8.9%	-24.90 [-48.41, -1.39]	2013
Zhao	38	17.9	30	85	40.8	30	9.4%	-47.00 [-62.94, -31.06]	2013
Subtotal (95%CI)			1107			2310	100.0%	-37.19 [-60.16, -14.23]	
Heterogeneity: Tau^2 = 1393.22; χ^2 = 195.54, df = 11 (P < 0.00001); I^2 = 94%									
Test for overall effect: Z = 3.17 (P = 0.002)									
1.1.3 Number of retrieved LNs									
Pugliese	25	4.5	16	31	8	48	9.5%	-6.00 [-9.16, -2.84]	2010
Kim	41.1	10.9	16	37.4	10	11	3.3%	3.70 [-4.27, 11.67]	2010
Woo	39	15.2	236	37.4	14.2	591	11.5%	1.60 [-0.65, 3.85]	2011
Kim	40.2	15.5	436	37.6	13.9	861	12.6%	2.60 [0.87, 4.33]	2012
Huang	32	13.7	39	26	12.4	64	5.9%	6.00 [0.74, 11.26]	2012
Eom	30.2	9.9	30	33.4	9.9	62	7.3%	-3.20 [-7.52, 1.12]	2012
Zhang	23.1	5.4	97	20	4.3	70	13.1%	3.10 [1.63, 4.57]	2012
Yoon	42.8	12.7	36	39.4	13.4	65	5.8%	3.40 [-1.87, 8.67]	2012
Hyun	32.8	123.8	38	32.6	13.3	83	5.9%	0.20 [-5.04, 5.44]	2013
Noshiro	44	19	21	40	15	160	3.0%	4.00 [-4.45, 12.45]	2013
Liu	34.7	5.2	48	32.5	4.7	48	12.1%	2.20 [0.22, 4.18]	2013
Zhao	30.8	6.3	30	28.1	5.1	30	10.1%	2.70 [-0.20, 5.60]	2013
Subtotal (95%CI)			1043			2093	100.0%	1.46 [-0.19, 3.10]	
Heterogeneity: Tau^2 = 4.80; χ^2 = 36.48, df = 11 (P = 0.0001); I^2 = 70%									
Test for overall effect: Z = 1.74 (P = 0.08)									

Mean difference IV, Random, 95%CI

-100 -50 0 50 100

Favours RAG Favours LAG

Test for subgroup differences: χ^2 = 48.00, df = 2 (P < 0.00001); I^2 = 95.8%

B

Study or Subgroup	RAG Mean	RAG Total	LAG Mean	LAG Total	Weight	Odds ratio M-H, Fixed, 95%CI	Year
2.1.1 Overall complications							
Pugliese	1	16	6	48	1.9%	0.47 [0.05, 4.20]	2010
Kim	0	16	1	11	1.1%	0.21 [0.01, 5.71]	2010
Woo	26	236	81	591	27.4%	0.78 [0.49, 1.25]	2011
Yoon	6	36	10	65	4.0%	1.10 [0.365, 3.32]	2012
Zhang	6	97	5	70	3.6%	0.86 [0.25, 2.93]	2012
Kang	14	100	29	282	8.7%	1.42 [0.72, 2.81]	2012
Kim	44	436	81	861	32.6%	1.08 [0.73, 1.59]	2012
Huang	6	39	10	64	4.3%	0.98 [0.33, 2.95]	2012
Eom	4	30	4	62	1.5%	2.23 [0.52, 9.62]	2012
Zhao	2	30	5	30	3.1%	0.36 [0.06, 2.01]	2013
Hyun	18	38	32	83	7.0%	1.43 [0.66, 3.11]	2013

Odds ratio M-H, Fixed, 95%CI

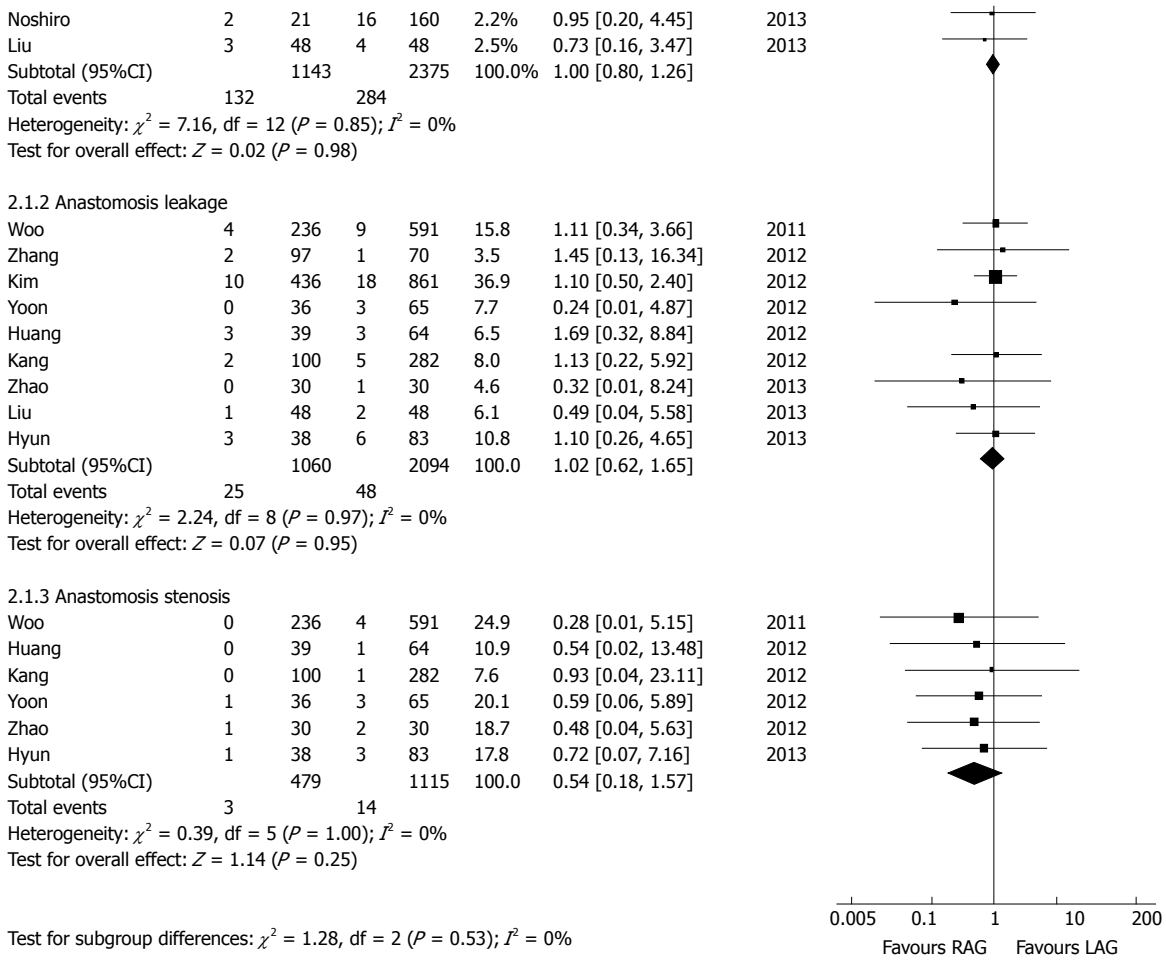


Figure 3 Forest plot comparing intraoperative (A) and postoperative (B) outcomes between robot-assisted gastrectomy and laparoscopy-assisted gastrectomy. RAG: Robot-assisted gastrectomy; LAG: Laparoscopy-assisted gastrectomy.

(I^2 = 0%). A fixed-effect model was selected. No significant difference between RAG and LAG group was found in the comparison of the incidences of overall complications (11.5% *vs* 12.0%, OR = 1.00, 95%CI: 0.80-1.26, P = 0.98).

The incidences of anastomosis leakages were reported in 9 studies[1,3,6,9,11,12,15,18,20]. In total, 1060 patients were treated with RAG and 2094 patients received LAG. No statistical heterogeneity was found (I^2 = 0%). There was no significant difference between RAG and LAG group in the comparison of the incidences of anastomosis leakages (2.4% *vs* 2.3%, OR = 1.02, 95%CI: 0.62-1.65, P = 0.95).

Six studies[1,9,12,15,18,20] involving 1594 subjects were included in the analysis of anastomosis stenosis rates. No statistical heterogeneity was found in this analysis (I^2 = 0%) and a fixed-effect model was selected. The results didn' t indicate statistical difference between the two groups in the comparison of the anastomosis stenosis rates (0.6% *vs* 1.3%, OR = 0.54, 95%CI: 0.18-1.57, P = 0.25).

Comparison of other clinical outcomes between RAG and LAG group

There were 7 studies[1,6,13,14,17-19] which reported of numbers of patients in LAG or RAG group who converted to open gastrectomy (Figure 4A). No statistical heterogeneity was found (I^2 = 0%). There was no statistical difference in the comparison of conversion to open gastrectomy between the two groups (OR = 2.10, 95%CI: 0.47-9.40, P = 0.33).

The incidences of bleeding events after operation were reported in 6 studies[1,3,9,11,12,18], involving 2897 subjects. No statistical heterogeneity was found in this analysis (I^2 = 29%) and a fixed-effect model was selected (Figure 4B). The results indicated that there was no significant difference in the comparison of bleeding rates (1.6% *vs* 1.4%, OR = 1.19, 95%CI: 0.63-2.24, P = 0.59).

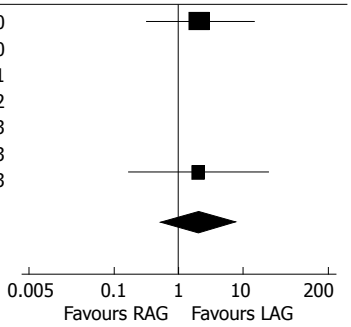
Seven studies[1,3,9,11,12,15,20] involving 2937 patients were included in the analysis of intestinal obstruction (Figure 4C). A fixed-effect model was selected (I^2 = 0%). No significant difference was found in the comparison of intestinal obstruction rates (1.1% *vs* 0.9%, OR = 1.17, 95%CI: 0.56-2.48, P = 0.67).

Eleven studies[1,3,6,9,12-15,17,19,20] had reported the mortalities (Figure 4D). There were 3230 subjects included (1008 in RAG group and 2222 in LAG group). No statistical heterogeneity was found (I^2 = 0%). The results indicated no significant difference of mortality between the two

A

Study or Subgroup	RAG Mean	RAG Total	LAG Mean	LAG Total	Weight	Odds ratio M-H, Fixed, 95%CI	Year
Pugliese	2	16	3	48	57.8%	2.14 [0.32, 14.14]	2010
Kim	0	16	0	11		Not estimable	2010
Woo	0	236	0	591		Not estimable	2011
Eom	0	30	0	62		Not estimable	2012
Noshiro	0	21	0	160		Not estimable	2013
Hyun	0	38	0	83		Not estimable	2013
Liu	2	48	1	48	42.2%	2.04 [0.18, 23.32]	2013
Subtotal (95%CI)		405		1003	100.0%	2.10 [0.47, 9.40]	
Total events	4		4				

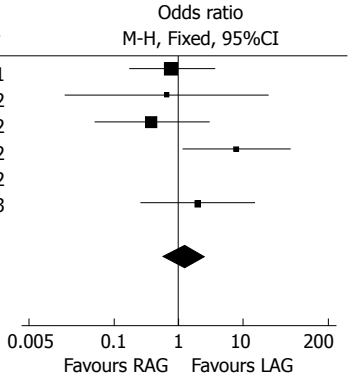
Heterogeneity: χ^2 = 0.00, df = 1 (P = 0.98); I^2 = 0%
Test for overall effect: Z = 0.97 (P = 0.33)



B

Study or Subgroup	RAG Mean	RAG Total	LAG Mean	LAG Total	Weight	Odds ratio M-H, Fixed, 95%CI	Year
Woo	4	236	12	591	39.5%	0.83 [0.27, 2.61]	2011
Zhang	1	97	1	70	6.7%	0.72 [0.04, 11.69]	2012
Kim	2	436	8	861	31.4%	0.49 [0.10, 2.32]	2012
Kang	5	100	3	282	8.8%	4.89 [1.15, 20.87]	2012
Huang	0	39	0	64		Not estimable	2012
Hyun	3	38	4	83	13.6%	1.69 [0.36, 7.97]	2013
Subtotal (95%CI)		946		1951	100.0%	1.19 [0.63, 2.24]	
Total events	15		28				

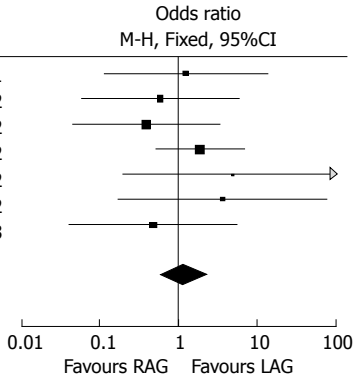
Heterogeneity: χ^2 = 5.60, df = 4 (P = 0.23); I^2 = 29%
Test for overall effect: Z = 0.54 (P = 0.59)



C

Study or Subgroup	RAG Mean	RAG Total	LAG Mean	LAG Total	Weight	Odds ratio M-H, Fixed, 95%CI	Year
Woo	1	236	2	591	9.1%	1.25 [0.11, 13.89]	2011
Yoon	1	36	3	65	16.7	0.59 [0.06, 5.89]	2012
Kim	1	436	5	861	26.9	0.39 [0.05, 3.38]	2012
Kang	4	100	6	282	24.2	1.92 [0.53, 6.94]	2012
Huang	1	39	0	64	2.9	5.03 [0.20, 126.47]	2012
Zhang	2	97	0	70	4.5	3.69 [0.17, 78.09]	2012
Zhao	1	30	2	30	15.5%	0.48 [0.04, 5.63]	2013
Subtotal (95%CI)		974		1963	100.0%	1.17 [0.56, 2.48]	
Total events	11		18				

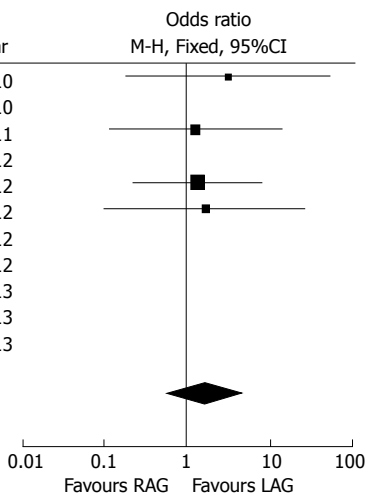
Heterogeneity: χ^2 = 3.72, df = 6 (P = 0.71); I^2 = 0%
Test for overall effect: Z = 0.42 (P = 0.67)



D

Study or Subgroup	RAG Mean	RAG Total	LAG Mean	LAG Total	Weight	Odds ratio M-H, Fixed, 95%CI	Year
Pugliese	1	16	1	48	10.8%	3.13 [0.18, 53.21]	2010
Kim	0	16	0	11		Not estimable	2010
Woo	1	236	2	591	26.1%	1.25 [0.11, 13.89]	2011
Yoon	0	36	0	65		Not estimable	2012
Kim	2	436	3	861	46.1%	1.32 [0.22, 7.92]	2012
Huang	1	39	1	64	17.0%	1.66 [0.10, 27.29]	2012
Kang	0	100	0	282		Not estimable	2012
Eom	0	30	0	62		Not estimable	2012
Zhao	0	30	0	30		Not estimable	2013
Liu	0	48	0	48		Not estimable	2013
Noshiro	0	21	0	160		Not estimable	2013
Subtotal (95%CI)		405		1003	100.0%	2.10 [0.47, 9.40]	
Total events	4		4				

Heterogeneity: χ^2 = 0.30, df = 1 (P = 0.96); I^2 = 0%
Test for overall effect: Z = 0.75 (P = 0.45)



E

Study or Subgroup	RAG Mean	RAG SD	RAG Total	LAG Mean	LAG SD	LAG Total	Weight	Mean difference IV, Random, 95%CI	Year
Pugliese	10	3	16	10	2.6	48	7.0%	0.00 [-1.64, 1.64]	2010
Kim	5.1	0.3	16	6.5	0.8	11	13.5%	-1.40 [-1.90, -0.90]	2010
Woo	7.7	17.2	236	7	5.7	591	4.8%	0.70 [-1.54, 2.94]	2011

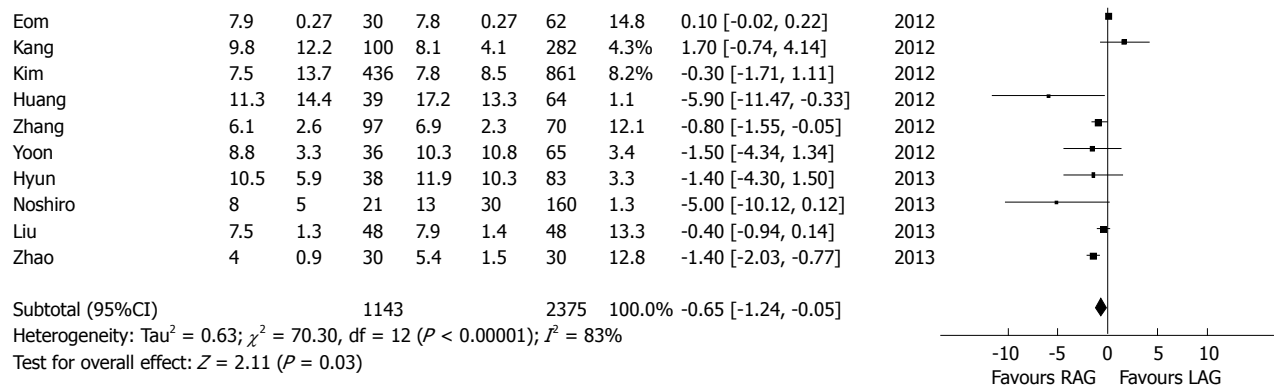


Figure 4 Forest plot. A: Comparing conversion to open gastrectomy between RAG and LAG; B: Comparing bleeding events between RAG and LAG; C: Comparing intestinal obstruction between RAG and LAG; D: Comparing mortality between RAG and LAG; E: Comparing postoperative hospital stay between RAG and LAG. RAG: Robot-assisted gastrectomy; LAG: Laparoscopy-assisted gastrectomy.

groups (0.5% *vs* 0.3%, OR = 1.55, 95%CI: 0.49-4.94, P = 0.45).

Thirteen studies[1,3,6,9,11-15,17-20] involving 3518 patients were included in the analysis of postoperative hospital stay (Figure 4E). A random-effect model was selected (I^2 = 83%). The RAG group had a shorter mean postoperative hospital stay than the LAG group [WMD = -0.65 d, 95%CI: -1.24-(-0.05), P = 0.03].

DISCUSSION

The findings from our meta-analysis suggest that RAG is effective and safe for gastric cancer compared to LAG. Overall, combining the available data RAG was associated with longer operative time, less blood loss and shorter postoperative hospital stay than LAG. Moreover, there was no significant difference in mortality, conversion, overall complications, postoperative bleeding events, intestinal obstruction, anastomosis leakage and anastomosis stenosis rates. There was also no significant difference in the numbers of retrieved LNs during the operation between RAG and LAG.

Previous studies have reported the application of RAG for the treatment of gastric cancer. Yoon *et al*[15] included 36 patients who underwent RAG and 65 patients who underwent LAG at the National Cancer Center in South Korea. The operative data, postoperative morbidity, and pathologic data were analyzed. They found that the mean postoperative hospital stay was 8.8 ± 3.3 d in the RAG group and 10.3 ± 10.8 d in the LATG group (P = 0.416). The mean operative time was 305.8 ± 115.8 min in the RAG group and 210.2 ± 57.7 min in the LAG group (P < 0.001). No significant differences were found in the comparison of mean number of dissected LNs and incidence of postoperative complications. Some other studies[6,18,20] and meta-analysis[21,22] have reported similar results. However, these studies have limited samples and most of them were retrospective. Therefore, we pooled relevant studies and conducted a meta-analysis to compare the short-term clinical outcomes of RAG with LAG systematically. Finally, 13 studies involving 3815 subjects were included. The quality of these studies was relatively high because their NOS scores ranged from 6 to 8. There was a significant heterogeneity among the included studies in the analysis of intraoperative outcomes. This may be explained by the differences in the stage of gastric cancer, resection scope, operation skill, gastric resection approach, extension of LN dissection and the standards for discharge among the studies. Further, according to the funnel plot, the publication bias was acceptable.

According to the results of our analysis, the operative time is much longer in RAG group. It may be related to the increased set-up time to position and the inexpert skill of surgeons. RAG was also associated with less estimated blood loss compared with LAG. It's more convenient for hemostatic treatment because RAG provides an excellent and stable visualization of the operative field[9,11]. Even though the mean postoperative hospital stay is 0.65 d shorter in RAG, we think that it is of little practical significance because it's too short. Moreover, there are no differences between RAG and LAG in the comparison of retrieved LNs and postoperative outcomes. Briefly, the results in the current study indicate that RAG is as safe and effective as LAG in the treatment of gastric cancer.

However, the costs of RAG are much higher than those of LAG. The mean cost of RAG is about $6000 to $11400 for gastric cancer, while only $2000 to $6000 in LAG group[9,11,17]. Consequently, before surgeons and patients make the decision, patients' economic condition should also be taken into consideration.

However, this study had some potential limitations. Firstly, there might be a certain degree of language bias because only publications in Chinese or English were searched in the databases. And then, the number of included subjects was relatively few in this study, which may lead to low statistical power. Moreover, most of them were retrospective designed and long-term outcomes were not reported. More high-quality randomized clinical studies are deserved to better evaluate both short and long-term outcomes of RAG. Further, the end points predetermined in the included studies were different. We can only partly extract the information from these

studies. As for study population, most participants were Asian. Studies in Western countries were relatively rare. Lastly, the differences in population characteristics (stage of gastric cancer, age, gender ratio, diabetes mellitus, hypertension, *etc*.), device and the duration of follow up among the included studies may also lead to a bias in a certain degree.

In conclusion, the synthesis of available evidence indicates that RAG is effective and safe in the treatment of gastric cancer. RAG is a promising alternative to laparoscopic surgery. Long-term randomized controlled studies with large scales and improved designs are needed to further evaluate the long-term outcomes.

ACKNOWLEDGEMENTS

The authors wish to thank Dan Huang for assistance with references.

COMMENTS

Background

Gastric cancer is the fourth leading cancer and second leading cause of cancer death in the world. Robot-assisted gastrectomy (RAG) is a new approach for gastric cancer and is reported to be safe and efficient. However, most of the studies were case control studies and their sample sizes were rather small.

Research frontiers

The purpose of this study was to perform a meta-analysis to compare the short-term clinical outcomes of RAG with laparoscopy-assisted gastrectomy (LAG) in gastric cancer patients systematically.

Innovations and breakthroughs

In the present study, the results indicate that RAG is associated with longer operative time [weighted mean difference (WMD) = 46.26 min, $P < 0.00001$], less blood loss (WMD = -37.19 mL, $P = 0.002$) and shorter postoperative hospital stay (WMD = -0.65 d, $P = 0.03$) than LAG. So far, this is a meta-analysis with most included studies and largest number of included subjects.

Applications

RAG is effective and safe in the treatment of gastric cancer. RAG is a promising alternative to laparoscopic surgery.

Terminology

RAG: Robot-assisted gastrectomy. Robot-assisted surgery is a kind of minimally invasive approaches. It can be used in gastrectomy and may enhance visualization of both the operative field and precision of the necessary techniques. The most popular one is da Vinci computer-enhanced surgical system.

Peer review

Lin *et al* in their manuscript present an interesting meta-analysis. This is a well done study.

REFERENCES

1 **Woo Y**, Hyung WJ, Pak KH, Inaba K, Obama K, Choi SH, Noh SH. Robotic gastrectomy as an oncologically sound alternative to laparoscopic resections for the treatment of early-stage gastric cancers. *Arch Surg* 2011; **146**: 1086-1092 [PMID: 21576595 DOI: 10.1001/archsurg.2011.114]

2 **Hohenberger P**, Gretschel S. Gastric cancer. *Lancet* 2003; **362**: 305-315 [PMID: 12892963 DOI: 10.1016/S0140-6736(03)13975-X]

3 **Kim KM**, An JY, Kim HI, Cheong JH, Hyung WJ, Noh SH. Major early complications following open, laparoscopic and robotic gastrectomy. *Br J Surg* 2012; **99**: 1681-1687 [PMID: 23034831 DOI: 10.1002/bjs.8924]

4 **Shehzad K**, Mohiuddin K, Nizami S, Sharma H, Khan IM, Memon B, Memon MA. Current status of minimal access surgery for gastric cancer. *Surg Oncol* 2007; **16**: 85-98 [PMID: 17560103 DOI: 10.1016/j.suronc.2007.04.012]

5 **Qiu J**, Pankaj P, Jiang H, Zeng Y, Wu H. Laparoscopy versus open distal gastrectomy for advanced gastric cancer: a systematic review and meta-analysis. *Surg Laparosc Endosc Percutan Tech* 2013; **23**: 1-7 [PMID: 23386142 DOI: 10.1097/SLE.0b013e3182747af7]

6 **Liu C**, Tang B, He Y, Shi Y, Zeng D, Luo H, Zhao Y, Qian F, Yu P. Surgical short-term outcomes of robotic gastrectomy vs laparoscopic gastrectomy: a case-control study. *Disan Junyi Daxue Xuebao* 2013, **35**: 1164-1166

7 **Kitano S**, Iso Y, Moriyama M, Sugimachi K. Laparoscopy-assisted Billroth I gastrectomy. *Surg Laparosc Endosc* 1994; **4**: 146-148 [PMID: 8180768]

8 **Etoh T**, Inomata M, Shiraishi N, Kitano S. Minimally invasive approaches for gastric cancer-Japanese experiences. *J Surg Oncol* 2013; **107**: 282-288 [PMID: 22504947 DOI: 10.1002/jso.23128]

9 **Huang KH**, Lan YT, Fang WL, Chen JH, Lo SS, Hsieh MC, Li AF, Chiou SH, Wu CW. Initial experience of robotic gastrectomy and comparison with open and laparoscopic gastrectomy for gastric cancer. *J Gastrointest Surg* 2012; **16**: 1303-1310 [PMID: 22450954 DOI: 10.1007/s11605-012-1874-x]

10 **Hashizume M**, Shimada M, Tomikawa M, Ikeda Y, Takahashi I, Abe R, Koga F, Gotoh N, Konishi K, Maehara S, Sugimachi K. Early experiences of endoscopic procedures in general surgery assisted by a computer-enhanced surgical system. *Surg Endosc* 2002; **16**: 1187-1191 [PMID: 11984681 DOI: 10.1007/s004640080154]

11 **Zhang X**, Jiang Z, Kun Z. Comparative study on clinical efficacy of robot-assisted and laparoscopic gastrectomy for gastric cancer. *Zhonhua Weichang Waike Zazhi* 2012, **15**: 804-806 [DOI: 10.3760/cma.j.issn.1671-0274.2012. 08.016]

12 **Kang BH**, Xuan Y, Hur H, Ahn CW, Cho YK, Han SU. Comparison of Surgical Outcomes between Robotic and Laparoscopic Gastrectomy for Gastric Cancer: The Learning Curve of Robotic Surgery. *J Gastric Cancer* 2012; **12**: 156-163 [PMID: 23094227 DOI: 10.5230/jgc.2012.12.3.156]

13 **Kim MC**, Heo GU, Jung GJ. Robotic gastrectomy for gastric cancer: surgical techniques and clinical merits. *Surg Endosc* 2010; **24**: 610-615 [PMID: 19688399 DOI: 10.1007/s00464-009-0618-9]

14 **Pugliese R**, Maggioni D, Sansonna F, Costanzi A, Ferrari GC, Di Lernia S, Magistro C, De Martini P, Pugliese F. Subtotal gastrectomy with D2 dissection by minimally invasive surgery for distal adenocarcinoma of the stomach: results and 5-year survival. *Surg Endosc* 2010; **24**: 2594-2602 [PMID: 20414682 DOI: 10.1007/s00464-010-1014-1]

15 **Yoon HM**, Kim YW, Lee JH, Ryu KW, Eom BW, Park JY, Choi IJ, Kim CG, Lee JY, Cho SJ, Rho JY. Robot-assisted total gastrectomy is comparable with laparoscopically assisted total gastrectomy for early gastric cancer. *Surg Endosc* 2012; **26**: 1377-1381 [PMID: 22083338 DOI: 10.1007/s00464-011-2043-0]

16 **Mao L**, Jian C, Changzhi L, Dan H, Suihua H, Wenyi T, Wei W. Cytochrome CYP2C19 polymorphism and risk of adverse clinical events in clopidogrel-treated patients: a meta-analysis based on 23,035 subjects. *Arch Cardiovasc Dis* 2013; **106**: 517-527 [PMID: 24080325 DOI: 10.1016/j.acvd.2013.06.055]

17 **Eom BW**, Yoon HM, Ryu KW, Lee JH, Cho SJ, Lee JY, Kim CG, Choi IJ, Lee JS, Kook MC, Rhee JY, Park SR, Kim YW. Comparison of surgical performance and short-term clinical outcomes between laparoscopic and robotic surgery in distal gastric cancer. *Eur J Surg Oncol* 2012; **38**: 57-63 [PMID: 21945625 DOI: 10.1016/j.ejso.2011.09.006]

18 **Hyun MH**, Lee CH, Kwon YJ, Cho SI, Jang YJ, Kim DH, Kim JH, Park SH, Mok YJ, Park SS. Robot versus laparoscopic gastrectomy for cancer by an experienced surgeon:

comparisons of surgery, complications, and surgical stress. *Ann Surg Oncol* 2013; **20**: 1258-1265 [PMID: 23080320 DOI: 10.1245/s10434-012-2679-6]

19 **Noshiro H**, Ikeda O, Urata M. Robotically-enhanced surgical anatomy enables surgeons to perform distal gastrectomy for gastric cancer using electric cautery devices alone. *Surg Endosc* 2014; **28**: 1180-1187 [PMID: 24202713]

20 **Zhao K**, Pan H, Wang G, Li M, Ruan H, Jiang Z, Li N, Li J. Contrast study of short-term effect between the Da Vinci surgical robot and laparoscopic technology in patients after distal gastric cancer surgery. *Zhonguo Shiyong Waike Zazhi* 2013, **33**: 325-327

21 **Marano A**, Hyung WJ. Robotic gastrectomy: the current state of the art. *J Gastric Cancer* 2012; **12**: 63-72 [PMID: 22792518 DOI: 10.5230/jgc.2013.13.3.136]

22 **Xiong J**, Nunes QM, Tan C, Ke N, Chen Y, Hu W, Liu X, Mai G. Comparison of short-term clinical outcomes between robotic and laparoscopic gastrectomy for gastric cancer: a meta-analysis of 2495 patients. *J Laparoendosc Adv Surg Tech A* 2013; **23**: 965-976 [PMID: 24093968 DOI: 10.1089/lap.2013.0279]

P- Reviewer: Bener A **S- Editor:** Wen LL **L- Editor**: A
E- Editor: Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.107

World J Meta-Anal 2014 August 26; 2(3): 107-126
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

META-ANALYSIS

Comparison of functional outcomes after retropubic, laparoscopic and robot-assisted radical prostatectomy: A meta-analysis

Ming-Jun Shi, Jie Yang, Xiang-Yu Meng, Sheng Li, Tao Liu, Zhi-Hai Fang, Rui Cao, Xing-Huan Wang

Ming-Jun Shi, Sheng Li, Tao Liu, Zhi-Hai Fang, Rui Cao, Xing-Huan Wang, Department of Urology, Zhongnan Hospital, Wuhan University, Wuhan 430071, Hubei Province, China
Jie Yang, Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China
Xiang-Yu Meng, Department of Hematology, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China
Author contributions: Shi MJ and Yang J contributed equally to this work; Shi MJ, Yang J, Li S and Meng XY performed the literature search, study selection and data extraction; Shi MJ, Wang XH, Liu T and Fang ZH analyzed the results; Shi MJ and Cao R did the statistical analysis; Shi MJ, Yang J and Meng XY drafted the manuscript; all authors provided input into the development of the manuscript and approved the final manuscript.
Supported by National Natural Science Foundation of China, No. 81172734
Correspondence to: Xing-Huan Wang, Professor, Department of Urology, Zhongnan Hospital, Wuhan University, 169 Donghu Road, Wuchang, Wuhan 430071, Hubei Province, China. wangxinghuan1965@gmail.com
Telephone: +86-27-67813104 Fax: +86-27-67813104
Received: March 26, 2014 Revised: May 10, 2014
Accepted: June 18, 2014
Published online: August 26, 2014

Abstract

AIM: To assess the 6-mo and 12-mo functional outcomes after retropubic, laparoscopic and robot-assisted laparoscopic radical prostatectomy retropubic radical prostatectomy (RRP) laparoscopic radical prostatectomy (LRP); robot-assisted laparoscopic prostatectomy (RARP).

METHODS: A literature search was conducted using the PubMed, EMBASE, The Cochrane Library and the Web of Knowledge databases updated to March, 2014 for relevant published studies. After data extraction and quality assessment *via* the Newcastle-Ottawa Scale or the Cochrane collaboration's tool for assessing risk of bias, meta-analysis was performed using RevMan 5.1. Either a random-effects model or a fixed-effects model was used. Potential publication bias was assessed using visual inspection of the funnel plots, and verified by the Egger linear regression test.

RESULTS: Thirty-seven studies were identified in total: 14 articles comparing LRP with RRP, 12 articles comparing RARP with RRP, and 11 articles comparing RARP with LRP. For urinary continence, a statistically significant advantage was observed in RARP compared with LRP or RRP both at 6 mo [odds ratio (OR) = 1.93; $P < 0.01$, OR = 2.23; $P < 0.05$, respectively] and 12 mo (OR = 1.47; $P < 0.01$, OR = 2.93; $P < 0.01$, respectively) postoperatively. The continence recovery rates after LRP and RRP, with obvious heterogeneity (6-mo: I^2 = 74%; 12-mo: I^2 = 75%), were equivalent (6-mo: $P = 0.52$; 12-mo: $P = 0.75$). In terms of potency recovery, for the first time, we ranked the three surgical approaches into a superiority level: RARP > LRP > RRP, with a statistically significant difference at 12 mo [RARP *vs* LRP (OR = 1.99; $P < 0.01$); RARP *vs* RRP (OR = 2.66; $P < 0.01$); LRP *vs* RRP (OR = 1.34; $P < 0.05$)], respectively. Meta-regression and subgroup analyses according to adjustment of the age, body mass index, prostate volume, Gleason score or prostate-specific antigen did not vary significantly.

CONCLUSION: Current evidence suggests that minimally invasive approaches (RARP or LRP) are effective procedures for functional recovery. However, more high-quality randomized control trials investigating the long-term functional outcomes are needed.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Prostate cancer; Radical prostatectomy; Urinary continence; Potency; Meta-analysis

Core tip: This review directly compared the functional

outcomes after retropubic, laparoscopic and robot-assisted radical prostatectomy, both at 6-mo and 12-mo follow-up. Compared with the previous meta-analysis which reported a comparable potency recovery of robot-assisted laparoscopic prostatectomy (RARP) *vs* laparoscopic radical prostatectomy (LRP), our review obviously included more studies and ranked the three techniques into a superiority level: RARP > LRP > RRP (retropubic radical prostatectomy). In addition, we performed a quality assessment of the studies, separated evaluation of randomized control trials (RCTs) and non-RCTs, and subgroup analyses or meta-regression as a supplement, thus the risk of methodological bias was reduced considerably.

Shi MJ, Yang J, Meng XY, Li S, Liu T, Fang ZH, Cao R, Wang XH. Comparison of functional outcomes after retropubic, laparoscopic and robot-assisted radical prostatectomy: A meta-analysis. *World J Meta-Anal* 2014; 2(3): 107-126 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/107.htm DOI: http://dx.doi.org/10.13105/wjma.v2.i3.107

INTRODUCTION

Prostate cancer (PCa) is now recognized as one of the most important medical problems in the male population. PCa accounted for almost 28% (238590) of all newly diagnosed cancer cases and it is the second cause of male cancer death (after lung cancer) in the United States, while in Europe, data show an incidence rate of 22.8% and a mortality of 9.5%[1,2]. With combined application of prostate-specific antigen (PSA) test and prostate biopsy, the percentage of early diagnosed PCa cases has increased .

Radical prostatectomy (RP) is one of the recommended standard treatments for clinically localized prostate cancer (cT1-cT2) patients with a life expectancy of more than 10 years[3]. The retropubic radical prostatectomy (RRP), since its first introduction by Walsh *et al*[4] in 1982, soon became the gold standard and the most widely used treatment for patients with localized PCa[5]. Recently, we have witnessed the emergence of laparoscopic radical prostatectomy (LRP) and robot-assisted laparoscopic prostatectomy (RARP). Facing all these surgical options, both patients and surgeons hesitate when a best treatment choice should be made. Although several experts have demonstrated that when compared with RRP, LRP and RARP have obvious advantages such as lower blood loss, less need for transfusion and shorter hospital-stay[6,7], but the lack of high-quality evidence and randomized control trials (RCTs) available precluded us from proving the superiority of any surgical approaches in terms of postoperative functional outcomes.

The increase in life expectancy in patients with localized PCa has made the post-treatment quality of life a key issue for PCa survivors, but some negative functional outcomes such as urinary incontinence and erectile dysfunction make the health-related quality of life worse. Relevant comparative studies showed 12-mo urinary continence recovery rates ranging from 47% to 96%, 48% to 97% and 88% to 97% after RRP, LRP and RARP, respectively. The previously published surgical series showed 12-mo potency recovery rates ranging from 39% to 72%, 41% to 81% and 61% to 87% after RRP, LRP and RARP, respectively. This apparent difference can be attributed to multiple definitions of urinary continence and potency, variations in population baseline, differences among surgical techniques and diverged data collection as well. In comparison with the only two meta-analyses evaluating functional outcomes after different surgical approaches, reported by the same author Ficarra *et al*[8,9] in August 2011, obviously our review included more studies and excluded two studies[10,11] which appeared to be ineligible since the presence of preoperative adjuvant hormonal therapy. Moreover, powerful quality assessment tools were utilized in this initial comparison of three key techniques (RRP, LRP and RARP) both at 6-mo and 12-mo follow-up.

MATERIALS AND METHODS

Literature search

A literature search of the following databases was performed: the PubMed, EMBASE, The Cochrane Library and the Web of Knowledge databases up to March, 2014. We used the following limits: humans, gender (male), and no restriction for languages. For each database, the same search terms "radical prostatectomy", "urinary continence", "incontinence", "potency" and "erectile function" were used. Although we also paid attention to two unpublished studies (gray literature) with relevant outcomes reported on the website "Clinical Trials.gov" and tried to contact the experts by e-mail, there has been no response so far, and therefore in this review only published papers were included.

Study selection

Our study followed the preferred reporting items for meta-analyses of observational studies in epidemiology (MOOSE) statement[12]. The inclusion criteria were as follows: (1) patient characteristics: localized PCa (cT1-cT2); comparable baseline demographics; preoperatively potent and continent; no obvious comorbidities; (2) surgical techniques: only pure RRP/RARP/LRP with or without modification; (3) methodologically: all studies comparing the postoperative outcomes as RRP/LRP, RRP/RARP or LRP/RARP and including at least one of the functional results; clear definition of urinary continence and potency; and (4) population-based studies, duplicated publications and meeting abstracts were excluded.

Data extraction

All eligible records were extracted independently by two reviewers and selected according to the inclusion criteria. We extracted the details of authors and publishing date;

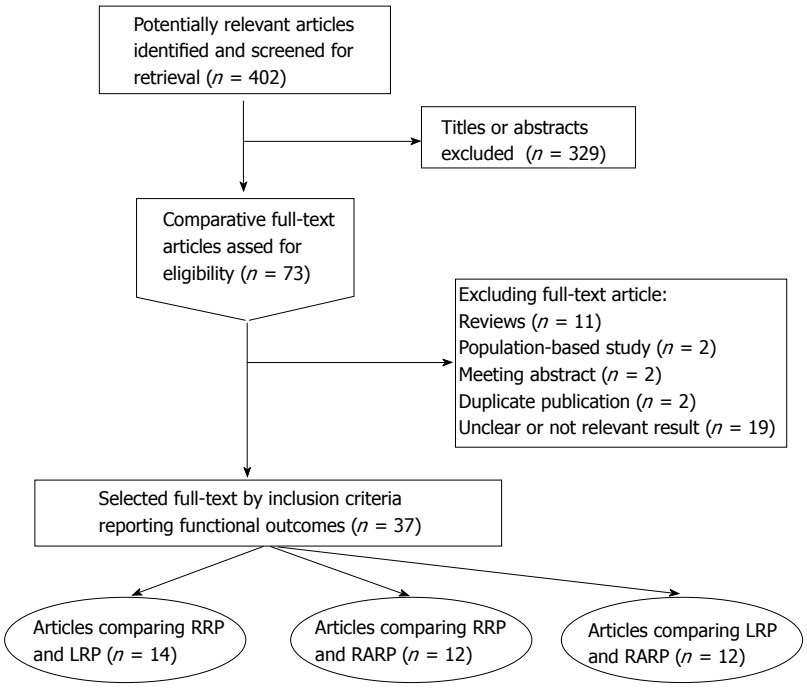


Figure 1 Flow diagram of the systematic review. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy.

surgical techniques and number of patients; study design; baseline mean age; BMI value; prostate volume; PSA level; urinary continence and potency definition; and 6- and 12-mo recovery rates of urinary continence and potency. Any uncertainties or discrepancies between the two reviewers were resolved by open discussion or consultation with the third reviewer.

Methodological quality assessment

The quality of cohort and case-control studies was assessed using the Newcastle-Ottawa quality scale (NOS) proposed by Wells *et al*[13]. This tool can be used either as a checklist or as a scale. The NOS scales were separately developed for cohort and case-control studies. Briefly, a star system was used for quality assessment of studies, and the NOS ranges from zero up to nine stars; studies were evaluated using items from three broad perspectives: selection of study groups (0-4 stars), comparability between groups (0-2 stars), and ascertainment of either the exposure or the outcome of interest (0-3 stars) for case-control or cohort studies, respectively.

The quality of each RCT was assessed using the Cochrane collaboration's tool for assessing risk of bias[14], which utilizes seven aspects: (1) details of randomization method; (2) allocation concealment; (3) blinding of participants and personnel; (4) blinding of outcome assessment; (5) incomplete outcome data; (6) selective outcome reporting; and (7) other sources of bias, to provide a qualification of risk of bias.

Statistical analysis

Statistical analyses were performed using the Cochrane Review Manager (RevMan) Version 5.1 software. Odds ratios (ORs) with 95% confidence intervals (CIs) for dichotomous variables were computed as summary statistics. According to the Higgins' I^2 statistic, a statistical heterogeneity of < 25, 25-50, and > 50% was defined as low, moderate, and high, respectively[15]. If no heterogeneity was found, a fixed-effects model using the Mantel-Haenszel method would be used[16,17]. If statistically significant heterogeneity was revealed, a random-effects model would be used[18]. The sensitivity analysis was also performed by two methods: (1) subgroup analysis, and (2) exclusion of the study accounting for the largest proportion; if no difference was detected then it could be confirmed that the outcomes were stable and reliable. The meta-regression analyses were performed by modeling on binary continence and potency outcomes, adjusting the age, BMI, prostate volume, mean Gleason score, and PSA level by using the STATA SE 12.0. For all statistical analyses, a *P*-value < 0.05 was set as the level of significance. The publication bias was examined using the funnel plot, the results of which were further verified by Egger's test[19].

RESULTS

Study identification

Figure 1 shows the flowchart of this review and summarizes the number of potential citations (Figure 1). The authors selected 73 full-text articles after a comprehensive review of 402 potential relevant citations. Among these, 14 articles compared RRP with LRP, consisting of seven prospective and seven retrospective studies[20-33]; 12 articles compared RRP with RARP, which consisted of six prospective and six retrospective studies[10,11,34-43]; 12 articles compared LRP with RARP, including two RCTs, one prospective and nine retrospective studies[39,44-54].

Quality of studies

Totally, there were 14 prospective studies and 21 retrospective studies included in this review. According to the

Table 1 Comparative studies evaluating urinary continence recovery after retropubic radical prostatectomy or laparoscopic radical prostatectomy

Quality	Case, *n*	Ref.	Country	Age (yr)	BMI (kg/m^2)	Prostate volume (mL, g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Continence definition	Data collection	Loss of follow-up (N/Y, %)	Urinary continence recovery, %(*n*)	
													6 mo	12 mo
3/2/2(H)	RRP, 70	Anastasiadis *et al*[20], 2003	France	64.8 ± 6.4	-	-	6.1 ± 1.1	11.2 ± 9.7	Prospective	0 pad	Nonvalidated questionnaire	Y, > 20%	43.3 (16/37)	77.7 (26/33)
	LRP, 230			64.1 ± 6.4			5.8 ± 1.2	10.7 ± 8.8					59.2[a] (67/113)	89.0 (94/106)
2/2/3(H)	RRP, 77	Roumeguere *et al*[21], 2003	Belgium	63.9 ± 5.5	-	42.0 ± 20.4	5.4 ± 1.5	10.5 ± 11.5	Prospective	0 pad	Interview	Y, > 20%	62.5 (40/64)	83.9 (47/56)
	LRP, 85			62.5 ± 6.0		37.3 ± 15.6	5.4 ± 1.5	8.6 ± 5.2					50.6 (37/73)	80.7 (42/52)
3/1/3(H)	RRP, 41	Remzi *et al*[22], 2005	Austria	60 ± 14	-	44 ± 18	4.7 ± 1.5	6.9 ± 4.4	Prospective	0 pad	Physician	N	-	80.3 (33/41)
	(a)tLRP, 39			61 ± 11		37 ± 16	5.1 ± 1.2	5.5 ± 3.7						84.6 (33/39)
	(b)eLRP, 41			59 ± 12		32 ± 14	5.5 ± 1.3	8.1 ± 6.1						87.8 (36/41)
3/2/3(H)	RRP, 75	Wagner *et al*[23], 2007	United States	59 ± 6.9	29 ± 4.5	-	-	8.1 ± 6.27	Prospective	0 pad	EPIC	Y, < 20%	-	47.0 (31/66)
	LRP, 75			58 ± 6.9	27 ± 3.0			6.2 ± 4.22						64.0[a] (43/67)
3/2/2(H)	RRP, 222	Touijer *et al*[24], 2008	United States	59 (54, 64)	-	-	-	5.3 (4.1, 7.3)	Prospective	0-1 safety pad	Institutional questionnaire	N	-	75.0[a] (167/222)
	LRP, 193			60 (55, 65)				5.3 (4.0, 7.5)						48.0 (93/193)
3/2/3(H)	RRP, 150	Greco *et al*[25], 2009	Italy	61.5 (49-74)	29 (25-33)	-	5 (3-7)	6.95 (3.4-10)	Prospective	0 pad	Validated questionnaire	N	76.0 (114/150)	91.0 (137/150)
	LRP, 150			60.5 (45-76)	32 (26-38)		5 (3-7)	6.3 (2.4-10)					89.3 (134/150)	97.0 (146/150)
3/2/2(H)	RRP, 102	Dahl *et al*[26], 2009	United States	59.9	-	-	-	-	Prospective	0 pad	Validated questionnaire	Y, > 20%	49.0 (38/78)	49.0 (35/72)
	LRP, 104			59.5									42.0 (31/74)	53.0 (41/78)
2/2/2(M)	RRP, 49	Egawa *et al*[27], 2003	Japan	67.0 ± 0.7	-	-	6.0 ± 0.2	8.3 ± 1.4	Retrospective	0 pad	Interview	Y, > 20%	84.1[a] (37/44)	92.9[a] (39/42)
	LRP, 34			68.0 ± 0.9			5.0 ± 0.2	6.6 ± 0.6					46.9 (15/32)	60.0 (12/20)
3/1/2(M)	RRP, 50	Artibani *et al*[28], 2003	Italy	64.28 ± 6.6	-	-	5.7 ± 1.2	11 ± 9	Retrospective	0 pad	Nonvalidated questionnaire	Y, > 20%	-	64.0 (9/14)
	LRP, 71			63.14 ± 5.8			5.8 ± 1.3	15.7 ± 17						40.0 (8/20)
4/2/2(H)	RRP, 70	Ghavamian *et al*[29], 2006	United States	57.8 ± 7.3	28.1	53.2 (19-135)	6.7 ± 1.3	9.9 ± 7.1	Retrospective	0 pad	Physician	Y, < 20%	71.4 (50/70)	87.6 (57/65)
	LRP, 70			60.8 ± 6.1	27.5	40.8 (20-114)	6.4 ± 0.8	7.6 ± 8.0					70.0 (49/70)	90.0 (63/70)
4/2/2(H)	RRP, 37	Takenaka *et al*[30], 2008	Japan	67.1 ± 6.0	23.5 ± 3.0	30.1 ± 26.9	6.9 ± 1.0	14.7 ± 11.9	Retrospective	0 pad	Nonvalidated questionnaire	N	77.0 (28/37)	91.0 (34/37)
	LRP, 109			66.1 ± 6.3	23.8 ± 2.5	32.2 ± 16.5	6.6 ± 0.7	11.0 ± 8.4					65.0 (71/109)	77.0 (84/109)
2/2/3(H)	RRP, 188	Simforoosh *et al*[31], 2009	Iran	62.1 (45-74)	-	-	-	13.6	Retrospective	0 pad	Physician	N	91.5 (172/188)	95.2 (179/188)
	LRP, 136			62.5 (45-76)				12.7					89.0 (121/136)	96.3 (131/136)
2/1/1(M)	RRP, 128	Springer *et al*[32], 2013	Germany	57.2 ± 7.4	28.3 ± 2.6	-	-	3.1 ± 1.7	Retrospective	0 pad	Validated questionnaire	N	73.4 (94/128)	86.4 (111/128)
	LRP, 125			56.8 ± 6.7	27.7 ± 3.8			3.2 ± 1.4					86.4 (108/125)[a]	96.8[a] (121/125)
3/2/2(H)	RRP, 168	Magheli *et al*[33], 2014	Germany	62.6 ± 5.4	-	58 ± 22	-	10.1 ± 11.9	Retrospective	0-1 safety pad	Validated questionnaire	Y, > 20%	-	83.2 (99/119)
	LRP, 171			62.3 ± 5.7		53 ± 20		9.2 ± 6.9						82.8 (96/116)

[a]$P < 0.05$. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; (a)tLRP: Transperitoneal laparoscopic radical; EPIC: Expanded prostate cancer index composite.

NOS scale (case-control studies) used for quality evaluation of the retrospective studies, 12 studies were in the high level (7-9 stars)[29-31,33,41-43,48,50-52,54], one study was in the low level (0-3 stars)[11], and the remaining eight studies were in the middle level (4-6 stars). As for the quality of the prospective studies, the NOS scale (cohort studies) was used, and 12 studies were in the high level (7-9 stars)[20-26,34-36,38,46], one study was in the middle level (4-6 stars)[37], and one study was in the low level (0-3 stars)[10].

The only available two RCTs were considered high quality by using the Cochrane collaboration's tool for assessing risk of bias.

Characteristics of included studies and meta-analyses on urinary continence recovery

Table 1 summarizes the results of urinary continence recovery rate between LRP and RRP. Among the 12 studies[20-33], 1427 patients treated with RRP and 1633 patients treated with LRP were included. Most of the selected studies had a very strict urinary continence definition as no pad. Only seven studies[25-27,29-32] provided the 6-mo urinary continence

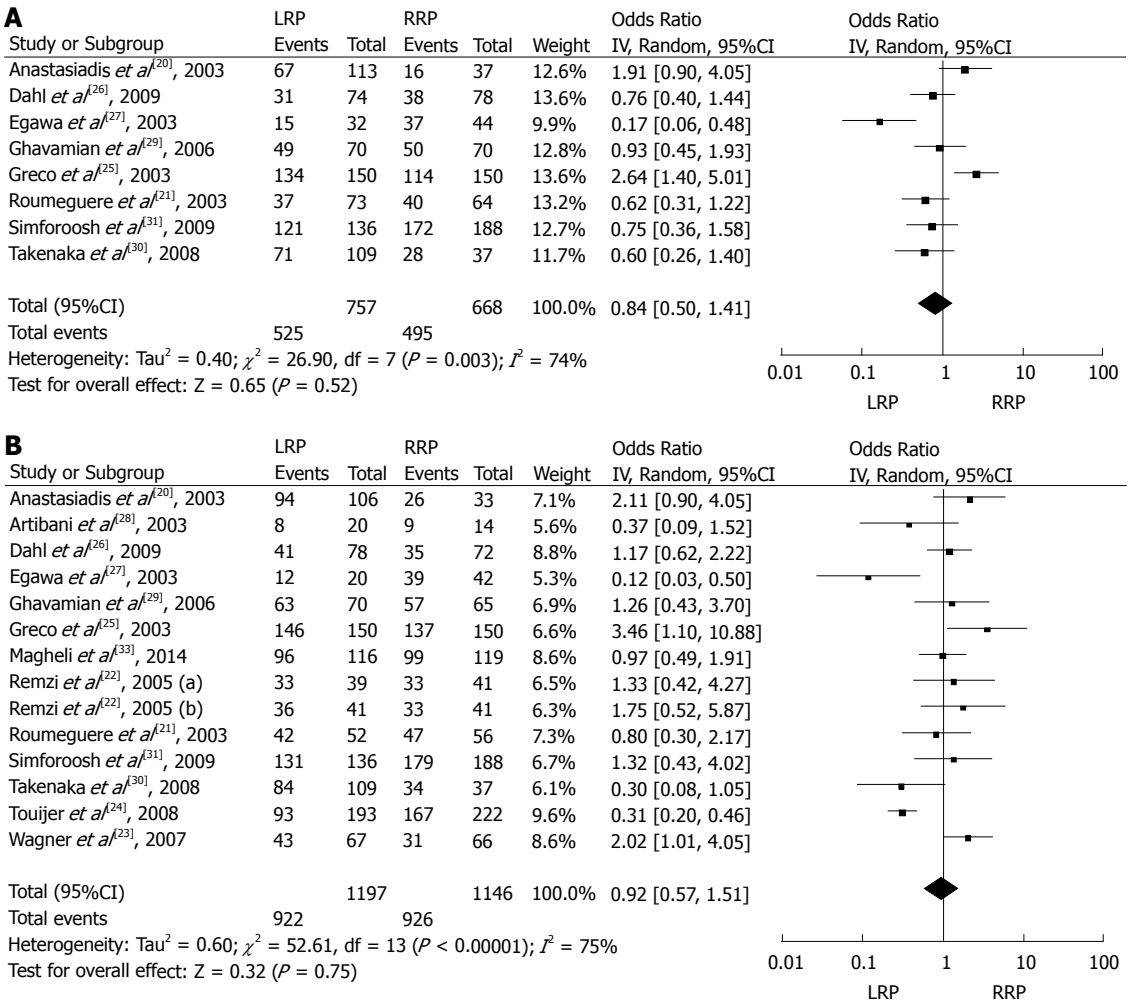


Figure 2 Forest plots and meta-analyses of laparoscopic radical prostatectomy and retropubic radical prostatectomy. A: 6-mo continence recovery; B: 12-mo continence recovery. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy.

rate. The 12-mo loss to follow-up rate was > 20% in six studies[20-21,26-28,33]. Although Springer *et al*[32] reported demonstrated a significant better outcome of LRP than ORP (96.8% *vs* 86.4%, $P < 0.05$), we did not include it in because of the preoperatively performed transurethral resection of the prostate in that report, which could potentially be an inconsistent factor among the groups. The mean urinary continence recovery rates at 6 and 12 mo were 56.6% (42%-70%) and 84.3% (48.0%-96.3%) after LRP; and 64.9% (43.3%-84.1%) and 77.8% (47.0%-95.2%) after RRP, respectively.

Six-month continence recovery after LRP and RRP: Statistically high heterogeneity ($I^2 = 74\%$, $P < 0.05$) was observed among the eight studies[20-21,25-27,29-31] included. The meta-analysis with a random-effects model showed no significant difference between LRP and RRP (OR = 0.84; 95%CI: 0.50-1.41; $P = 0.52$) (Figure 2).

Twelve-month continence recovery after LRP and RRP: Fourteen studies were included in the meta-analysis[20-31,33], and there was a statistical heterogeneity ($I^2 = 75\%$, $P < 0.05$). No significant difference was found between LRP and RRP by using a random-effects model (OR = 0.92; 95%CI: 0.57-1.51; $P = 0.75$) (Figure 2).

Table 2 summarizes the results of urinary continence recovery rate between RARP and RRP. A total of 1942 patients who received RRP and 1882 patients who received RARP were included. Half of the included studies had a very strict urinary continence definition as no pad. Only two studies[37,40] had a high loss to follow-up rate (> 20%) at 12 mo. Tewari *et al*[34] reported that the median urinary continence recovery was significantly better after RARP compared with after RRP (44 d *vs* 160 d, $P < 0.05$), and Kim *et al*[10] drew the same conclusion, while Krambeck *et al*[11] presented an opposite result in the comparison of RARP and RRP (91.8% *vs* 93.7%, respectively). However, compared with the previous meta-analysis[8], Kim *et al*[10] and Krambeck *et al*[11] results were excluded in our review because of their preoperative adjuvant hormonal therapy, which would undoubtedly cause difference.

Six-month continence recovery after RARP and RRP: Statistically significant heterogeneity was observed among the eight included studies ($I^2 = 73\%$, $P < 0.05$)[36-43],

Table 2 Comparative studies evaluating urinary continence recovery after retropubic radical prostatectomy or robot-assisted radical prostatectomy

Quality	Case, *n*	Ref.	Country	Age (yr)	BMI (kg/m^2)	Prostate volume (mL, g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Continence definition	Data collection	Loss of follow-up (N/Y,%)	Urinary continence recovery, %(*n*)	
													6 mo	12 mo
3/2/3(H)	RRP, 100 RARP, 200	Tewari *et al*[34], 2003	United States	63.1 (42.8-72) 59.9 (40-72)	27.6 (17-41) 27.7 (19-38)	48.4 (24.2-70) 58.8 (18-140)	-	7.3 (1.9-35) 6.4 (0.6-41)	Prospective	0-1 safety pad	Interview	-	Median:160 d Median:44 d[a]	
3/2/2(H)	RRP, 105 RARP, 103	Ficarra *et al*[35], 2008	Italy	65 (61-69) 61 (57-67)	26 (24-28) 26 (24-28)	40 (30-47) 37.5 (30-48)	-	6 (5-10) 6.4 (4.6-9)	Prospective	0 pad	ICIQ-UI	N	-	88.0 (92/105) 97.0[a] (100/103)
3/2/3(H)	RRP, 110 RARP, 188	Ham *et al*[36], 2008	South Korea	66.9 ± 6.0 67.3 ± 6.2	23.6 ± 1.8 23.6 ± 2.3	-	-	55.2 ± 23.7 22.3 ± 34.3	Prospective	0 pad	Validated questionnaire	N	75.5 (83/110) 87.2 (164/188)	81.8 (90/110) 92.0[a] (173/188)
3/1/2(M)	RRP, 75 RARP, 75	Di Pierro *et al*[37], 2010	Switzerland	64.3 (59.1-68.0) 62.8 (58.4-67.0)	-	-	-	7.57 (5.1-10.4) 7.72 (5.6-12.1)	Prospective	0 pad	Institutional questionnaire	Y, > 20%	83.0 (62/75) 95.0[a] (71/75)	80.0 (60/75) 89[a] (40/45)
1/1/1(L)	RRP, 235 RARP, 528	Kim *et al*[10], 2011	South Korea	66.5 ± 5.7 64.2 ± 7.3	-	18.2 ± 23.4 15.2 ± 20.2	-	14.6 ± 22.1 10.4 ± 16.0	Prospective	0 pad	Validated questionnaire	-	Median: 4.3 mo Median: 3.7 mo	
4/2/3(H)	RRP, 109 RARP, 61	Geraerts *et al*[38], 2013	Belgium	62.22 ± 6.12 61.48 ± 6.08	-	-	-	-	Prospective	24h pad test	Validated questionnaire	N	94.0 (102/109) 95.0 (58/61)	96.0 (105/109) 97.0 (59/61)
2/1/2(M)	RRP, 62 RARP, 60	Caballero *et al*[39], 2008	Spain	66.5 (62-69) 56 (56-65.25)	-	41 (30.15-52) 29.5 (23-40)	-	9.66 (7-16.6) 7 (5.7-10)	Retrospective	0 pad	Unspecified	Y, < 20%	45.9 (28/61) 60.0 (30/50)[a]	-
2/0/1(L)	RRP, 588 RARP, 294	Krambeck *et al*[11], 2008	United States	61.0 (41.0-77.0) 61.0 (38.0-76.0)	-	-	-	5.0 (0.6-39.7) 4.9 (0.5-33.5)	Retrospective	0 pad	Institutional questionnaire	Y, < 20%		93.7 (446/476) 91.8 (224/244)
3/1/2(M)	RRP, 240 RARP, 120	Rocco *et al*[40], 2009	Italy	63 (46-77) 63 (47-76)	-	-	6 (4-10) 6 (4-9)	6.7 (0.7-22.0) 6.9 (0.4-23.0)	Retrospective	0-1 safety pad	Interview	Y, > 20%	83.0 (189/229) 93.0[a] (102/110)	88.0 (191/217) 97.0[a] (77/79)
3/1/3(H)	RRP, 30 RARP, 30	Ou *et al*[41], 2009	United States	70.03 ± 6.10 67.27 ± 6.21	24.09 ± 3.28 24.22 ± 3.16	15.89 ± 14.15 16.45 ± 18.80	6.22 ± 1.62 6.13 ± 0.9	-	Retrospective	0-1 safety pad	Unspecified	N	83.3 (25/30) 96.7 (29/30)	96.6 (29/30) 100.0 (30/30)
3/2/3(H)	RRP, 176 RARP, 77	Choo *et al*[42], 2013	South Korea	67 ± 6.25 66 ± 7.75	24 ± 2.73 24 ± 2.55	42 ± 18.82 41 ± 15.77	-	7.6 ± 19.33 7.2 ± 13.19	Retrospective	0-1 safety pad	Validated questionnaire	N	92.0 (162/176) 84.0 (65/77)	96.0 (169/176) 94.0 (72/77)
3/1/3(H)	RRP, 112 RARP, 146	Son *et al*[43], 2013	South Korea	65.0 ± 6.1 65.5 ± 6.7	24.3 ± 2.4 24.5 ± 2.5	41.3 ± 30.0 45.9 ± 16.3	-	-	Retrospective	0 pad	Validated questionnaire	Y, < 20%	51.7 (49/94) 87.5[a] (107/122)	70.7 (66/94) 94.5[a] (115/122)

[a] $P < 0.05$. RRP: Retropubic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; ICIQ-UI: International consultation of incontinence questionnaire-urinary incontinence.

and the meta-analysis with a random-effects model showed a significant advantage after RARP than after RRP (OR = 2.23; 95%CI: 1.20-4.14; $P < 0.05$) (Figure 3).

Twelve-month continence recovery after RARP and RRP: Totally, eight studies were included to compare RARP and RRP[35-38,40-43]. No evidence of statistical heterogeneity was observed ($I^2 = 49\%$, $P = 0.06$) and pooled analysis with a fixed-effects model demonstrated a statistically significant advantage in favor of RARP (OR = 2.93; 95%CI: 1.99-4.32; $P < 0.01$) (Figure 3).

Table 3 summarizes the results of urinary continence recovery rate between LRP and RARP. A total of 2195 patients treated with LRP and 1940 patients treated with RARP were included. Both of the only two RCTs (high quality) revealed that the urinary continence recovery rates were significantly higher at 6 and 12 mo after RARP, in comparison with those after LRP ($P < 0.05$)[44,45]. The evidence with the largest sample size, reported by Ploussard *et al*[46], was the only prospective study of high quality (7 stars) and showed similar results with the two RCTs. Almost all of the remaining retrospective studies also indicated better outcomes after RARP. Only one study[52] had a high loss to follow-up rate (> 20%) at 12 mo. The most crucial difference between our pooled-analysis and the previous meta-analysis[8] was that the RCTs were evaluated separately with non-RCTs (NRCTs) in this review, since they had totally different data types.

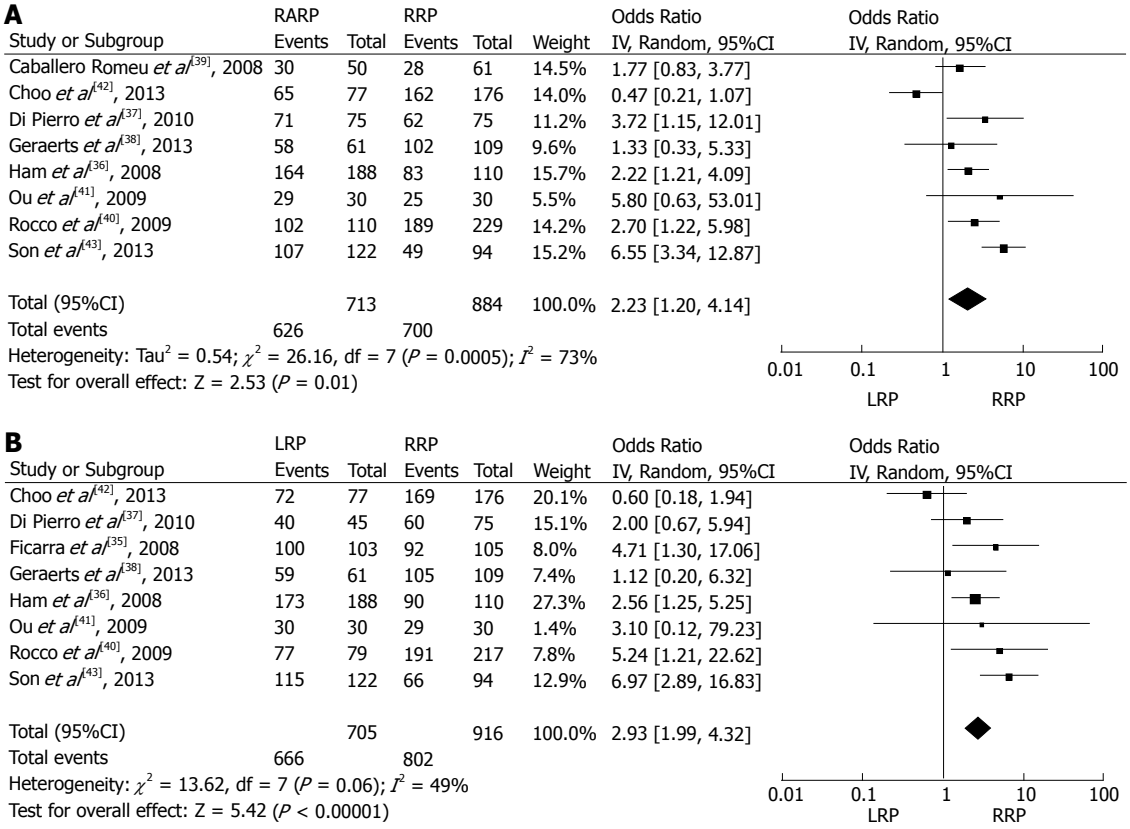


Figure 3 Forest plots and meta-analyses of robot-assisted radical prostatectomy and retropubic radical prostatectomy. A: 6-mo continence recovery; B: 12-mo continence recovery. RRP: Retropubic radical prostatectomy; RARP: Robot-assisted radical prostatectomy.

Six-month continence recovery after RARP and LRP: The two RCTs[44,45] showed no heterogeneity (I^2 = 0%, P = 0.92), and supported the advantage after RARP with a fixed-effects model (OR = 2.66; 95%CI: 1.31-5.40; P < 0.01) (Figure 4). In the cumulative analysis of 10 NRCTs[39,46-54], no heterogeneity was found (I^2 = 38%, P = 0.11), so a fixed-effects model was performed. The result also demonstrated a statistically significant advantage in favor of RARP (OR = 1.93; 95%CI: 1.67-2.23; P < 0.01) (Figure 4).

Twelve-month continence recovery after RARP and LRP: No evidence of statistical heterogeneity was observed in both of the two RCTs (I^2 = 0%, P = 0.88) or the seven NRCTs (I^2 = 0%, P = 0.44), and the pooled analyses with a fixed-effects model either for the RCTs or the NRCTs showed a statistically significant advantage in favor of RARP [(OR = 3.52; 95%CI: 1.36-9.13; P < 0.05); (OR = 1.47; 95%CI: 1.25-1.74; P < 0.01), respectively] (Figure 4).

Characteristics of included studies and meta-analyses on potency recovery

Table 4 summarizes the results of potency recovery rate between LRP and RRP. Among the 10 studies, 907 patients treated with RRP and 1004 patients treated with LRP were included. Eight of them had a very strict potency definition as erection sufficient for intercourse (ESI). The 12-mo loss to follow-up rate was > 20% in three studies[20,26,33]. Springer *et al*[32] report was not included in the meta-analysis because of its preoperative surgery. The nerve sparing (NS) procedures were not clearly mentioned in two studies[24,26], and the remaining studies either used the bilateral or unilateral nerve sparing measures. The mean potency recovery rates at 6 and 12 mo were 30.6% (23.0%-38.1%) and 45.8% (32.0%-54.5%) after RRP; and 42.5% (37%-48%) and 55% (41%-66%) after LRP, respectively.

Six-month potency recovery after LRP and LRP: No statistical heterogeneity was observed (I^2 = 0%, P = 0.67) in the included four studies[21,26,28,29]. The meta-analysis evaluating potency with a fixed-effects model suggested no statistically significant difference between LRP and RRP (OR = 1.48; 95%CI: 0.94-2.34; P = 0.09) (Figure 5).

Twelve-month potency recovery after LRP and LRP: Eight studies were included[20-21,23-26,29,33] and no statistical heterogeneity was observed (I^2 = 0%, P = 0.50). The pooled analysis with a fixed-effects model showed a statistically significant advantage in favor of LRP (OR = 1.34; 95%CI: 1.05-1.70; P < 0.05) (Figure 5).

Table 5 summarizes the results of potency recovery rate after RARP and RRP. A total of 1278 patients treated with RRP and 1309 patients treated with RARP were included. In half of them, the NS procedures were not

Table 3 Comparative studies evaluating urinary continence recovery after laparoscopic radical prostatectomy or robot-assisted radical prostatectomy

Quality	Case, *n*	Author, yr	Country	Age (yr)	BMI (kg/m^2)	Prostate volume (mL,g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Continence definition	Data collection	Loss of follow-up (N/Y, %)	Urinary continence recovery, %(*n*)	
													6 mo	12 mo
High	LRP, 60	Asimakopoulos *et al*[44], 2011	Italy	61.1 ± 5.1	26.3 ± 2.2	-	-	7.37 (1.5-9.15)	RCT	0 pad	ICS-MSF	N	75.0 (45/60)	83.0 (50/60)
	RARP, 52			59.6 ± 5.4	25.8 ± 2.6			8.9 (5.8-9.2)					88.0 (46/52)	94.0 (49/52)
High	LRP, 60	Porpiglia *et al*[45], 2012	Italy	64.7 ± 5.9	26.8 ± 2.9	37.7 ± 14.1	-	8.3 ± 6.5	RCT	0-1 pad	EPIC	N	73.3 (44/60)	83.3 (50/60)
	RARP, 60			63.9 ± 6.7	26.2 ± 2.5	36.2 ± 12.6		6.9 ± 4.2					88.3a (53/60)	95.0[a] (57/60)
3/1/3(H)	LRP, 1377	Ploussard *et al*[46], 2012	France	62.7	26.6	-	-	9.8	Prospective	0 pad	Validated questionnaire	N	58.9 (811/1377)	68.5 (943/1377)
	RARP, 1009			62.7	26.5			9.2					72.0[a] (726/1009)	75.4 (761/1009)
2/1/2(M)	LRP, 50	Joseph *et al*[47], 2005	United Kingdom	61.8 ± 1.6	-	-	6 ± 0.14	6.0 ± 0.83	Retrospective	0 pad	Interview	N	92.0 (46/50)	-
	RARP, 50			59.6 ± 1.6			6 ± 0.15	7.3 ± 1.2					90.0 (45/50)	
2/1/2(M)	LRP, 70	Caballero *et al*[39], 2008	Spain	66.5 (62-69)	-	41 (30.15-52)	-	9.66 (7-16.6)	Retrospective	0 pad	Unspecified	Y, < 20%	36.4 (24/66)	-
	RARP, 60			56 (56-65.25)		29.5 (23-40)		7 (5.7-10)					60.0 (30/50)	
3/1/3(H)	LRP, 31	Lee *et al*[48], 2009	South Korea	63.0 ± 8.52	25.2 ± 2.59	37.4 ± 13.05	6.5 ± 1.23	11.7 ± 13.72	Retrospective	0-1 safety pad	Institutional questionnaire	N	80.6 (25/31)	-
	RARP, 21			64.6 ± 6.79	25.5 ± 2.64	39.9 ± 15.54	6.6 ± 0.97	8.1 ± 7.01					81.0 (17/21)	
3/1/2(M)	LRP, 60	Cho *et al*[49], 2009	South Korea	66.5 (57-75)	23.65 (18.1-28.4)	39.7 (19-72)	6.81 (5-9)	11.04 (2.72-36.6)	Retrospective	0-1 safety pad	Interview	N	71.7 (43/60)	100.0 (60/60)
	RARP, 60			66.3 (50-77)	24.61 (19.9-26.3)	36.6 (22-92.8)	6.83 (5-8)	9.98 (2.91-26.3)					93.3 (56/60)	100.0 (60/60)
4/2/3(H)	LRP, 75	Hakimi *et al*[50], 2009	United States	59.6 (43-72)	-	-	-	7.5	Retrospective	0 pad	IPSS	N	65.3 (49/75)	89.3 (67/75)
	RARP, 75			59.8 (42-71)				8.4					74.7 (56/75)	93.3 (70/75)
4/2/2(H)	LRP, 45	Trabulsi *et al*[51], 2010	United States	58.1 (43-74)	-	-	-	6.2	Retrospective	0-1 safety pad	Validated questionnaire	N	71.0 (32/45)	82.0 (37/45)
	RARP, 205			59.9 (42-76)				6.4					91.0[a] (187/205)	94.0[a] (193/205)
3/2/2(H)	LRP, 161	Willis *et al*[52], 2011	United States	58.0 ± 6.7	27.0 ± 3.4	35.2 ± 10.1	-	5.7 ± 2.9	Retrospective	0 pad	Validated questionnaire	Y, > 20%	55.0 (64/117)	72.0 (84/116)
	RARP, 121			58.1 ± 6.3	26.7 ± 3.3	41.5 ± 15.2		5.0 ± 2.2					66.0 (50/76)	75.0 (33/44)
3/1/2(M)	LRP, 62	Park *et al*[53], 2011	South Korea	65.7 (38-77)	24.6 (19.4-31.4)	30.1 (12.0-56.0)	-	9.14 (2.65-30.77)	Retrospective	0-1 safety pad	Interview	N	76.3 (47/62)	95.0 (59/62)
	RARP, 44			62.7 (46-71)	26.0 (19.7-39.4)	32.9 (15.5-66.8)		6.32 (1.86-29.5)					93.5 (41/44)	94.4 (42/44)
3/2/3(H)	LRP, 144	Park *et al*[54], 2013	South Korea	67 (38-77)	24.2 (17.2-31.4)	28.8 (12.0-74.0)	-	5.84 (0.08-41.26)	Retrospective	0 pad	Interview	N	65.5 (94/144)	78.1 (112/144)
	RARP, 183			63 (44-75)	24.7 (16.4-39.4)	30.3 (15.5-82.8)		4.98 (0.05-51.46)					83.5[a] (153/183)	87.4 (160/183)

[a]$P < 0.05$. LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; RCT: Randomized controlled trial; IPSS: International prostate symptom score; EPIC: Expanded prostate cancer index composite; ICS-MSF: International continence society-male short form questionnaire.

clearly mentioned. Three studies[11,37,40] had a high loss to follow-up rate (> 20%) at 12 mo. Tewari *et al*[34] reported that the median potency recovery was significantly better after RARP than after RRP (180 d *vs* 440 d, $P < 0.05$). The mean 12-mo potency recovery rate ranged from 40% to 50% after RRP and from 54% to 87.5% after RARP.

Six-month potency recovery after RARP and RRP: A statistically significant heterogeneity was observed among the three included studies ($I^2 = 68\%$, $P = 0.05$)[37,40,42], and the pooled analysis with a random-effects model suggested a statistically significance in favor of RARP (OR = 2.77; 95%CI: 1.23-6.21; $P < 0.05$) (Figure 6).

Twelve-month potency recovery after RARP and RRP: Six studies were included[35-37,40-42] and no statistical heterogeneity was observed ($I^2 = 0\%$, $P = 0.61$). The cumulative analysis with a fixed- effects model showed a statistically significant advantage in favor of RARP (OR = 2.66; 95%CI: 1.96-3.60; $P < 0.01$) (Figure 6).

Table 6 summarizes the results of potency recovery rate after RARP with LRP. Among these eight studies[44-46,49,50,52-54], 1322 patients who received LRP and 1203 patients who received RARP were included, and all these studies performed the NS techniques (bilateral or unilateral) except the one by Asimakopoulos *et al*[44]. Most of the studies

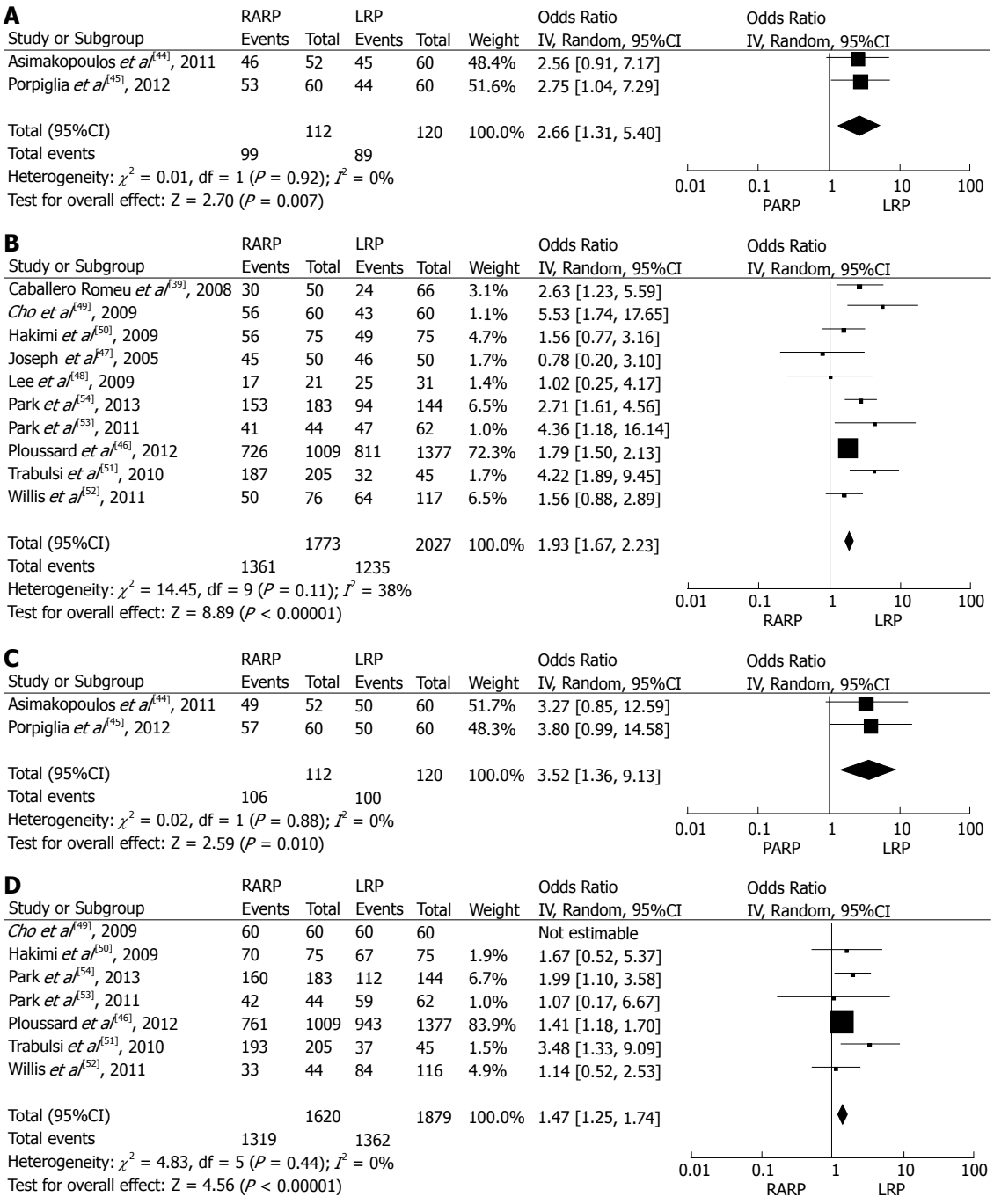


Figure 4 Forest plots and meta-analyses of robot-assisted radical prostatectomy and laparoscopic radical prostatectomy. A: 6-mo continence recovery based on randomized control trials (RCTs); B: 6-mo continence recovery based on non-randomized control trials (NRCTs); C: 12-mo continence recovery based on RCTs; D: 12-mo continence recovery based on NRCTs. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy.

used a strict potency definition as ESI. In addition, two retrospective studies[52,53] had a high loss to follow-up rate (> 20%) at 12 mo. The RCTs were evaluated separately with NRCTs. For NRCTs, the mean potency recovery rates at 6 and 12 mo were 33.8% (20.4%-48.5%) and 43.2% (31.6%-65.5%) after LRP; and 55.5% (31.1%-75%) and 65.1% (36.5%-80.0%) after RARP.

Six-month potency recovery after RARP and LRP: The two RCTs[44,45] showed a statistical heterogeneity (I^2 = 84%, P < 0.05), and demonstrated comparable result between RARP and LRP with a random-effects model (OR = 4.75; 95%CI: 0.92-24.54; P = 0.06) (Figure 7). In the cumulative analysis of five NRCTs[46,49-50,52,54], no heterogeneity was found (I^2 = 0%, P = 0.50), so a fixed-effects model was utilized. The result demonstrated a statistically significant advantage in favor of RARP (OR = 2.56; 95%CI: 2.11-3.10; P < 0.01) (Figure 7).

Twelve-month potency recovery after RARP and LRP: No evidence of statistical heterogeneity was observed in the two RCTs (I^2 = 17%, P = 0.27) and

Table 4 Comparative studies evaluating potency recovery after retropubic radical prostatectomy or laparoscopic radical prostatectomy

Quality	Case, *n*	Author, yr	Country	Age (yr)	BMI (kg/m^2)	Prostate volume (mL, g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Potency definition	Data collection	Loss of follow-up (N/Y, %)	Potency recovery(UNS/BNS),%(*n*)		Potency recovery(unclear NS), %(*n*)	
													6 mo	12 mo	6 mo	12 mo
3/2/2(H)	RRP, 70	Anastasiadis *et al*[20], 2003	France	64.8 ± 6.4	-	-	6.1 ± 1.1	11.2 ± 9.7	Prospective	ESI	Nonvalidated questionnaire	Y, > 20%	-	71.0 (23/33)	-	30.0 (10/33)
	LRP, 230			64.1 ± 6.4			5.8 ± 1.2	10.7 ± 8.8						98.0 (104/106)		41.0 (43/106)
2/2/3(H)	RRP, 33	Roumeguere *et al*[21], 2003	Belgium	63.9 ± 5.5	-	42.0 ± 20.4	5.4 ± 1.5	10.5 ± 11.5	Prospective	ESI	IIEF-5	N	33.3 (11/33)	54.5 (18/33)	-	-
	LRP, 26			62.5 ± 6.0		37.3 ± 15.6	5.4 ± 1.5	8.6 ± 5.2					34.6 (9/26)	65.3 (17/26)		
3/2/3(H)	RRP, 25	Wagner *et al*[23], 2007	United States	59 ± 6.9	29 ± 4.5	-	-	8.1 ± 6.27	Prospective	ESI	EPIC	N	-	44.0 (11/25)	-	-
	LRP, 37			58 ± 6.9	27 ± 3.0			6.2 ± 4.22						41.0 (15/37)		
3/2/2(H)	RRP, 164	Touijer *et al*[24], 2008	United States	59 (54, 64)	-	-	-	5.3 (4.1, 7.3)	Prospective	ESI	Institutional questionnaire	N	-	-	-	58.5 (96/164)
	LRP, 132			60 (55, 65)				5.3 (4.0, 7.5)								56.2 (73/130)
3/2/3(H)	RRP, 150	Greco *et al*[25], 2009	Italy	61.5 (49-74)	29 (25-33)	-	5 (3-7)	6.95 (3.4-10)	Prospective	ESI	IIEF-5	N	-	51.0 (77/150)	-	
	LRP, 150			60.5 (45-76)	32 (26-38)		5 (3-7)	6.3 (2.4-10)						66.0[a] (99/150)		
3/2/2(H)	RRP, 102	Dahl *et al*[26], 2009	United States	59.9	-	-	-	-	Prospective	ESI	Validated questionnaire	Y, > 20%	-	-	23.0 (18/77)	32.0 (23/73)
	LRP, 104			59.5											37.0 (28/75)	43.0 (33/77)
3/1/2(M)	RRP, 50	Artibani *et al*[28], 2003	Italy	64.28 ± 6.6	-	-	5.7 ± 1.2	11 ± 9	Retrospective	ESI	Nonvalidated questionnaire	Y, < 20%	-	-	10.0 (4/40)	-
	LRP, 71			63.14 ± 5.8			5.8 ± 1.3	15.7 ± 17							8.8 (5/57)	
4/2/2(H)	RRP, 42	Ghavamian *et al*[29], 2006	United States	57.8 ± 7.3	28.1	53.2 (19-135)	6.7 ± 1.3	9.9 ± 7.1	Retrospective	ESI	IIEF-5	N	38.1 (16/42)	52.5 (21/40)	-	-
	LRP, 50			60.8 ± 6.1	27.5	40.8 (20-114)	6.4 ± 0.8	7.6 ± 8.0					48.0 (24/50)	64.0 (32/50)		
2/1/1(M)	RRP, 128	Springer *et al*[32], 2013	Germany	57.2 ± 7.4	28.3 ± 2.6	-	-	3.1 ± 1.7	Retrospective	IIEF-5 > 22	IIEF-5	N	-	53.1 (68/128)	-	-
	LRP, 125			56.8 ± 6.7	27.7 ± 3.8			3.2 ± 1.4						74.4[a] (93/125)		
3/2/2(H)	RRP, 143	Magheli *et al*[33], 2014	Germany	62.6 ± 5.4	-	58 ± 22	-	10.1 ± 11.9	Retrospective	IIEF-5 > 17	Validated questionnaire	Y, > 20%	-	29.0 (18/62)	-	-
	LRP, 79			62.3 ± 5.7		53 ± 20		9.2 ± 6.9						28.0 (7/25)		

[a]$P < 0.05$. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; ESI: Erection sufficient for intercourse; IIEF: International index of erectile function; EPIC: Expanded prostate cancer index composite; UNS: Unilateral nerve sparing; PSA: Prostate-specific antigen; BNS: Bilateral nerve sparing.

the pooled analyses with a fixed-effects model showed a statistically significant advantage in favor of RARP (OR = 5.35; 95%CI: 2.77-10.31; $P < 0.01$) (Figure 7). In the six NRCTs[46,49-50,52-54], a statistical heterogeneity ($I^2 = 52\%$, $P = 0.27$) was found, and the cumulative analysis also demonstrated a statistically significant advantage in favor of RARP by using a random-effects model (OR = 1.99; 95%CI: 1.35-2.93; $P < 0.01$) (Figure 7).

Sensitivity analysis and meta-regression analysis

Sensitivity analysis was preformed to verify the reliability and stability of the evidence when a statistical heterogeneity existed. The subgroup analyses of the 6- or 12-mo urinary continence recovery following LRP and RRP did not vary significantly by source of country ($P > 0.05$), continence definition ($P > 0.05$), study design ($P > 0.05$) and loss of follow-up rate ($P > 0.05$) (Tables 7 and 8). While in the subgroup analyses of 6-mo urinary continence recovery following RARP and RRP, the results were unstable, with Western country and strict definition indicating better outcomes in favor of RARP (OR = 2.32; 95%CI: 1.47-3.67; $P < 0.01$ and OR = 3.09; 95%CI: 1.65-5.80; $P < 0.01$, respectively) (Table 9). Table 10 independently evaluates the most important factors (NS procedures) for 12-mo potency recovery among different techniques. Since all the included studies comparing RARP and LRP had taken unilateral or bilateral NS procedures, only the subgroups comparing LRP/RRP and RARP/RRP were analyzed. Our results again

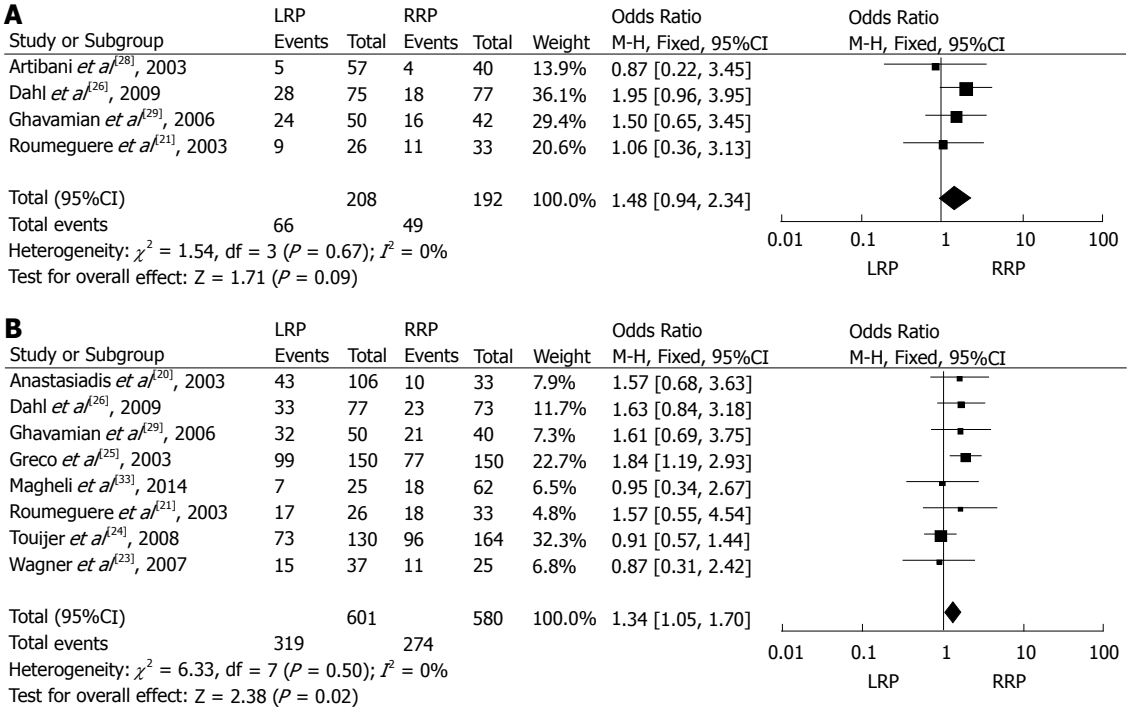


Figure 5 Forest plots and meta-analyses of laparoscopic radical prostatectomy and retropubic radical prostatectomy. A: 6-mo potency recovery; B: 12-mo potency recovery. RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy.

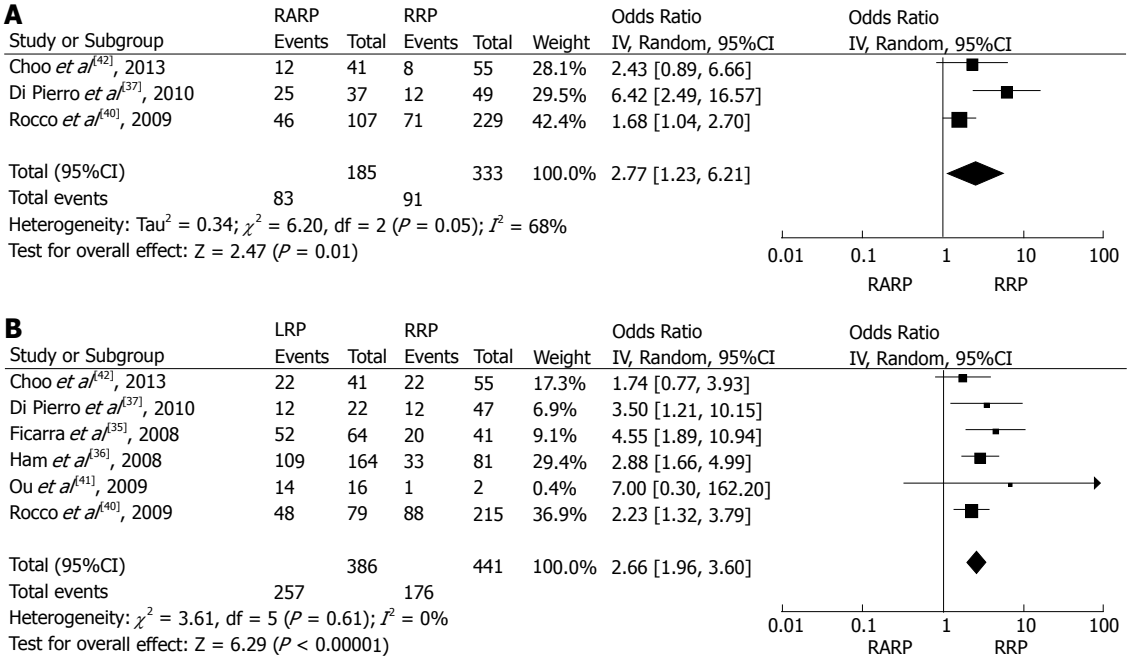


Figure 6 Forest plots and meta-analyses of robot-assisted radical prostatectomy and retropubic radical prostatectomy. A: 6-mo potency recovery; B: 12-mo potency recovery. RRP: Retropubic radical prostatectomy; RARP: Robot-assisted radical prostatectomy.

confirmed that the NS measures were advantageous factors to potency recovery ($P < 0.05$). All of the other remaining outcomes were proved to be stable and reliable by either using model conversion or exclusion of the study with the largest proportion.

Regrettably, in our meta-regression analyses, none of the adjustments such as age, BMI, prostate volume, Gleason score or PSA, achieved a statistical significance ($P < 0.05$) (Tables 11 and 12); however, the l'Abbé graphs showed an overall trend either as a positive correlation or a negative correlation between those potential factors and different surgical techniques. The older age, lower BMI and lower PSA level were associated with lower odds of different technical groups (Figure 8). The prostate vol-

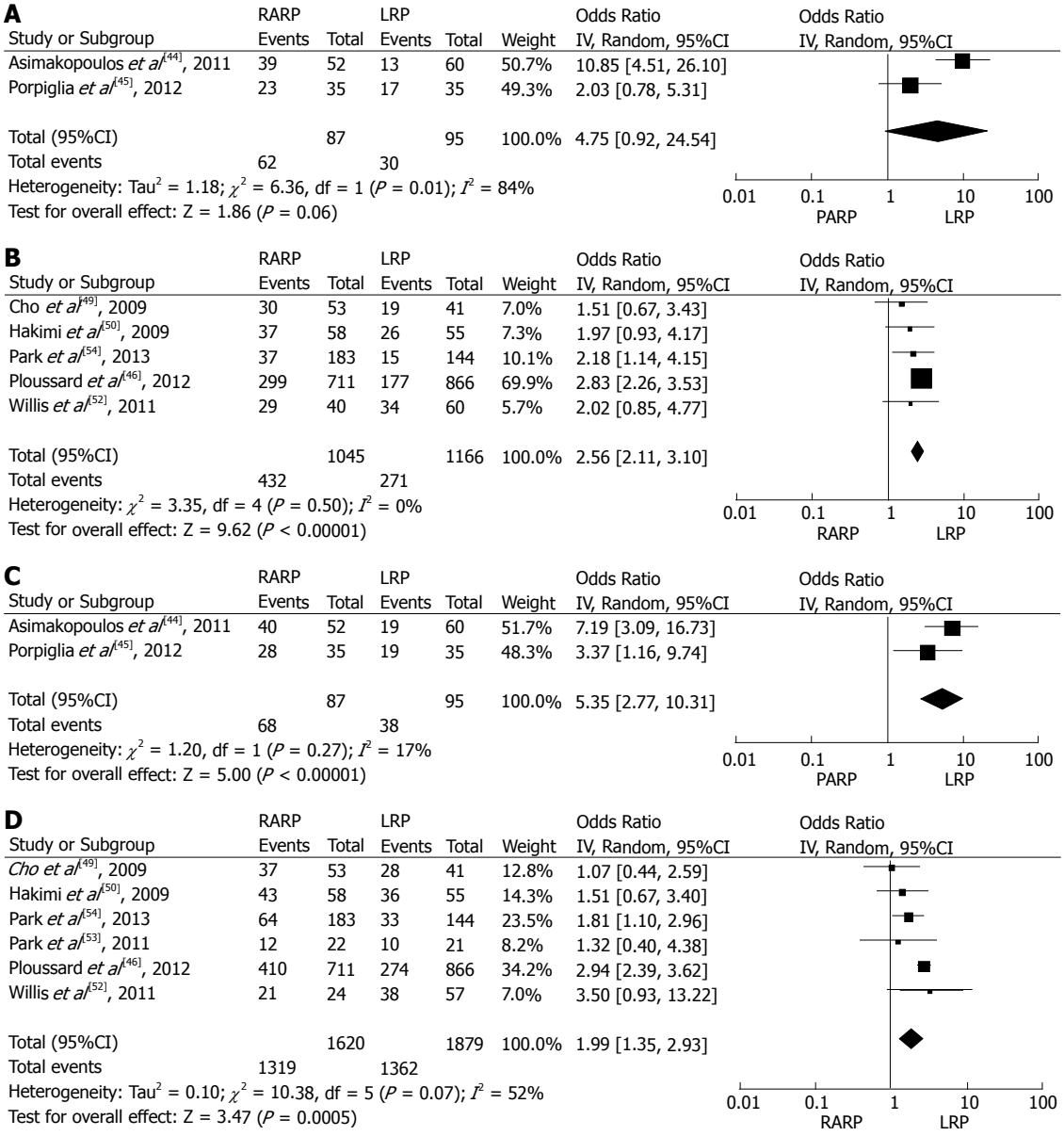


Figure 7 Forest plots and meta-analyses of robot-assisted radical prostatectomy and laparoscopic radical prostatectomy. A: 6-mo potency recovery based on randomized control trials (RCTs); B: 6-mo potency recovery based on non-randomized control trials (NRCTs); C: 12-mo potency recovery based on RCTs; D: 12-mo potency recovery based on NRCTs. LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy.

ume and Gleason score did not demonstrate any trend between the different methods of surgery (Figure 8).

Publication bias

The funnel plots of two comparative results (6-mo potency recovery after LRP/RRP and after RARP/LRP) were asymmetrical (Figure 9), indicating the existence of publication bias; this was also confirmed by Egger linear regression test ($P = 0.024$ and $P = 0.013$, respectively). All the other comparisons demonstrated symmetrical funnel plots and found no statistical significance ($P > 0.05$) by using the Egger's test, indicating no publication bias.

DISCUSSION

This meta-analysis was designed in accordance with the MOOSE reporting guidelines[10]. In 2011, Ficarra *et al*[8,9] had performed two meta-analyses which tried to compare the superiority of techniques concerning RARP *vs* RRP and RARP *vs* LRP. However, a deep investigation focusing on the deficiencies of these two studies made them possibly inconvincible: (1) limited number of studies included; (2) the lack of credible quality assessment tool for the included studies; (3) as for the comparison between RARP and LRP, it did not correspond with the methodological rules of a meta-analysis to integrate the RCT with the NRCT studies to analyze the outcomes, as they were totally two different level of evidences; (4) in the few included studies, Kim *et al*[10] and Krambeck *et al*[11] results were not available for the comparison between RARP and RRP; and (5) though all the outcomes of these two studies were apparently heterogeneous, the

Table 5 Comparative studies evaluating potency recovery after retropubic radical prostatectomy or robot-assisted radical prostatectomy

Quality	Case, *n*	Author, yr	Country	Age (y)	BMI (kg/m^2)	Prostate volume (mL, g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Potency definition	Data collection	Loss of follow-up (N/Y,%)	Potency recovery(UNS/BNS), %(*n*)		Potency recovery (unclear NS), %(*n*)	
													6 mo	12 mo	6 mo	12 mo
3/2/3(H)	RRP, 100	Tewari *et al*[34], 2003	United States	63.1 (42.8-72)	27.6 (17-41)	48.4 (24.2-70)	-	7.3 (1.9-35)	Prospective	Presence of erection	Interview	-	Median: 440 d		Median:440 d	
	RARP, 200			59.9 (40-72)	27.7 (19-38)	58.8 (18-140)		6.4 (0.6-41)					Median: 180 d[a]		Median:180 d[a]	
3/2/2(H)	RRP, 41	Ficarra *et al*[35], 2008	Italy	65 (61-69)	26 (24-28)	40 (30-47)	-	6 (5-10)	Prospective	IIEF-5 > 17	IIEF-5	N	-	49.0 (20/41)	-	-
	RARP, 64			61 (57-67)	26 (24-28)	37.5 (30-48)		6.4 (4.6-9)						81.0[a] (52/64)		
3/2/3(H)	RRP, 81	Ham *et al*[36], 2008	South Korea	66.9 ± 6.0	23.6 ± 1.8	-	-	55.2 ± 23.7	Prospective	ESI	IIEF-5	N	-	40.7 (33/81)	-	-
	RARP, 164			67.3 ± 6.2	23.6 ± 2.3			22.3 ± 34.3						66.5 (109/164)		
3/1/2(M)	RRP, 49	Di Pierro *et al*[37], 2010	Switzerland	64.3 (59.1-68.0)	-	-	-	7.57 (5.1-10.4)	Prospective	ESI	Institutional questionnaire	Y, > 20%	-	-	25.0 (12/49)	26.0 (12/47)
	RARP, 37			62.8 (58.4-67.0)				7.72 (5.6-12.1)							68.0 (25/37)	55.0 (12/22)
1/1/1(L)	RRP, 122	Kim *et al*[10], 2011	South Korea	66.5 ± 5.7	-	18.2 ± 23.4	-	14.6 ± 22.1	Prospective	ESI	Validated questionnaire	N	-	-	6.7 (8/122)	28.1 (34/122)
	RARP, 373			64.2 ± 7.3		15.2 ± 20.2		10.4 ± 16.0							33.0 (123/373)	57.1 (213/373)
2/0/1(L)	RRP, 588	Krambeck *et al*[11], 2008	United States	61.0 (41.0-77.0)	-	-	-	5.0 (0.6-39.7)	Retrospective	ESI	Institutional questionnaire	Y, > 20%	-	-	-	62.8 (262/417)
	RARP, 294			61.0 (38.0-76.0)				4.9 (0.5-33.5)								70.0 (142/203)
3/1/2(M)	RRP, 240	Rocco *et al*[40], 2009	Italy	63 (46-77)	-	-	6 (4-10)	6.7 (0.7-22.0)	Retrospective	ESI	Interview	Y, > 20%	-	-	31.0 (71/229)	41.0 (88/215)
	RARP, 120			63 (47-76)			6 (4-9)	6.9 (0.4-23.0)							43.0 (46/107)	61.0 (48/79)
3/1/3(H)	RRP, 2	Ou *et al*[41], 2009	United States	70.03 ± 6.10	24.09 ± 3.28	15.89 ± 14.15	6.22 ± 1.62		Retrospective	Presence of erection	Unspecified	N	-	50.0 (1/2)	-	-
	RARP, 16			67.27 ± 6.21	24.22 ± 3.16	16.45 ± 18.80	6.13 ± 0.9							87.5.0 (14/16)		
3/2/3(H)	RRP, 55	Choo *et al*[42], 2013	South Korea	67 ± 6.25	24 ± 2.73	42 ± 18.82		7.6 ± 19.33	Retrospective	ESI	IIEF-5	N	15.0 (8/55)	40.0 (22/55)	-	-
	RARP, 41			66 ± 7.75	24 ± 2.55	41 ± 15.77		7.2 ± 13.19					29.0 (12/41)	54.0 (22/41)		

[a]$P < 0.05$. RRP: Retropubic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; ESI: Erection sufficient for intercourse; PSA: Prostate-specific antigen; IIEF-5: International index of erectile function; UNS: Unilateral nerve sparing; BNS: Bilateral nerve sparing.

authors did not use any sensitivity analysis or subgroup analysis to explain the source of heterogeneity.

In contrast, our meta-analysis directly compared these three surgical approaches for the 6- and the 12-mo functional outcomes following radical prostatectomy (RP). In 2009, Ficarra *et al*[6] conducted a meta-analysis including only 6 studies and reported the 12-mo continence recovery following LRP and RRP, whose result was consistent with ours (OR = 0.87; 95%CI: 0.54-1.39; $P = 0.56$ and OR = 0.92; 95%CI: 0.57-1.51; $P = 0.75$, respectively). However, in our meta-analysis, the study by Rassweiler *et al*[55] was excluded because of its preoperative surgery and neo-adjuvant therapy and 13 eligible studies were included. Moreover, we evaluated the 6-mo continence recovery (OR, 0.84; 95%CI: 0.50-1.41; $P = 0.52$), so this result would be more convincible and complete. Compared with the previous meta-analysis by Ficarra *et al*[8], whose results for 12-mo urinary continence recovery based on a pooled analysis of 5 studies comparing RARP *vs* RRP, and 5 studies comparing RARP *vs* LRP identified the advantage in favor of RARP (OR = 1.53; 95%CI: 1.04-2.25; $P < 0.05$ and OR = 2.39; 95%CI: 1.29-4.45; $P < 0.01$, respectively), our meta-analyses identified the similar advantage in favor of RARP both at 6-mo and 12-mo follow-up. A critical review by Coelho *et al*[7] also indicated better outcomes after RARP compared with RRP (92% *vs* 80%) or with LRP (92% *vs* 84%). Obviously,

Table 6 Comparative studies evaluating potency recovery after laparoscopic radical prostatectomy or robot-assisted radical prostatectomy

Quality	Case, *n*	Author, yr	Country	Age (yr)	BMI (kg/m^2)	Prostate Volume (mL,g)	Gleason score (biopsy)	PSA (ng/mL)	Study design	Potency definition	Data collection	Loss of follow-up (N/Y, %)	Potency recovery (UNS/BNS), %(*n*)		Potency recovery (unclear NS), %(*n*)	
													6 mo	12 mo	6 mo	12 mo
High	LRP, 60	Asimakopoulos *et al*[44], 2011	Italy	61.1 ± 5.1	26.3 ± 2.2	-	-	7.37 (1.5-9.15)	RCT	ESI	IIEF-6	N	-	-	22.0 (13/60)	32.0 (19/60)
	RARP, 52			59.6 ± 5.4	25.8 ± 2.6			8.9 (5.8-9.2)							75.0[a] (39/52)	77.0[a] (40/52)
High	LRP, 35	Porpiglia *et al*[45], 2012	Italy	64.7 ± 5.9	26.8 ± 2.9	37.7 ± 14.1	-	8.3 ± 6.5	RCT	IIEF-5 > 17	IIEF-5	N	48.5 (17/35)	54.2 (19/35)	-	-
	RARP, 35			63.9 ± 6.7	26.2 ± 2.5	36.2 ± 12.6		6.9 ± 4.2					65.7 (23/35)	80.0[a] (28/35)		
3/1/3(H)	LRP, 866	Ploussard *et al*[46], 2012	France	62.7	26.6	-	-	9.8	Prospective	ESI	IIEF-5	N	20.4 (177/866)	31.6 (274/866)	-	-
	RARP, 711			62.7	26.5			9.2					42.1 (299/711)	57.7 (410/711)		
3/1/2(M)	LRP, 41	Cho *et al*[49],	South Korea	66.5 (57-75)	23.65 (18.1-28.4)	39.7 (19-72)	6.81 (5-9)	11.04 (2.72-36.6)	Retrospective	ESI	Interview	N	46.3 (19/41)	68.3 (28/41)	-	-
	RARP, 53			66.3 (50-77)	24.61 (19.9-26.3)	36.6 (22-92.8)	6.83 (5-8)	9.98 (2.91-26.3)					56.6 (30/53)	69.8 (37/53)		
4/2/3(H)	LRP, 55	Hakimi *et al*[50], 2009	United States	59.6 (43-72)	-	-	-	7.5	Retrospective	Presence of	IIEF-5	N	47.3 (26/55)	65.5 (36/55)	-	-
	RARP, 58			59.8 (42-71)				8.4		Erection			63.8 (37/58)	74.1 (43/58)		
3/2/2(H)	LRP, 86	Willis *et al*[52], 2011	United States	58.0 ± 6.7	27.0 ± 3.4	35.2 ± 10.1	-	5.7 ± 2.9	Retrospective	ESI	Validated	Y, > 20%	57.0 (34/60)	67.0 (38/57)	-	-
	RARP, 74			58.1 ± 6.3	26.7 ± 3.3	41.5 ± 15.2		5.0 ± 2.2			questionnaire		73.0 (29/40)	88.0 (21/24)		
3/1/2(M)	LRP, 35	Park *et al*[53], 2011	South Korea	65.7 (38-77)	24.6 (19.4-31.4)	30.1 (12.0-56.0)	-	9.14 (2.65-30.77)	Retrospective	ESI	Interview	Y, > 20%	-	47.6 (10/21)	-	-
	RARP, 37			62.7 (46-71)	26.0 (19.7-39.4)	32.9 (15.5-66.8)		6.32 (1.86-29.5)						54.5 (12/22)		
3/2/3(H)	LRP, 144	Park *et al*[54], 2013	South Korea	67 (38-77)	24.2 (17.2-31.4)	28.8 (12.0-74.0)	-	5.84 (0.08-41.26)	Retrospective	ESI	Interview	N	30.8 (26/83)	32.7 (27/83)	10.2 (15/144)	22.9 (33/144)
	RARP, 183			63 (44-75)	24.7 (16.4-39.4)	30.3 (15.5-82.8)		4.98 (0.05-51.46)					31.1 (49/156)	36.5 (57/156)	20.1 (37/183)	35.0 (64/183)

[a]$P < 0.05$. LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; RCT: Randomized controlled trial; ESI: Erection sufficient for intercourse; IIEF: International index of erectile function. EPIC: Expanded prostate cancer index composite; SHIM: Sexual health inventory for men; UNS: Unilateral nerve sparing; PSA: Prostate-specific antigen; BNS: Bilateral nerve sparing.

except the inclusion of more studies and the exclusion of two studies[10,11], our meta-analyses separated the RCT from the NRCT studies to analyze the outcomes, therefore, the result was subjected to fewer confounding and biases of study design.

In terms of potency recovery, for the first time, with 8 studies included, our meta-analysis supported the superiority of LRP than RRP at 12-mo follow-up (OR = 1.34; 95%CI: 1.05-1.70; $P < 0.05$). Compared with the previous meta-analysis by Ficarra *et al*[9], whose results for 12-mo potency recovery based on pooled 6 studies comparing RARP *vs* RRP, and 4 studies comparing RARP *vs* LRP demonstrated a better outcome in favor of RARP against RRP (OR = 2.84; 95%CI: 1.48-5.43; $P < 0.01$) and a non-statistically significant trend between RARP and LRP (OR = 1.89; 95%CI: 0.70-5.05; $P = 0.21$), our meta-analyses showed a statistically significant advantage in favor of RARP *vs* RRP (6-mo: $P < 0.05$ and 12-mo: $P < 0.01$, respectively) and also showed a better result in favor of RARP *vs* LRP (6-mo: $P < 0.01$ and 12-mo: $P < 0.01$, respectively). In ad-

Table 7 Subgroup analyses of 6-mo urinary continence recovery after laparoscopic radical prostatectomy or retropubic radical prostatectomy

Subgroup	Study	Sample size	Heterogeneity I^2(%)	*P*-value	Meta-analysis	
					OR	95%CI
Country	Asia	553	63	0.06	0.45	0.20-1.04
	America	346	0	0.45	0.83	0.51-1.34
	Europe	763	80	0.40	1.46	0.60-3.55
Continence definition	0 pad	1662	74	0.52	0.84	0.50-1.41
	0-1 pad	0	-	-	-	-
Study design	prospective	968	77	0.55	1.24	0.61-2.50
	retrospective	694	59	0.08	0.56	0.29-1.07
Loss of follow-up	≤ 20%	911	71	0.87	1.06	0.53-2.09
	> 20%	751	78	0.32	0.66	0.29-1.51

OR: Odds ratio; CI: Confidence

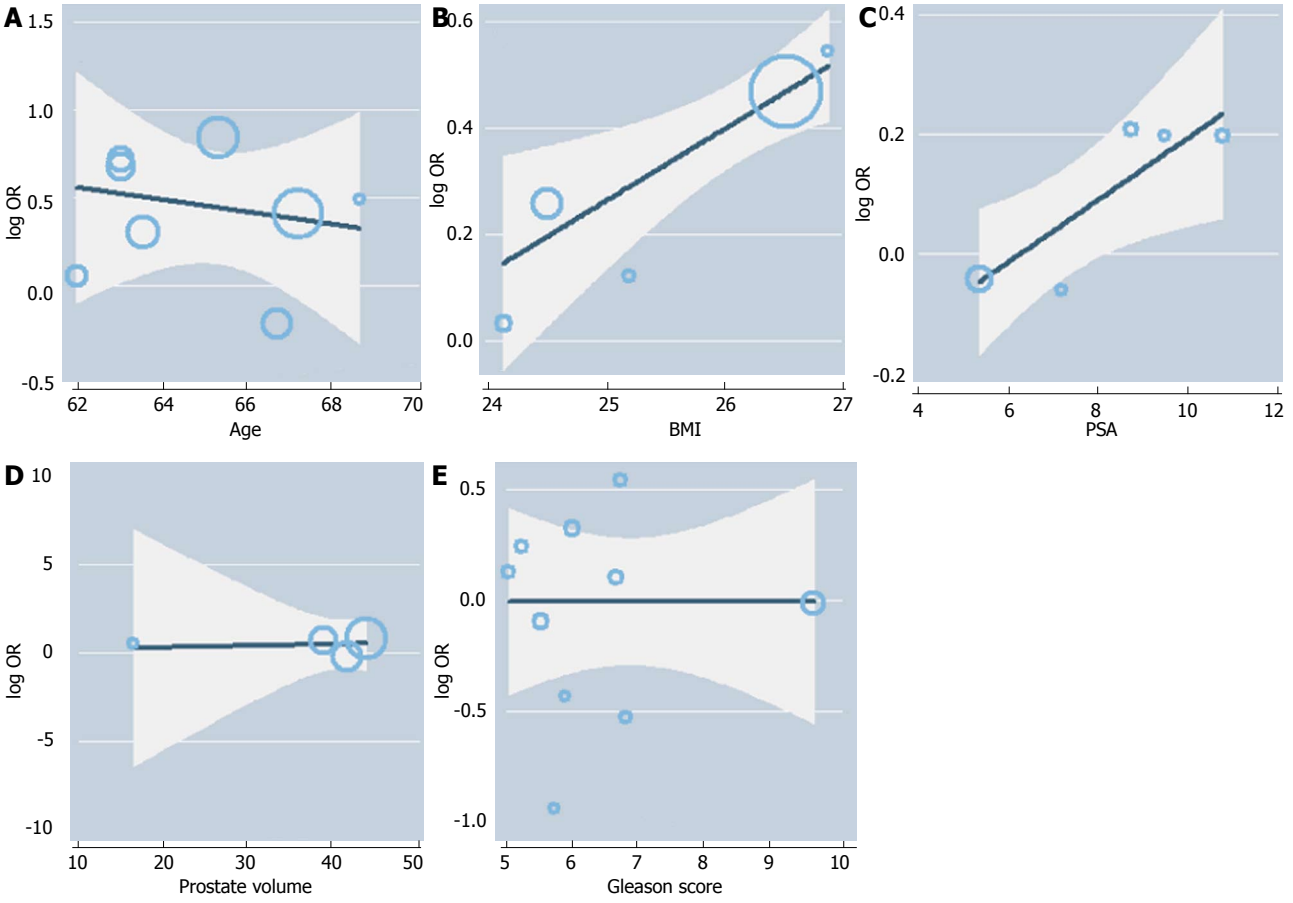


Figure 8 Representative l'Abbé plots show the overall trend. A: 12-mo continence of robot-assisted radical prostatectomy (RARP) and retropubic radical prostatectomy (RRP); B: 12-mo potency of RARP and laparoscopic radical prostatectomy (LRP); C: 12-mo potency of LRP and RRP; D: 12-mo continence of RARP and RRP; E: 12-mo continence of LRP and RRP. PSA: Prostate-specific antigen; BMI: Body mass index.

dition, there were some potential biases in Ficarra *et al*[9] meta-analysis which included only 6 studies, and two of them[10,11] were considered ineligible. While in our meta-analyses, the increased study number and the separation of the RCT and the NRCT studies, would be helpful to minimize the confounding of study design. Briefly, we supported a dramatic grading by superiority level for different comparisons of potency: RARP > LRP > RRP.

In this review, statistically significant heterogeneity was observed for several comparisons. So the subgroup analyses were added according to adjustment for country, continence or potency definition, study design and the NS procedures. We found that Western country and strict definition indicated better outcomes in favor of RARP against RRP ($P < 0.01$) for 6-mo urinary continence recovery. This difference may be explained by the popularity of robotic technique in Western countries. As the classic NS technique was repeatedly proved to be a sig-

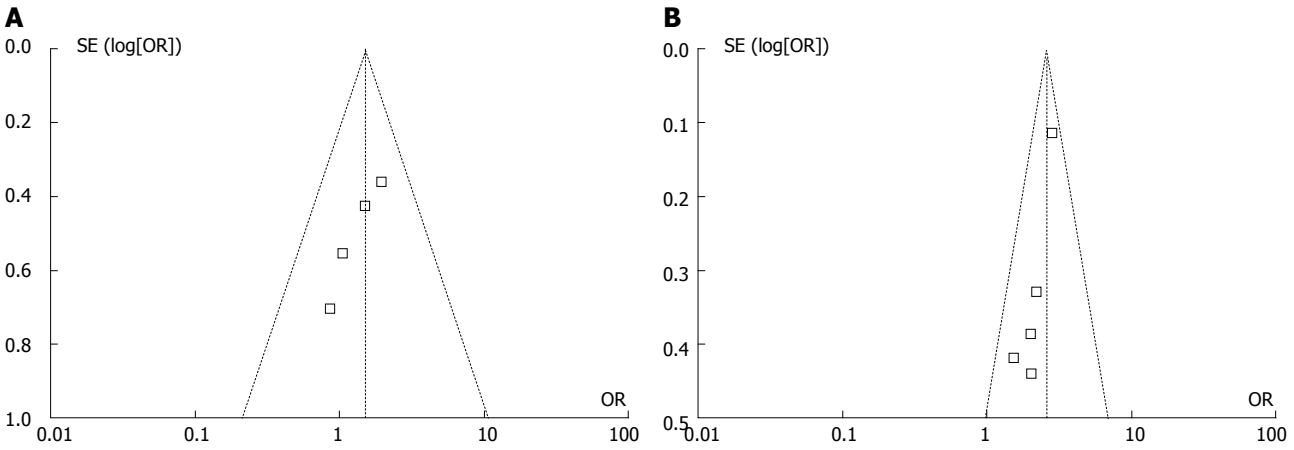


Figure 9 Funnel plots for 6-mo potency recovery. A: Comparison of laparoscopic radical prostatectomy (LRP) and retropubic radical prostatectomy (RRP); B: Comparison of robot-assisted radical prostatectomy (RARP) and LRP based on non-randomized control trials (NRCTs).

Table 8 Subgroup analyses of 12-mo urinary continence recovery after laparoscopic radical prostatectomy or retropubic radical prostatectomy

Subgroup	Study	Sample size	Heterogeneity I^2 (%)	*P*-value	Meta-analysis	
					OR	95%CI
Country	Asia	553	72	0.18	0.38	0.09-1.54
	America	911	89	0.91	0.95	0.35-2.55
	Europe	1343	29	0.33	1.26	0.79-2.02
Continence	0 pad	908	55	0.75	1.08	0.68-1.69
definition	0-1 pad	754	88	0.27	0.53	0.17-1.63
Study design	prospective	509	83	0.51	1.26	0.63-2.53
	retrospective	1153	57	0.15	0.60	0.30-1.20
Loss of follow-up	≤ 20%	451	82	0.82	1.09	0.51-2.33
	> 20%	1211	59	0.45	0.79	0.43-1.46

OR: Odds ratio; CI: Confidence interval.

Table 9 Subgroup analyses of 6-mo urinary continence recovery after robot-assisted radical prostatectomy or retropubic radical prostatectomy

Subgroup	Study	Sample size	Heterogeneity I^2 (%)	*P*-value	Meta-analysis	
					OR	95%CI
Country	Asia	809	92	0.35	1.93	0.48-7.70
	Europe/America	862	0	< 0.01	2.32	1.47-3.67
Continence	0 pad	828	63	< 0.01	3.09	1.65-5.80
definition	0-1 pad	673	82	0.52	1.62	0.37-7.06
Study design	prospective	448	0	< 0.01	2.48	1.44-4.26
	retrospective	1223	80	0.1	2.07	0.87-4.95
Loss of follow-up	≤ 20%	1161	80	0.1	2.00	0.88-4.53
	> 20%	510	0	< 0.01	2.99	1.55-5.77

OR: Odds ratio; CI: Confidence interval.

nificant predictor of return of potency by Coelho *et al*[7], by Ayyathurai *et al*[56] and by Briganti *et al*[57], this review independently evaluated it for 12-mo potency recovery between different techniques, and we confirmed again that the NS measures were advantageous factors to potency recovery ($P < 0.05$). Furthermore, the other factors such as age, BMI, prostate volume, Gleason score or PSA could also be a source of heterogeneity. Standford *et al*[58] found that urinary function varied with age and sexual function with age and race. Shikanov *et al*[59] emphasized other factors influencing continence and potency, such as baseline status, surgical technique, extent of NS and adjuvant therapy. In this review, we performed meta-regression analyses to explore the correlation between these factors and different techniques. Though no obviously statistical significance was found, the l'Abbé graphs predicted the trends that better functional outcomes were more easily achieved in patients with younger age, larger

Table 10 Subgroup analyses of 12-mo potency recovery after nerve sparing procedures

Techniques	Subgroup	Sample size	Heterogeneity I^2 (%)	*P*-value	Meta-analysis	
					OR	95%CI
LRP *vs* RRP	uni/bilateral NS	735	0	< 0.05	1.52	1.09-2.13
	unclear NS	802	22	0.37	1.17	0.83-1.65
RARP *vs* RRP	uni/bilateral NS	464	0	< 0.01	2.83	1.90-4.22
	unclear NS	446	0	< 0.01	2.43	1.52-3.90

RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; OR: Odds ratio; CI: Confidence interval; NS: Nerve sparing.

Table 11 Meta-regression of 12-mo continence recovery

Techniques	Factors	Sample, *n*	Coefficient	*P* value	95%CI	
					Lower CI	Upper CI
LRP *vs* RRP	Age	14	-0.0422414	0.480	-0.1685084	0.0840256
	Prostate Volume	7	0.0004602	0.976	-0.0367033	0.0376237
	Gleason Score	10	-0.0002758	0.998	-0.2325786	0.2320269
	PSA	11	0.0381884	0.508	-0.0871645	0.1635414
RARP *vs* RRP	Age	8	-0.0347693	0.763	-0.3038441	0.2343054
	BMI	5	0.178217	0.604	-0.8030416	1.159476
	Prostate Volume	4	0.0076432	0.912	-0.2556839	0.2709703
	PSA	5	0.0028508	0.882	-0.053367	0.0590685
RARP *vs* LRP	Age	6	-0.0026949	0.968	-0.1735327	0.1789224
	BMI	4	0.0709043	0.680	-0.7088789	0.5670703
	PSA	6	0.0275948	0.661	-0.1898594	0.1346698

RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; CI: Confidence interval; PSA: Prostate-specific antigen; BMI: Body mass index.

Table 12 Meta-regression of 12-mo potency recovery

Techniques	Factors	Sample, *n*	Coefficient	*P* value	95%CI	
					Lower CI	Upper CI
LRP *vs* RRP	Age	8	-0.0334222	0.682	-0.156947	0.2237914
	Gleason Score	5	-0.0059256	0.732	-0.5614423	0.4429304
	PSA	5	0.0509797	0.558	-0.1961242	0.2980837
RARP *vs* RRP	Age	6	-0.006352	0.939	-0.2221039	0.2093999
	PSA	5	0.0018209	0.892	-0.0373331	0.0409749
RARP *vs* LRP	Age	6	-0.0437647	0.535	-0.2229024	0.1353731
	BMI	5	0.1340739	0.315	-0.220684	0.4888318
	Prostate Volume	4	-0.0080152	0.894	-0.2365214	0.2204911
	PSA	6	0.0350044	0.588	-0.1301063	0.2001150

RRP: Retropubic radical prostatectomy; LRP: Laparoscopic radical prostatectomy; RARP: Robot-assisted radical prostatectomy; CI: Confidence interval; PSA: Prostate-specific antigen; BMI: Body mass index.

BMI or higher PSA level in the RARP group than the other two groups (LRP or RRP), while it was difficult to judge the superiority of any technique in patients with different prostate volumes and Gleason scores.

Some potential limitations should be noted. First, moderate heterogeneity was found in several comparisons. Except the potential confounding factors controlled by the inclusion criteria and analyzed with subgroup stratification as described above, surgeon's experience and the means of modification varied from one to another, which could also influence the functional outcomes and were difficult to control. Second, contrary to expectation, due to the inclusion of few eligible studies for each comparison and the lack of data in available studies, all the meta-regression analyses presented non-statistically significant differences, which limited us to reach an exact correlation between those potential factors and the three techniques, this result still needs to be identified by further research. Third, the quality of eligible studies could potentially be another confounding factor. RCTs are powerful tools, which provide the highest level of evidence; however, because many patients refuse to participate in the randomization and the blinding degree is less, surgical RCTs are difficult to conduct. Only two RCTs were included for the comparison between RARP and LRP, and the remaining studies were all observational

comparative studies. In addition, the NOS tool itself has imperfections[60]. Finally, publication bias still existed. The failed acquisition of gray literature may contribute to this publication bias.

The superiority of a certain surgical approach in terms of functional outcomes is always a pivotal controversy. These outcomes were influenced by multiple factors including patient characteristics, surgical techniques and methodology used for data collection. In summary, concerning the urinary continence recovery, only RARP showed an advantage when compared with LRP or with RRP, and the result was comparable between LRP and RRP. In terms of potency recovery, for the first time, we ranked the three surgical approaches into a superiority level: RARP > LRP > RRP, which showed a statistically significant advantage both at 6 and 12 mo postoperatively. However, the limitation of this meta-analysis and potential factors should be taken into consideration and our results also need to be validated by further high quality multi-center RCTs with strict design and large sample size.

COMMENTS

Background

Radical prostatectomy (RP) is one of the recommended standard treatments for clinically localized prostate cancer (cT1-cT2) patients. The retropubic radical prostatectomy (RRP) was considered as the gold standard and the most widely used treatment for patients with localized prostate cancer (PCa). Recently, the authors have witnessed the emergence of laparoscopic radical prostatectomy (LRP) and robot-assisted laparoscopic prostatectomy (RARP).

Research frontiers

Several experts have demonstrated that when compared with RRP, LRP and RARP have obvious advantages such as lower blood loss, less need for transfusion and shorter hospital-stay, but the lack of high-quality evidence and RCTs available precluded us from proving the superiority of any surgical option in terms of postoperative functional outcomes.

Innovations and breakthroughs

In terms of potency recovery, for the first time, we ranked the three surgical approaches into a superiority level: RARP > LRP > RRP, which showed a statistically significant advantage both at 6 and 12 mo postoperatively.

Applications

Current evidence suggests that minimally invasive approaches (RARP or LRP) are effective procedures for functional recovery. However, more high-quality randomized, controlled trials investigating the long-term functional outcomes are required to determine the advantages of RARP.

Terminology

RARP means robot-assisted laparoscopic prostatectomy. LRP means laparoscopic radical prostatectomy. RRP means retropubic radical prostatectomy. The principal postoperative functional outcomes for patients with prostatectomy are urinary continence and potency recovery.

Peer review

This manuscript compared the functional outcomes among three radical prostatectomy procedures. The project design and analyses of the data are acceptable. The figures and tables well summarize the existing data. Overall the manuscript is well written.

REFERENCES

1 **Siegel R**, Naishadham D, Jemal A. Cancer Statistics, 2013. *CA Cancer J Clin* 2013; **63**: 11-30

2 **Ferlay J**, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013; **49**: 1374-1403 [PMID: 23485231 DOI: 10.1016/j.ejca.2012.12.027]

3 **Heidenreich A**, Bastian PJ, Bellmunt J, Bolla M, Joniau S, van der Kwast T, Mason M, Matveev V, Wiegel T, Zattoni F, Mottet N. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. *Eur Urol* 2014; **65**: 124-137 [PMID: 24207135]

4 **Walsh PC**, Donker PJ. Impotence following radical prostatectomy: Insight into etiology and prevention. *J Urol* 1982; **128**: 492-497

5 **Bill-Axelson A**, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, Nordling S, Häggman M, Andersson SO, Bratell S, Spångberg A, Palmgren J, Steineck G, Adami HO, Johansson JE. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2011; **364**: 1708-1717 [PMID: 21542742 DOI: 10.1056/NEJMoa1011967]

6 **Ficarra V**, Novara G, Artibani W, Cestari A, Galfano A, Graefen M, Guazzoni G, Guillonneau B, Menon M, Montorsi F, Patel V, Rassweiler J, Van Poppel H. Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a systematic review and cumulative analysis of comparative studies. *Eur Urol* 2009; **55**: 1037-1063 [PMID: 19185977 DOI: 10.1016/j.eururo.2009.01.036]

7 **Coelho RF**, Rocco B, Patel MB, Orvieto MA, Chauhan S, Ficarra V, Melegari S, Palmer KJ, Patel VR. Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a critical review of outcomes reported by high-volume centers. *J Endourol* 2010; **24**: 2003-2015 [DOI: 10.1089/end.2010.0295]

8 **Ficarra V**, Novara G, Rosen RC, Artibani W, Carroll PR, Costello A, Menon M, Montorsi F, Patel VR, Stolzenburg JU, Van der Poel H, Wilson TG, Zattoni F, Mottrie A. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. *Eur Urol* 2012; **62**: 405-417 [PMID: 22749852 DOI: 10.1016/j.eururo.2012.05.045]

9 **Ficarra V**, Novara G, Ahlering TE, Costello A, Eastham JA, Graefen M, Guazzoni G, Menon M, Mottrie A, Patel VR, Van der Poel H, Rosen RC, Tewari AK, Wilson TG, Zattoni F, Montorsi F. Systematic review and meta-analysis of studies reporting potency rates after robot-assisted radical prostatectomy. *Eur Urol* 2012; **62**: 418-430 [DOI: 10.1016/j.eururo.2012.05.046]

10 **Kim SC**, Song C, Kim W, Kang T, Park J, Jeong IG, Lee S, Cho YM, Ahn H. Factors determining functional outcomes after radical prostatectomy: robot-assisted versus retropubic. *Eur Urol* 2011; **60**: 413-419 [PMID: 21612859 DOI: 10.1016/j.eururo.2011.05.011]

11 **Krambeck AE**, DiMarco DS, Rangel LJ, Bergstralh EJ, Myers RP, Blute ML, Gettman MT. Radical prostatectomy for prostatic adenocarcinoma: a matched comparison of open retropubic and robot-assisted techniques. *BJU Int* 2009; **103**: 448-453 [DOI: 10.1111/j.1464-410X.2008.08012.x]

12 **Stroup DF**, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**: 2008-2012 [PMID: 10789670]

13 **Wells GA**, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Cited: 2013-10-10. Available from: URL: http: //www.ohri.ca/programs/clinical_epidemiology/oxford.htm

14 **Higgins JPT**, Altman DG, Sterne JAC, editors. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of InterventionsVersion 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from: URL: http: //www.cochrane-handbook.org

15 **Higgins JP**, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539
16 **Mantel N**, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719-748 [PMID: 13655060]
17 **Fidler V**, Nagelkerke N. The Mantel-Haenszel procedure revisited: models and generalizations. *PLoS One* 2013; **8**: e58327 [PMID: 23516463 DOI: 10.1371/journal.pone.0058327]
18 **Deeks JJ**, Higgins JPT, Altman DG, editors. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from: URL: http: //www.cochrane-handbook.org
19 **Egger M**, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629-634 [PMID: 9310563]
20 **Anastasiadis AG**, Salomon L, Katz R, Hoznek A, Chopin D, Abbou CC. Radical retropubic versus laparoscopic prostatectomy: a prospective comparison of functional outcome. *Urology* 2003; **62**: 292-297 [PMID: 12893338]
21 **Roumeguere T**, Bollens R, Vanden Bossche M, Rochet D, Bialek D, Hoffman P, Quackels T, Damoun A, Wespes E, Schulman CC, Zlotta AR. Radical prostatectomy: a prospective comparison of oncological and functional results between open and laparoscopic approaches. *World J Urol* 2003; **20**: 360-366 [PMID: 12682770]
22 **Remzi M**, Klingler HC, Tinzl MV, Fong YK, Lodde M, Kiss B, Marberger M. Morbidity of laparoscopic extraperitoneal versus transperitoneal radical prostatectomy verus open retropubic radical prostatectomy. *Eur Urol* 2005; **48**: 83-89; discussion 89 [PMID: 15967256]
23 **Wagner AA**, Link RE, Trock BJ, Sullivan W, Pavlovich CP. Comparison of open and laparoscopic radical prostatectomy outcomes from a surgeon's early experience. *Urology* 2007; **70**: 667-671 [PMID: 17991534]
24 **Touijer K**, Eastham JA, Secin FP, Romero Otero J, Serio A, Stasi J, Sanchez-Salas R, Vickers A, Reuter VE, Scardino PT, Guillonneau B. Comprehensive prospective comparative analysis of outcomes between open and laparoscopic radical prostatectomy conducted in 2003 to 2005. *J Urol* 2008; **179**: 1811-1817; discussion 1817 [PMID: 18353387 DOI: 10.1016/j.juro.2008.01.026]
25 **Greco F**, Wagner S, Hoda MR, Kawan F, Inferrera A, Lupo A, Reichelt O, Jurczok A, Hamza A, Fornara P. Laparoscopic vs open retropubic intrafascial nerve-sparing radical prostatectomy: surgical and functional outcomes in 300 patients. *BJU Int* 2010; **106**: 543-547 [PMID: 20067455 DOI: 10.1111/j.1464-410X.2009.09157.x]
26 **Dahl DM**, Barry MJ, McGovern FJ, Chang Y, Walker-Corkery E, McDougal WS. A prospective study of symptom distress and return to baseline function after open versus laparoscopic radical prostatectomy. *J Urol* 2009; **182**: 956-965 [PMID: 19616252 DOI: 10.1016/j.juro.2009.05.044]
27 **Egawa S**, Kuruma H, Suyama K, Iwamura M, Baba S. Delayed recovery of urinary continence after laparoscopic radical prostatectomy. *Int J Urol* 2003; **10**: 207-212 [PMID: 12657100]
28 **Artibani W**, Grosso G, Novara G, Pecoraro G, Sidoti O, Sarti A, Ficarra V. Is laparoscopic radical prostatectomy better than traditional retropubic radical prostatectomy? An analysis of peri-operative morbidity in two contemporary series in Italy. *Eur Urol* 2003; **44**: 401-406 [PMID: 14499672]
29 **Ghavamian R**, Knoll A, Boczko J, Melman A. Comparison of operative and functional outcomes of laparoscopic radical prostatectomy and radical retropubic prostatectomy: single surgeon experience. *Urology* 2006; **67**: 1241-1246 [PMID: 16678887]
30 **Takenaka A**, Soga H, Kurahashi T, Miyake H, Tanaka K, Fujisawa M. Early recovery of urinary continence after laparoscopic versus retropubic radical prostatectomy: evaluation of preoperative erectile function and nerve-sparing procedure as predictors. *Int Urol Nephrol* 2009; **41**: 587-593 [PMID: 18810650 DOI: 10.1007/s]
31 **Simforoosh N**, Javaherforooshzadeh A, Aminsharifi A, Tabibi A. Early continence after open and laparoscopic radical prostatectomy with sutureless vesicourethral alignment: an alternative technique, 8 years' experience. *Urol J* 2009; **6**: 163-169 [PMID: 19711268]
32 **Springer C**, Inferrera A, Pini G, Mohammed N, Fornara P, Greco F. Laparoscopic versus open bilateral intrafascial nerve-sparing radical prostatectomy after TUR-P for incidental prostate cancer: surgical outcomes and effect on postoperative urinary continence and sexual potency. *World J Urol* 2013; **31**: 1505-1510 [PMID: 23400788 DOI: 10.1007/s00345-013-1036-0]
33 **Magheli A**, Busch J, Leva N, Schrader M, Deger S, Miller K, Lein M. Comparison of surgical technique (open vs. laparoscopic) on pathological and long term functional outcomes following radical prostatectomy. *BMC Urol* 2014; **14**: 18 [PMID: 24506815 DOI: 10.1186/1471-2490-14-18]
34 **Tewari A**, Srivasatava A, Menon M. A prospective comparison of radical retropubic and robot-assisted prostatectomy: experience in one institution. *BJU Int* 2003; **92**: 205-210 [PMID: 12887468]
35 **Ficarra V**, Novara G, Fracalanza S, D'Elia C, Secco S, Iafrate M, Cavalleri S, Artibani W. A prospective, non-randomized trial comparing robot-assisted laparoscopic and retropubic radical prostatectomy in one European institution. *BJU Int* 2009; **104**: 534-539 [PMID: 19281468 DOI: 10.1111/j.1464-410X.2009.08419.x]
36 **Ham WS**, Park SY, Kim WT, Koo KC, Lee YS, Choi YD. Open versus robotic radical prostatectomy: a prospective analysis based on a single surgeon's experience. *J Robotic Surg* 2008; **2**: 235-241[DOI: 10.1007/s11701-008-0111-9]
37 **Di Pierro GB**, Baumeister P, Stucki P, Beatrice J, Danuser H, Mattei A. A prospective trial comparing consecutive series of open retropubic and robot-assisted laparoscopic radical prostatectomy in a centre with a limited caseload. *Eur Urol* 2011; **59**: 1-6 [PMID: 21035248 DOI: 10.1016/j.eururo.2010.10.026]
38 **Geraerts I**, Van Poppel H, Devoogdt N, Van Cleynenbreugel B, Joniau S, Van Kampen M. Prospective evaluation of urinary incontinence, voiding symptoms and quality of life after open and robot-assisted radical prostatectomy. *BJU Int* 2013; **112**: 936-943 [PMID: 23937206 DOI: 10.1111/bju.12258]
39 **Caballero Romeu JP**, Palacios Ramos J, Pereira Arias JG, Gamarra Quintanilla M, Astobieta Odriozola A, Ibarluzea González G. [Radical prostatectomy: evaluation of learning curve outcomes laparoscopic and robotic-assisted laparoscopic techniques with radical retropubic prostatectomy]. *Actas Urol Esp* 2008; **32**: 968-975 [PMID: 19143287]
40 **Rocco B**, Matei DV, Melegari S, Ospina JC, Mazzoleni F, Errico G, Mastropasqua M, Santoro L, Detti S, de Cobelli O. Robotic vs open prostatectomy in a laparoscopically naive centre: a matched-pair analysis. *BJU Int* 2009; **104**: 991-995 [PMID: 19426191 DOI: 10.1111/j.1464-410X.2009.08532.x]
41 **Ou YC**, Yang CR, Wang J, Cheng CL, Patel VR. Comparison of robotic-assisted versus retropubic radical prostatectomy performed by a single surgeon. *Anticancer Res* 2009; **29**: 1637-1642 [PMID: 19443379]
42 **Choo MS**, Choi WS, Cho SY, Ku JH, Kim HH, Kwak C. Impact of prostate volume on oncological and functional outcomes after radical prostatectomy: robot-assisted laparoscopic versus open retropubic. *Korean J Urol* 2013; **54**: 15-21 [PMID: 23362442 DOI: 10.4111/kju.2013.54.1.15]
43 **Son SJ**, Lee SC, Jeong CW, Jeong SJ, Byun SS, Lee SE. Comparison of Continence Recovery Between Robot-Assisted Laparoscopic Prostatectomy and Open Radical Retropubic Prostatectomy: A Single Surgeon Experience. *Korean J Urol*

2013; **54**: 598-602 [DOI: 10.4111/kju.2013.54.9.598]
44 **Asimakopoulos AD**, Pereira Fraga CT, Annino F, Pasqualetti P, Calado AA, Mugnier C. Randomized comparison between laparoscopic and robot-assisted nerve-sparing radical prostatectomy. *J Sex Med* 2011; **8**: 1503-1512 [PMID: 21324093 DOI: 10.1111/j.1743-6109.2011.02215.x]
45 **Porpiglia F**, Morra I, Lucci Chiarissi M, Manfredi M, Mele F, Grande S, Ragni F, Poggio M, Fiori C. Randomised controlled trial comparing laparoscopic and robot-assisted radical prostatectomy. *Eur Urol* 2013; **63**: 606-614 [PMID: 22840353 DOI: 10.1016/j.eururo.2012.07.007]
46 **Ploussard G,** de la Taille A, Moulin M, Vordos D, Hoznek A, Abbou CC, Salomon L. Comparisons of the perioperative, functional, and oncologic outcomes after robot-assisted versus pure extraperitoneal laparoscopic radical prostatectomy. *Eur Urol* 2014; **65**: 610-619 [PMID: 23245815 DOI: 10.1016/j.eururo.2012.11.049]
47 **Joseph JV**, Vicente I, Madeb R, Erturk E, Patel HR. Robot-assisted vs pure laparoscopic radical prostatectomy: are there any differences? *BJU Int* 2005; **96**: 39-42 [PMID: 15963117]
48 **Lee HW**, Lee HM, Seo SI. Comparison of initial surgical outcomes between laparoscopic radical prostatectomy and robot-assisted laparoscopic radical prostatectomy performed by a single surgeon. *Korean J Urol* 2009; **50**: 468-474 [DOI: 10.4111/kju.2009.50.5.468]
49 **Cho JW**, Kim TH, Sung GT. Laparoscopic radical prostatectomy versus robot-assisted laparoscopic radical prostatectomy: a single surgeon's experience. *Korean J Urol* 2009; **50**: 1198-202 [DOI: 10.4111/kju.2009. 50.12.1198]
50 **Hakimi AA**, Blitstein J, Feder M, Shapiro E, Ghavamian R. Direct comparison of surgical and functional outcomes of robotic-assisted versus pure laparoscopic radical prostatectomy: single-surgeon experience. *Urology* 2009; **73**: 119-123 [PMID: 18952268 DOI: 10.1016/j.urology.2008.08.491]
51 **Trabulsi EJ**, Zola JC, Gomella LG, Lallas CD. Transition from pure laparoscopic to robotic-assisted radical prostatectomy: a single surgeon institutional evolution. *Urol Oncol* 2010; **28**: 81-85 [PMID: 20123354 DOI: 10.1016/j.urolonc.2009.07.002]
52 **Willis DL**, Gonzalgo ML, Brotzman M, Feng Z, Trock B, Su LM. Comparison of outcomes between pure laparoscopic vs robot-assisted laparoscopic radical prostatectomy: a study of comparative effectiveness based upon validated quality of life outcomes. *BJU Int* 2012; **109**: 898-905 [PMID: 21933328 DOI: 10.1111/j.1464-410X]
53 **Park JW**, Won Lee H, Kim W, Jeong BC, Jeon SS, Lee HM, Choi HY, Seo SI. Comparative assessment of a single surgeon's series of laparoscopic radical prostatectomy: conventional versus robot-assisted. *J Endourol* 2011; **25**: 597-602 [PMID: 21438677 DOI: 10.1089/end.2010.0229]
54 **Park B**, Kim W, Jeong BC, Jeon SS, Lee HM, Choi HY, Seo SI. Comparison of oncological and functional outcomes of pure versus robotic-assisted laparoscopic radical prostatectomy performed by a single surgeon. *Scand J Urol* 2013; **47**: 10-18 [PMID: 22835035 DOI: 10.3109/00365599.2012.696137]
55 **Rassweiler J**, Seemann O, Schulze M, Teber D, Hatzinger M, Frede T. Laparoscopic versus open radical prostatectomy: a comparative study at a single institution. *J Urol* 2003; **169**: 1689-1693 [PMID: 12686809]
56 **Ayyathurai R**, Manoharan M, Nieder AM, Kava B, Soloway MS. Factors affecting erectile function after radical retropubic prostatectomy: results from 1620 consecutive patients. *BJU Int* 2008; **101**: 833-836 [PMID: 18190627 DOI: 10.1111/j.1464-410X.2007.07409.x]
57 **Briganti A**, Gallina A, Suardi N, Capitanio U, Tutolo M, Bianchi M, Passoni N, Salonia A, Colombo R, Di Girolamo V, Guazzoni G, Rigatti P, Montorsi F. Predicting erectile function recovery after bilateral nerve sparing radical prostatectomy: a proposal of a novel preoperative risk stratification. *J Sex Med* 2010; **7**: 2521-2531 [PMID: 20487236 DOI: 10.1111/j.1743-6109.2010.01845.x]
58 **Stanford JL**, Feng Z, Hamilton AS, Gilliland FD, Stephenson RA, Eley JW, Albertsen PC, Harlan LC, Potosky AL. Urinary and sexual function after radical prostatectomy for clinically localized prostate cancer: the Prostate Cancer Outcomes Study. *JAMA* 2000; **283**: 354-360 [PMID: 10647798]
59 **Shikanov S**, Desai V, Razmaria A, Zagaja GP, Shalhav AL. Robotic radical prostatectomy for elderly patients: probability of achieving continence and potency 1 year after surgery. *J Urol* 2010; **183**: 1803-1807 [DOI: 10.1016/j.juro.2010.01.016]
60 **Stang A**. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 2010; **25**: 603-605

P- Reviewer: Liu YY, Papatsoris AG **S- Editor:** Ji FF
L- Editor: Wang TQ **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.13105/wjma.v2.i3.127

World J Meta-Anal 2014 August 26; 2(3): 127-134
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

META-ANALYSIS

Decision-tree analysis for cost-effective management of solitary pulmonary nodules in China

Bei Lu, Li-Xin Sun, Xi Yan, Zhen-Zhong Ai, Jin-Zhi Xu

Bei Lu, Li-Xin Sun, Xi Yan, Zhen-Zhong Ai, Jin-Zhi Xu, Department of Thoracic Surgery, the Fourth Affiliated Hospital of Harbin Medical University, Harbin 150001, Heilongjiang Province, China

Author contributions: This research paper was written by Lu B as the first author under supervision of Xu JZ; Sun LX retrieved and analyzed all medical literatures; Yan X and Ai ZZ prepared the illustrations and the statistical analysis.

Correspondence to: Dr. Jin-Zhi Xu, Department of Thoracic Surgery, the Fourth Affiliated Hospital of Harbin Medical University, No.37, Yi yuan Street, Nan gang District, Harbin 150001, Heilongjiang Province, China. 13836027105@139.com
Telephone: +86-451-82576963
Received: February 25, 2014 Revised: May 17, 2014
Accepted: July 18, 2014
Published online: August 26, 2014

Abstract

AIM: To analyze the cost-effectiveness of the diagnosis of solitary pulmonary nodule (SPN) in China.

METHODS: Decision analysis models were constructed to assess the cost-effectiveness of four strategies for the management of SPN: computed tomography (CT) alone, CT plus CT-guided automated cutting needle biopsy (ACNB), CT plus positron emission tomography/computed tomography (PET/CT), CT plus diffusion-weighted magnetic resonance imaging (DWI) plus PET/CT.

RESULTS: The prevalence of lung cancer among SPN discovered in the clinical setting was approximately 50%. The CT plus ACNB strategy had higher diagnostic accuracies (87% *vs* 81%), with a cost saving of ￥1945 RMB per patient, and reducing unnecessary thoracotomy by 16.5%; this was associated with a 4.5% missed diagnosis rate. CT plus DWI plus PET/CT strategy also had higher accuracies (95% *vs* 81%), with a cost saving of ￥590 RMB per patient, and reducing unnecessary thoracotomy by 13.5%; this was accompanied by 0.3% missed diagnosis rate. CT plus PET strategy is cost effective at a prevalence rate of 0-34%, but there was a larger prevalence range of lung cancer for CT plus ACNB strategy (from 0 to 0.6) and CT plus DWI plus PET/CT strategy(from 0 to 0.64).

CONCLUSION: CT plus DWI plus PET/CT strategy was cost-effective, and had a higher accuracy accompanied by a lower missed diagnosis rate than CT plus ACNB strategy.

© 2014 Baishideng Publishing Group Inc. All rights reserved.

Key words: Solitary pulmonary nodules; Diffusion-weighted Magnetic resonance imaging; Computed tomography-guided automated cutting needle biopsy; Positron emission tomography/computed tomography; Cost effectiveness

Core tip: It has become a major concern how to manage solitary pulmonary nodule (SPN) discovered in the clinical setting with low cost and high accuracy and reduce the morbidity and mortality associated with radiation, biopsy, and surgical procedures. However, up to now there has not been an analysis of cost-effectiveness of various strategies for the diagnosis and management of SPN in China. Recent studies in other countries may not hold true in a Chinese hospital because of differences in health-care systems and diagnostic strategy. We are the first to perform such an analysis for the cost-effective management of solitary pulmonary nodules in China.

Lu B, Sun LX, Yan X, Ai ZZ, Xu JZ. Decision-tree analysis for cost-effective management of solitary pulmonary nodules in China. *World J Meta-Anal* 2014; 2(3): 127-134 Available from: URL: http://www.wjgnet.com/2308-3840/full/v2/i3/127.htm
DOI: http://dx.doi.org/10.13105/wjma.v2.i3.127

INTRODUCTION

Lung cancer has become the leading cause of cancer deaths in China because the incidence and mortality are rapidly increasing. Although earlier detection of peripheral lung cancer may reduce the mortality rate, population-based screening using plain chest radiography has been not carried out in China. Therefore, most solitary pulmonary nodules (SPN), which are usually discovered in the clinical setting on hospital in China, have a higher malignant prevalence[1,2]. It has become a major concern how to manage SPN discovered in the clinical setting with low cost and high accuracy and reduce the morbidity and mortality associated with radiation, biopsy, and surgical procedures.

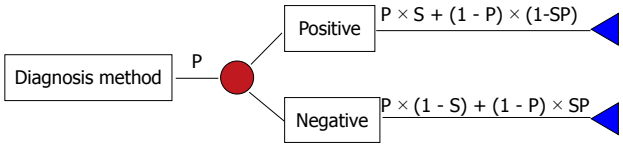
However, up to now there has not been an analysis of cost-effectiveness of various strategies for the diagnosis and management of SPN in China. Recent studies in other countries may not hold true in a Chinese hospital because of differences in health-care systems and diagnostic strategy[3-7]. The cost of surgery is not obviously higher than that of automated cutting needle biopsy (ACNB, six times) and positron emission tomography/computed tomography (PET/CT, three times) in China. But in the United States, Japan and European countries the cost of surgery was twenty times as much as that of biopsy and PET/CT. Because of high cost, PET/CT has not gained widespread popularity in most areas of China. A recent study showed that diffusion-weighted magnetic resonance imaging (DWI) may be able to be used in place of PET/CT for discriminating malignant from benign pulmonary nodules and is associated with significantly less false positives and lower costs than PET/CT[8]. Furthermore, in general, physicians in Chinese hospitals do not perform thoracoscopy with local anesthesia in patients who have or are suspected of having lung cancer.

According to the upper analysis, it is necessary to borrow previous study methods to evaluate the role of CT, ACNB, DWI, and PET/CT in the investigation of SPN which are discovered in the clinical setting in hospital in China. We used decision-tree analysis models and compared four strategies to understand under what conditions various strategies should be cost-effective and have high diagnostic accuracy.

MATERIALS AND METHODS

In the current study, we have defined SPN as spherical intrapulmonary x-ray densities less than 4.0 cm in diameter with no calcium visible on a standard chest X-ray. There is no associated atelectasis, hilar enlargement, or pleural effusion. We do not include patients with any evidence of metastasis or a recent primary malignancy outside of the chest. We also exclude cases that have previous radiographs that have already established the stability of the rate of growth of the nodule. Multiple nodules are also not considered in the current analysis.

The analysis for cost-effectiveness was performed using quantitative methods of decision analysis. Decision-tree models were constructed with multiple competing strategies, and reported values of prior probability of cancer, and sensitivity and specificity of each diagnostic modality were applied to them using Chinese health care costs. The average cost per patient for each strategy, including all diagnostic tests and surgery when undertaken, was calculated. The medical literature was surveyed to obtain the performance of each diagnostic test. For the calculations, Decision Analysis Add-In for Microsoft Excel (Version1.0.6, Palisade Corporation) was used. The decision tree was constructed using the choices and potential outcomes of the choices. All conditional probabilities of each outcome in the tree were calculated and obtained as a function of the variables listed in Tables 1 and 2 by using Bayesian analysis, Calculation Methods:



Where P = prevalence, S = sensitivity, and Sp = specificity.

Calculations of overall cost of competing strategies were calculated by summing the products of the probabilities and values of the outcome of each strategy. Overall costs per patient in each strategy can be calculated automatically by the software and listed under "chance" in Figure 1. Accuracies (A) of each diagnosis method in each strategy were calculated by formula:

$A = (S - Sp)P + Sp$

False positive (FP), false negatives (FN), true positive (TP) and true negatives (TN) of each diagnosis method in each strategy were calculated by the formula:

$FP = 1/\{[(Sp-A)S/(1-Sp)(A-S)]+1\}$

$FP = 1/\{[(S-A)Sp/(1-S)(A-Sp)]+1\}$

Overall accuracies of each strategy were calculated by summing the products of the probabilities of the outcome of each strategy and TP plus TN in each strategy; The overall misdiagnosis rate of each strategy was calculated by summing the products of the probabilities of the outcome of each strategy and FP in each strategy; the overall missed diagnosis rate of each strategy was calculated by summing the products of the probabilities of the outcome of each strategy and FN in each strategy.

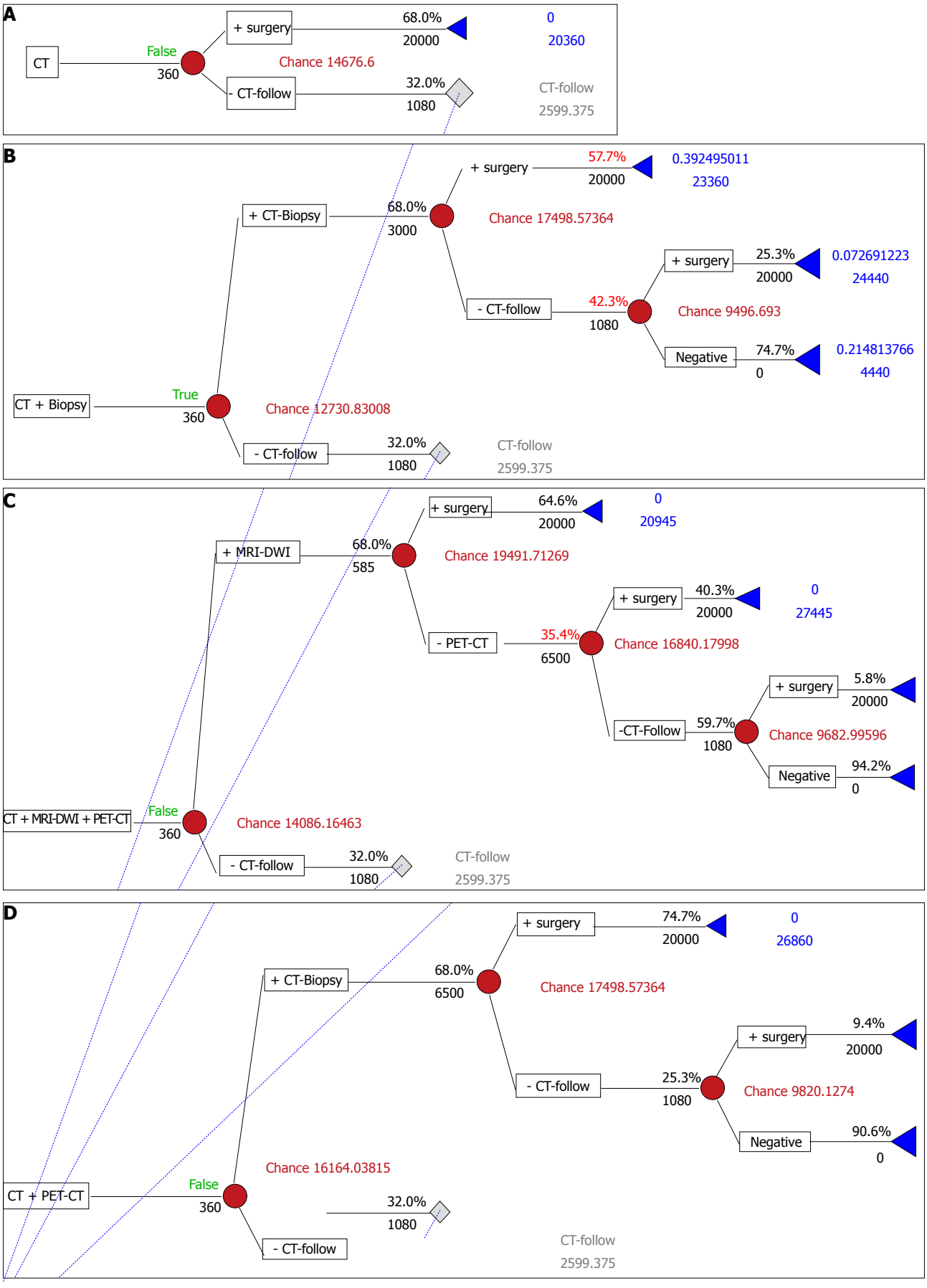
Cost-effectiveness was expressed as the incremental cost-accuracy ratio (ICAR) where:

$ICAR = (Cost_{strat}-Cost_{bl}/(Accuracy_{strat}-Accuracy_{bl})$ (RMB/%)[9].

$Cost_{strat}$ and $Accuracy_{strat}$ are the average cost per patient and accuracy of the strategy being compared, and $Cost_{bl}$ and $Accuracy_{bl}$ are the cost per patient and accuracy of a baseline strategy, which was the CT alone strategy in this study. A negative ICAR resulting from a negative numerator and positive denominator indicates that the strategy being compared is clearly preferred. In the case of a positive ICAR, a criterion must be chosen that gives the acceptable cost to get a higher accuracy.

Decision-tree analysis models

We compared four strategies for the management of


A
CT
False
360
+ surgery
68.0%
20000
0
20360
Chance 14676.6
- CT-follow
32.0%
1080
CT-follow
2599.375
B
CT + Biopsy
True
360
+ CT-Biopsy
68.0%
3000
+ surgery
57.7%
20000
0.392495011
23360
Chance 17498.57364
- CT-follow
42.3%
1080
+ surgery
25.3%
20000
0.072691223
24440
Chance 9496.693
Negative
74.7%
0
0.214813766
4440
Chance 12730.83008
- CT-follow
32.0%
1080
CT-follow
2599.375
C
CT + MRI-DWI + PET-CT
False
360
+ MRI-DWI
68.0%
585
+ surgery
64.6%
20000
0
20945
Chance 19491.71269
- PET-CT
35.4%
6500
+ surgery
40.3%
20000
0
27445
Chance 16840.17998
-CT-Follow
59.7%
1080
+ surgery
5.8%
20000
Chance 9682.99596
Negative
94.2%
0
Chance 14086.16463
- CT-follow
32.0%
1080
CT-follow
2599.375
D
CT + PET-CT
False
360
+ CT-Biopsy
68.0%
6500
+ surgery
74.7%
20000
0
26860
Chance 17498.57364
- CT-follow
25.3%
1080
+ surgery
9.4%
20000
Chance 9820.1274
Negative
90.6%
0
Chance 16164.03815
- CT-follow
32.0%
1080
CT-follow
2599.375

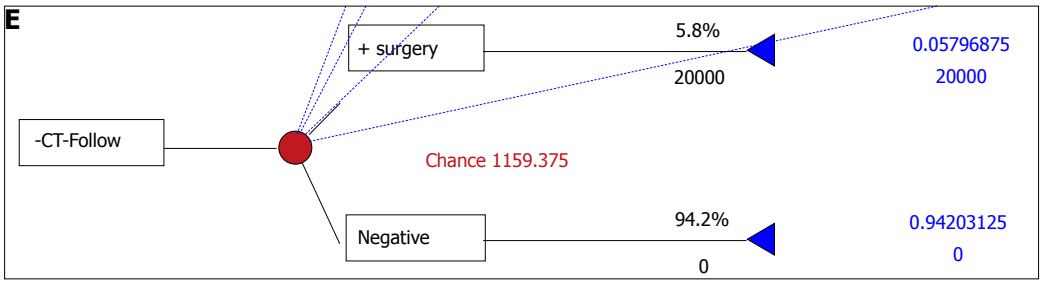


Figure 1 Strategies for the management of solitary pulmonary nodule. A: CT alone strategy (base line); B: CT plus CT-guided needle biopsy strategy; C: CT plus DWI plus PET strategy; D: CT plus PET strategy. PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging.

Table 1 Performance parameters used in the decision-tree analyses

	Sensitivity	Specificity	Ref.
Chest CT	0.99	0.63	[5]
PET-CT	0.97	0.85	[5]
ACNB	0.769	0.936	[5]
CT-Follow	0.56	0.95	[3]
DWI	0.7	0.97	[3]

PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging; ACNB: Automated cutting needle biopsy.

Table 2 Medical costs in China used in the decision-tree analyses

	Cost (RMB)
Chest CT without contrast enhancement	￥360
PET-CT	￥6500
ACNB[1]	￥3000
CT-Follow: continuous three Chest CT	￥1080
DWI	￥585
Surgical resection of SPN	￥20000

[1]Including the costs of possible treatment of pneumothorax. PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging; SPN: Solitary pulmonary nodule; ACNB: Automated cutting needle biopsy.

SPN (Figure 1): (1) CT alone strategy (baseline), (2) CT plus PET/CT strategy, (3) CT plus ACNB strategy, and (4) a CT-plus-DWI plus PET/CT strategy.

In these strategies, all patients initially undergo a chest CT examination without contrast enhancement. If an initial CT diagnosis of benign pulmonary nodule is made or no SPN is demonstrated on CT, the patients are followed up using an unenhanced chest CT. When follow-up CT shows the growth of the SPN, the diagnosis of lung cancer is made: (1) CT alone strategy: The CT alone strategy simulates a simple diagnostic approach to the management of patients with SPN. In this strategy, all patients, in whom SPN is diagnosed as lung cancer on the initial chest CT, proceed to surgical resection without pathological confirmation; (2) CT plus PET strategy: In this strategy, all patients who are CT positive (*i.e.*, an SPN is diagnosed as lung cancer) undergo chest PET-CT. If the PET/CT is then also positive, the patients proceed to surgical resection. If a chest PET/CT examination is negative after a positive CT examination, the patients are followed up by an unenhanced chest CT; (3) CT plus ACNB strategy: If the initial CT is positive, it is followed directly by ACNB. If the biopsy is then also positive, patient undergo surgical resection. If ACNB is negative, the patients are followed up by an unenhanced chest CT; (4) CT plus DWI plus PET/CT strategy: In this strategy, all patients who are CT positive undergo a chest CT plus DWI strategy. If the CT plus DWI strategy is then also positive, the patients proceed to surgical resection. If a chest a CT plus DWI strategy examination is negative after a positive CT examination, the patients are given an extra CT-plus-PET strategy above-mentioned.

Prevalence of cancer and diagnostic performance (sensitivity and specificity)

Although there were no large sample statistics for epidemiology, some small sample studies and our experience indicated the prevalence of lung cancer among SPN was approximately 50% in China[1,2]. Therefore, we applied this value (50%) to the decision-tree analyses.

The diagnostic performance (sensitivity and specificity) of each diagnostic test are gleaned from the literature and entered into the models (Table 1). Although multi-slice spiral CT is taking the place of conventional thin slice CT gradually, the diagnostic performance of multi-slice spiral CT on SPN still lack of large sample statistics. Tsubamoto *et al*[10] reported that accuracy of the final diagnosis based on coronal multi-planar reconstruction of with a multi-detector-row CT scanner (74%) was almost equal to that based on transverse thin-section CT (71%) ($P = 0.3$) for the evaluation of solitary pulmonary nodules. As a result, we still used CT sensitivity of 99% and specificity of 63% in the largest series published by Siegelman *et al*[11], their study included 634 nodules, using conventional thin slice CT. From many published reports of PET/CT imaging of SPN, we adopted the results of Yi *et al*[12] and Kim *et al*[13]. We applied PET/CT sensitivity of 97% and specificity of 85% to the decision-tree analyses. The diagnostic accuracy of CT-guided transthoracic ACNB has been reported based on comparisons with

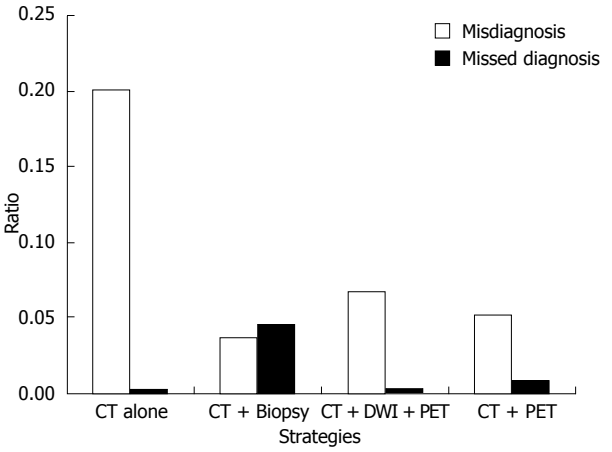


Figure 2 Ratio of the patients who undergo unnecessary surgery for a benign solitary pulmonary nodule and in whom lung cancer is misdiagnosed as a benign nodule. The strategies using CT + DWI + PET and CT + ACNB dramatically decrease the ratio of the patients who undergo unnecessary surgery for a benign SPN, although the ratio of cancer which is missed at diagnosis is increased. PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging; SPN: Solitary pulmonary nodule; ACNB: Automated cutting needle biopsy.

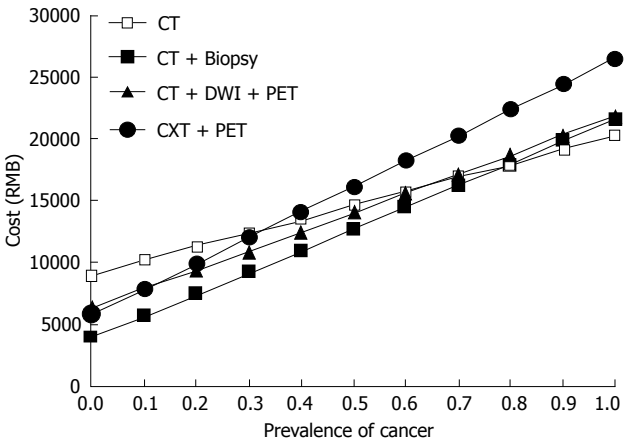


Figure 3 Correlation between the prevalence of cancer and total costs per patient. The total costs per patients were increased regardless of the strategies and the difference among the strategies became small, although the CT+PET alone strategy had the highest cost at the prevalence of 50%. PET: Positron emission tomography; CT: Computed tomography.

fine-needle aspiration biopsy. In China, ACNB has been often used by physicians, usually without combining the use of a long-throw biopsy needle, higher mean number of needle passes, and tandem system like Satoh *et al*[14]; as a result we selected ACNB sensitivity of 76.9% and specificity of 93.6% by Tsukada *et al*[15]. Although serial CT is probably the most common follow-up methodology to be adopted in the practice, there is little data on the utility of follow-up chest CT for the diagnosis of SPN. Takashima *et al*[16] reported the sensitivity of 0.56 and specificity of 0.95, when the nodule showing an increase of 0.5 mm or more in the maximum or perpendicular diameter or both on high-resolution CT images was diagnosed as a lung cancer. We applied these values to our analysis. There is a few data on the utility of DWI for the diagnosis of SPN. The study by Mori *et al*[8] reported the sensitivity of 0.70 and specificity of 0.97, which is similar to our unpublished results, we applied these values to our analysis.

Cost of each medical procedure

The mean costs of diagnostic tests and thoracotomy in China are shown in Table 2. The cost of thoracotomy in RMB was based on the bills in our hospital during 2009 (n = 20). The costs of diagnostic examinations include the costs of diagnostic procedure and radiological or/and pathological interpretations based on the data of the Harbin Price Bureau (These costs are uniform in China according to the Ministry of Public Health of the People's Republic of China). In China, patients undergoing ACNB do not stay overnight in a hospital. Although major complications are rare, pneumothorax is the most common complication after ACNB with a reported rate of 22%-54%[14,17,18]. According to our study, chest tube placement was not necessary for the incidence of pneumothorax. A central venous catheter attached to a negative pressure aspirator was used for the treatment of pneumothorax in our hospital. The costs of possible treatment of pneumothorax were included in the cost of CT-guided needle biopsy.

Statistical analysis

The decision tree methodology can be extended by performing sensitivity analysis to determine the conditions under which the new test remains cost-effective. This was achieved by entering different values for the disease prevalence (0-100%) into the decision-tree models. Although the costs of medical procedures will clearly vary among countries, within China these costs are fixed and therefore need not constitute a variable in the sensitivity analysis.

RESULTS

The prevalence of lung cancer among SPN discovered in hospital in the clinical setting was approximately 50%. In this prevalence, the strategies using CT plus ACNB and CT plus DWI plus PET/CT were the cost-effective alternatives to the CT alone strategy. The CT plus ACNB strategy had higher accuracies (87% *vs* 81%), with a cost saving of ￥1945 RMB per patient, and reduced the number of candidates who underwent unnecessary thoracotomy for a benign SPN by 16.5%; this was accompanied by a rate of missed diagnosis of 4.5%. The strategies using CT plus DWI plus PET/CT had also higher accuracies (95% *vs* 81%), with a cost saving of ￥590 RMB per patient, and reduced unnecessary thoracotomy by 13.5%; this was accompanied by a rate of missed diagnosis of 0.3% (Table 3, Figure 2).

At a prevalence of cancer up to 34%, CT plus PET strategy is cost effective with a higher accuracy than CT alone strategy. But CT plus ACNB strategy and CT plus DWI plus PET/CT strategy had a better cost-effective-

Table 3 Overall preoperative diagnostic accuracy, costs per patient and incremental cost-accuracy ratio in each strategy, when the prevalence of cancer is 50%

Strategy	Accuracy	Cost per patient	ICAR (RMB/%)
CT alone	0.81	14676	Baseline
CT + PET-CT	0.952	16164	8196
CT + ACNB	0.867	12731	-17583
CT + DWI + PET-CT	0.949	14086	-4247

PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging; ICAR: Incremental cost-accuracy ratio; ACNB: Automated cutting needle biopsy.

ness over a larger prevalence of lung cancer, ranging from 0 to 0.60 (CT plus ACNB) and from 0 to 0.63 (CT plus DWI plus PET/CT).

The total costs per patient were increased regardless of the strategies, and the difference among the strategies became small (Figure 3). By reducing the number of the candidates who undergo unnecessary thoracotomy for a benign SPN (Figure 2), CT plus ACNB strategy and CT plus DWI plus PET/CT strategy decreased the total costs per patients. The total costs per patient of CT plus DWI plus PET/CT strategy were reduced not only by reducing unnecessary thoracotomy by 13.5%, but also by giving 44% of the candidates a chance to avoid undergoing expensive PET/CT, meanwhile, the overall accuracy of the CT plus DWI plus PET/CT strategy was not lower than that of CT plus PET strategy (Figure 4).

ICAR of the strategies using CT plus ACNB and CT plus DWI plus PET/CT were negative values with higher accuracies, clearly representing cost-effectiveness compared to the CT alone strategy (Figure 5). Although CT plus ACNB strategy is more cost saving, CT plus DWI plus PET/CT strategy has a lower rate of missed diagnosis (Figure 2) and a higher overall accuracy (Figure 4).

Figure 5 plots the prevalence of cancer versus ICAR. Since the CT alone strategy had a higher cost at the prevalence of cancer between 0 and 64% (Figure 3), in our analysis, a negative ICAR means that CT plus ACNB strategy and CT plus DWI plus PET/CT strategy have lower costs with higher accuracies. At the prevalence of cancer up to 34%, all strategies were cost-effective.

DISCUSSION

From the analyses performed here, adding ACNB, DWI, or PET/CT is advantageous in terms of greater accuracy and cost-effectiveness. The results of our sensitivity analyses showed that CT plus ACNB strategy and CT plus DWI plus PET/CT strategy have a wide range of prevalence of cancer, and CT alone strategy becomes more cost-effective only when the prevalence of cancer exceeds approximately 65%.

In the study, we assumed that SPN evaluated were discovered in the clinical setting in hospital. In our daily clinical setting, the prevalence of lung cancer among SPN was approximately 50%, far higher than those detected

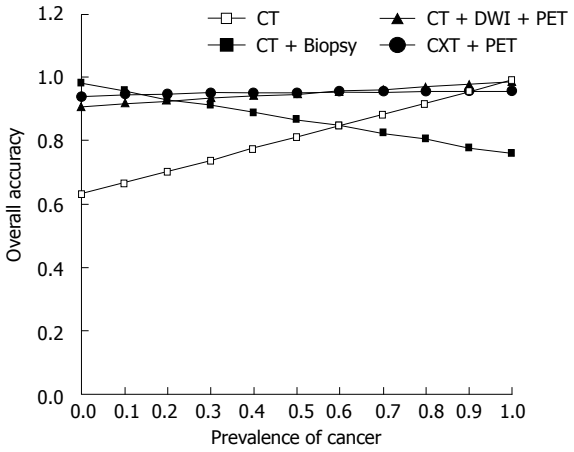


Figure 4 Correlation between the prevalence of cancer and the overall accuracy. The use of CT+PET or CT+DWI+PET resulted in a higher accuracy between the prevalence of 20%-90% compared to CT alone strategy and CT+Biopsy. PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging.

on lung cancer screening. For SPN discovered in the lung cancer screening, for a lower prevalence (approximately 10%), reducing false-positives to avoid more morbidity, mortality and increased costs for unnecessary thoracotomy is of importance; but for SPN discovered in the clinical setting in hospital, with a very high pretest probability, avoiding a missed diagnosis for false-negatives is as important as decreasing the misdiagnosis rate and reducing the cost. In the current study, CT plus ACNB strategy had a more cost saving and lower misdiagnosis rate, but also accompanied by higher rate of missed diagnosis up to 4.5%. The strategies using CT plus DWI plus PET/CT had higher accuracies (95% *vs* 81%), accompanied by a lower rate of missed diagnosis of 0.3%. This means there is an increase in life expectancy for decreasing rate of missed diagnosis.

The current analysis implies that the use of CT plus ACNB strategy or CT plus DWI plus PET/CT strategy can provide an advantage of a lower cost with higher accuracy. In the prevalence of cancer of 50%, the CT plus DWI plus PET/CT strategy is the optimal choice. In the United States, Japan and European countries the cost of surgery was twenty times as much as that of PET/CT. But in China the cost of surgery is only three times as much as that of PET/CT. Therefore, by reducing the number of the candidates who undergo unnecessary thoracotomy for a benign SPN, CT plus PET/CT strategy only decreased limited total costs per patients, which was hard to compensate for expensive cost of PET/CT, especially when the prevalence of lung cancer was up to 34%. A recent study showed that DWI may be able to be used in place of PET/CT for discriminating malignant from benign pulmonary nodules and is associated with significantly less false positive and lower cost than PET/CT[8]. In our unpublished study, a false-positive of DWI for SPN was only 3.5%. We assumed that most malignant SPN first could be identified by undergoing DWI previous to PET/CT, then negative nodules were added to

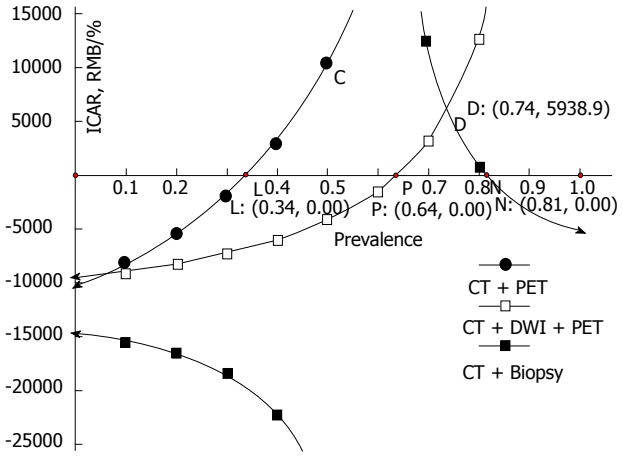


Figure 5 Incremental cost-accuracy ratio compared with computed tomography alone strategy as the baseline. A negative ICAR means cost-effectiveness. At the prevalence of cancer up to 34%, all strategies being compared are cost effective; in particular, the CT plus PET plus DWI strategy is cost-effective at the prevalence up to 63%. ICAR: Incremental cost-accuracy ratio; PET: Positron emission tomography; CT: Computed tomography; DWI: Diffusion-weighted magnetic resonance imaging.

PET/CT examination at a relatively low prevalence of lung cancer. The current study showed that the CT plus DWI plus PET/CT strategy decreased the total costs per patients not only by reducing unnecessary thoracotomy by 13.5%, but also by giving 44% of the candidates a chance to avoid undergoing expensive PET/CT; meanwhile, overall accuracies of CT plus DWI plus PET/CT strategy was not significantly lower than that of CT plus PET strategy.

ICAR of CT plus ACNB strategy are lower than that of the CT plus DWI plus PET/CT strategy, suggesting better cost-effectiveness of ACNB for SPN diagnosis, but we considered CT plus ACNB strategy is not an optimal option for the evaluation of SPN according to the analysis of current results. Most importantly, overall accuracies of CT plus ACNB strategy decreased gradually with the prevalence of lung cancer increasing, and was similar to that of CT alone strategy at a prevalence of lung cancer up to 50%. With the accompaniment of this, CT plus ACNB strategy had a rate of missed diagnosis of 4.5% for the evaluation of SPN, which was significantly higher than 0.3% of CT plus DWI plus PET/CT strategy. The high rate of missed diagnosis will inevitably reduce life expectancy. Secondly, unlike those discovered by screening chest radiography, most SPN discovered in the clinical setting in hospital are usually larger than 1 cm in diameter and with a high pretest probability. For such a large SPN, PET/CT has a higher negative predictive value and a lower rate of missed diagnosis; if this is accompanied with DWI with a higher positive predictive value and lower misdiagnosis rate, there will be an increase in cost-effectiveness. Our results underscored the point. In addition, although major complications are rare, pneumothorax is the most common complication with a reported rate of 22%-54% with ACNB[14,17,18]. A central venous catheter attached to negative pressure aspirator was used for the treatment of pneumothorax in our hospital. These invasions not only increased medical costs, but also caused some physical or mental injuries.

The current study is entirely a statistical simulation, and individual variation should and will inevitably occur in real medical practice. Therefore, the utility of a decision tree analysis must always be limited. Patient preferences and concerns must be considered when determining how to manage an individual patient. Although, according to this analysis, the CT plus DWI plus PET strategy was the optimal choice, the selection of strategy depends not only on the cost-effectiveness, but also on variable patient factors and accessibility to the modalities. Risk-taking attitudes of the patient and physician will also influence the choice of testing strategies. Because PET/CT is not yet widely available in China, with a difficulty in accessing it for many patients, CT alone strategy often is used for discrimination between malignant and benign SPN. Because of a high prevalence of lung cancer among SPN discovered in hospital in China, especially with a pretest probability up to 64%, in terms of ICAR in Figure 5 CT alone strategy is cost-effective. Between a prevalence of 80%-100%, CT plus ACNB is also cost-effective (Figure 5), but is often not used because of a rather low accuracy (Figure 4).

There were some limits to our study. First, SPN which are diagnosed as lung cancer were assumed to require surgery regardless of cancer staging. However, many factors interact to determine the needs of further examinations of the lung lesions and its resectability, including not only tumor staging, but also performance status and cardiopulmonary status. These factors were not traced in the models. Secondly, because of the lack of large sample study in China, the diagnostic performance (sensitivity and specificity) of each diagnostic test was gleaned from the literature which were published in the United States, Japan and European countries. We used those that were similar to results of a small sample study in China.

In conclusion, the introduction of CT plus DWI plus PET/CT strategy for the evaluation of SPN, which are discovered on chest radiography in the clinical setting, is potentially cost-effective in China with higher accuracy, over a large prevalence of cancer. When the prevalence of cancer rises up to 65%, the introduction of CT alone strategy for the evaluation of SPN is potentially cost-effective.

COMMENTS

Background

It has become a major concern about how to reduce the mortality rate of lung cancer, for it has become the leading cause of cancer deaths in China because the incidence and mortality are rapidly increasing. Although the authors have noted earlier detection of peripheral lung cancer, population-based screening using plain chest radiograph has been not carried out in China. Therefore, most solitary pulmonary nodules, which are usually discovered in the clinical setting in hospital in China, have a higher malignant prevalence.

Research frontiers

Recent studies in other countries may not work for the Chinese hospital setting because of differences in health-care systems and diagnostic strategy. The cost

of surgery is not obviously higher than that of automated cutting needle biopsy (ACNB, six times) and positron emission tomography/computed tomography (PET/CT, three times) in China. But in the United States, Japan and European countries the cost of surgery was twenty times as much as that of biopsy and PET/CT. Because of expensive cost, PET/CT has not gained widespread popularity in most areas of China. A recent study showed that diffusion-weighted magnetic resonance imaging (DWI) may be able to be used in place of PET/CT for discriminating malignant from benign pulmonary nodules and is associated with significantly less false positives and lower cost than PET/CT.

Innovations and breakthroughs

Up to now there has not been an analysis of cost-effectiveness of various strategies for the diagnosis of solitary pulmonary nodule (SPN) in China. Through the four strategies for the management of SPN: CT alone, CT plus CT-guided ACNB, CT plus PET/CT, CT plus DWI plus PET/CT, we analyzed the cost-effectiveness and chose CT plus DWI plus PET/CT strategy as an optimal option for the evaluation of SPN in China, because this strategy was not only cost-effective, but also had a higher accuracy accompanied by a lower missed diagnosis rate than CT plus ACNB strategy.

Applications

According to this analysis, the CT plus DWI plus PET strategy was the optimal choice, which can be applied on chest radiograph in the clinical setting, having potential cost-effectiveness in China with higher accuracy, over a large prevalence of cancer.

Terminology

SPN is the round or oval opaque areas of solitary pulmonary with the diameter < 3 cm performed on the X-line. The standards of the diagnosis of lung cancer by CT: there is a blood supply in the malignant tumor, while the benign tumor has lower or no blood supply; the malignant tumor has an irregular shape and sublobe with burrs on the edge.

Peer review

In the manuscript Lu *et al* present the results of an analysis of cost-effectiveness of various strategies for the diagnosis of SPN in China. They have assessed the cost-effectiveness of four strategies for the management of SPN: CT alone, CT plus CT-guided ACNB, CT plus PET/CT, CT plus DWI plus PET/CT. The English of the text is of high quality, clear and easy to follow. The tables and the figures are clear and self explanatory. The statistical approach meets the current standards.

REFERENCES

1 **Zhang L**, Wang M, Wang Y, Li L. [Clinico-pathological study of 98 patients with pulmonary solitary nodule]. *Zhonghua Zhong Liu Za Zhi* 2002; **24**: 491-493 [PMID: 12485507]

2 **Chen W**, Liu J, Chen Q, Li W, Xiong Z, Long X. [Bayes analysis in clinical decision-making for solitary pulmonary nodules]. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 2009; **34**: 401-405 [PMID: 19483287 DOI: 10.3736/jcim20090501]

3 **Tsushima Y**, Endo K. Analysis models to assess cost effectiveness of the four strategies for the work-up of solitary pulmonary nodules. *Med Sci Monit* 2004; **10**: MT65-MT72 [PMID: 15114278]

4 **Keith CJ**, Miles KA, Griffiths MR, Wong D, Pitman AG, Hicks RJ. Solitary pulmonary nodules: accuracy and cost-effectiveness of sodium iodide FDG-PET using Australian data. *Eur J Nucl Med Mol Imaging* 2002; **29**: 1016-1023 [PMID: 12173015 DOI: 10.1007/s00259-002-0833-2]

5 **Gambhir SS**, Shepherd JE, Shah BD, Hart E, Hoh CK, Valk PE, Emi T, Phelps ME. Analytical decision model for the cost-effective management of solitary pulmonary nodules. *J Clin Oncol* 1998; **16**: 2113-2125 [PMID: 9626211]

6 **Kosuda S**, Ichihara K, Watanabe M, Kobayashi H, Kusano S. Decision-tree sensitivity analysis for cost-effectiveness of chest 2-fluoro-2-D-[(18)F]fluorodeoxyglucose positron emission tomography in patients with pulmonary nodules (non-small cell lung carcinoma) in Japan. *Chest* 2000; **117**: 346-353 [PMID: 10669673 DOI: 10.1378/chest.117.2.346]

7 **Gould MK**, Sanders GD, Barnett PG, Rydzak CE, Maclean CC, McClellan MB, Owens DK. Cost-effectiveness of alternative management strategies for patients with solitary pulmonary nodules. *Ann Intern Med* 2003; **138**: 724-735 [PMID: 12729427]

8 **Mori T**, Nomori H, Ikeda K, Kawanaka K, Shiraishi S, Katahira K, Yamashita Y. Diffusion-weighted magnetic resonance imaging for diagnosing malignant pulmonary nodules/masses: comparison with positron emission tomography. *J Thorac Oncol* 2008; **3**: 358-364 [PMID: 18379353 DOI: 10.1097/JTO.0b013e318168d9ed]

9 **Dietlein M**, Weber K, Gandjour A, Moka D, Theissen P, Lauterbach KW, Schicha H. Cost-effectiveness of FDG-PET for the management of solitary pulmonary nodules: a decision analysis based on cost reimbursement in Germany. *Eur J Nucl Med* 2000; **27**: 1441-1456 [PMID: 11083532 DOI: 10.1007/s002590000324]

10 **Tsubamoto M**, Johkoh T, Kozuka T, Honda O, Koyama M, Murai S, Inoue A, Sumikawa H, Tomiyama N, Hamada S, Yamamoto S, Nakamura H, Kudo M. Coronal multiplanar reconstruction view from whole lung thin-section CT by multidetector-row CT: determination of malignant or benign lesions and differential diagnosis in 68 cases of solitary pulmonary nodule. *Radiat Med* 2003; **21**: 267-271 [PMID: 14743900]

11 **Siegelman SS**, Khouri NF, Leo FP, Fishman EK, Braverman RM, Zerhouni EA. Solitary pulmonary nodules: CT assessment. *Radiology* 1986; **160**: 307-312 [PMID: 3726105 DOI: 10.1148/radiology.160.2.3726105]

12 **Yi CA**, Lee KS, Kim BT, Choi JY, Kwon OJ, Kim H, Shim YM, Chung MJ. Tissue characterization of solitary pulmonary nodule: comparative study between helical dynamic CT and integrated PET/CT. *J Nucl Med* 2006; **47**: 443-450 [PMID: 16513614]

13 **Kim SK**, Allen-Auerbach M, Goldin J, Fueger BJ, Dahlbom M, Brown M, Czernin J, Schiepers C. Accuracy of PET/CT in characterization of solitary pulmonary lesions. *J Nucl Med* 2007; **48**: 214-220 [PMID: 17268017]

14 **Satoh S**, Ohdama S, Matsubara O, Okochi Y, Tanaka R, Kimula Y. CT-guided automated cutting needle biopsy by a combined method for accurate specific diagnosis of focal lung lesions. *Radiat Med* 2005; **23**: 30-36 [PMID: 15786749]

15 **Tsukada H**, Satou T, Iwashima A, Souma T. Diagnostic accuracy of CT-guided automated needle biopsy of lung nodules. *AJR Am J Roentgenol* 2000; **175**: 239-243 [PMID: 10882279 DOI: 10.2214/ajr.175.1.1750239]

16 **Takashima S**, Sone S, Li F, Maruyama Y, Hasegawa M, Kadoya M. Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions. *AJR Am J Roentgenol* 2003; **180**: 1255-1263 [PMID: 12704034 DOI: 10.2214/ajr.180.5.1801255]

17 **Klein JS**, Salomon G, Stewart EA. Transthoracic needle biopsy with a coaxially placed 20-gauge automated cutting needle: results in 122 patients. *Radiology* 1996; **198**: 715-720 [PMID: 8628859 DOI: 10.1148/radiology.198.3.8628859]

18 **Li H**, Boiselle PM, Shepard JO, Trotman-Dickenson B, McLoud TC. Diagnostic accuracy and safety of CT-guided percutaneous needle aspiration biopsy of the lung: comparison of small and large pulmonary nodules. *AJR Am J Roentgenol* 1996; **167**: 105-109 [PMID: 8659351 DOI: 10.2214/ajr.167.1.8659351]

P- Reviewer: Komócsi A **S- Editor:** Song XX
L- Editor: O'Neill M **E- Editor:** Lu YJ

World Journal of Meta-Analysis

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
www.wjgnet.com

World J Meta-Anal 2014 August 26; 2(3): I-V
ISSN 2308-3840 (online)
© 2014 Baishideng Publishing Group Inc. All rights reserved.

INSTRUCTIONS TO AUTHORS

GENERAL INFORMATION

World Journal of Meta-Analysis (*World J Meta-Anal*, *WJMA*, online ISSN 2308-3840, DOI: 10.13105) is a peer-reviewed open access (OA) academic journal that aims to guide clinical practice and improve diagnostic and therapeutic skills of clinicians, with a specific focus on meta-analysis, systematic review, mixed-treatment comparison, meta-regression, overview of reviews.

Aim and scope

WJMA covers a variety of clinical medical fields including allergy, anesthesiology, cardiac medicine, clinical genetics, clinical neurology, critical care, dentistry, dermatology, emergency medicine, endocrinology, family medicine, gastroenterology and hepatology, geriatrics and gerontology, hematology, immunology, infectious diseases, internal medicine, obstetrics and gynecology, oncology, ophthalmology, orthopedics, otolaryngology, pathology, pediatrics, peripheral vascular disease, psychiatry, radiology, rehabilitation, respiratory medicine, rheumatology, surgery, toxicology, transplantation, and urology and nephrology, while maintaining its unique dedication to systematic reviews and meta-analyses.

We encourage authors to submit their manuscripts to *WJMA*. We will give priority to manuscripts that are supported by major national and international foundations and those that are of great basic and clinical significance.

WJMA is edited and published by Baishideng Publishing Group (BPG). BPG has a strong professional editorial team composed of science editors, language editors and electronic editors. BPG currently publishes 43 OA clinical medical journals, including 42 in English, has a total of 15471 editorial board members or peer reviewers, and is a world first-class publisher.

Columns

The columns in the issues of *WJMA* will include: (1) Editorial: The editorial board members are invited to make comments on an important topic in their field in terms of its current research status and future directions to lead the development of this discipline; (2) Frontier: The editorial board members are invited to select a highly cited cutting-edge original paper of his/her own to summarize major findings, the problems that have been resolved and remain to be resolved, and future research directions to help readers understand his/her important academic point of view and future research directions in the field; (3) Field of Vision: The editorial board members are invited to write commentaries on classic articles, hot topic articles, or latest articles to keep readers at the forefront of research and increase their levels of clinical research. Classic articles refer to papers that are included in Web of Knowledge and have received a large number of citations (ranking in the top 1%) after being published for more than 2 years, reflecting the quality and impact of papers. Hot topic articles refer to papers that are included in Web of Knowledge and have received a large number of citations after being published for no more than 2 years, reflecting cutting-edge trends in scientific research. Latest articles refer to the latest published high-quality papers that are included in PubMed, reflecting the latest research trends. These commentary articles should focus on the status quo of research, the most important research topics, the problems that have now been resolved and remain to be resolved, and future research directions. Basic information about the article to be commented will be provided as well (including authors, article title, journal name, year, volume, and inclusive page numbers; (4) Minireviews: The editorial board members are invited to write short reviews on recent advances and trends in research to provide readers; (5) Review: To make a systematic review to focus on the status quo of research, the most important research topics, the problems that have now been resolved and remain to be resolved, and future research directions; (6) Topic Highlight: The editorial board members are invited to write a series of articles (7-10 articles) to comment and discuss a hot topic; (7) Meta-Analysis: Covers the systematic review, mixed-treatment comparison, meta-regression, and overview of reviews. To summarize a given quantitative effect, *e.g.*, the clinical effectiveness and safety of clinical treatments by combining data from two or more randomized controlled trials, thereby providing more precise and externally valid estimates than those which would stem from each individual dataset if analyzed separately from the others; (8) Medical Ethics: The editorial board members are invited to write articles about medical ethics to increase readers' knowledge of medical ethics. The topic covers international ethics guidelines, animal studies, clinical trials, organ transplantation, *etc.*; (9) Letters to the Editor: To discuss and make reply to the contributions published in *WJMA*, or to introduce and comment on a controversial issue of general interest; (10) Book Reviews: To introduce and comment on quality monographs; and (11) Autobiography: The editorial board members are invited to write their autobiography to provide readers with stories of success or failure in their scientific research career. The topic covers their basic personal information and information about when they started doing research work, where and how they did research work, what they have achieved, and their lessons from success or failure.

Name of journal

World Journal of Meta-Analysis

ISSN

ISSN 2308-3840 (online)

Launch date

May 26, 2013

Frequency

Quarterly

Editors-in-Chief

Giuseppe Biondi-Zoccai, MD, Assistant Professor, Department of Medico-Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina 04100, Italy

Editorial office

Jin-Lei Wang, Director
Xiu-Xia Song, Vice Director
World Journal of Meta-Analysis
Room 903, Building D, Ocean International Center,
No. 62 Dongsihuan Zhonglu, Chaoyang District,
Beijing 100025, China
Telephone: +86-10-85381891
Fax: +86-10-85381893

E-mail: editorialoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com

Publisher
Baishideng Publishing Group Inc
8226 Regency Drive,
Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com

Instructions to authors
Full instructions are available online at http://www.wjgnet.com/2308-3840/g_info_20100722180909.htm.

Indexed and Abstracted in
Digital Object Identifier.

SPECIAL STATEMENT

All articles published in journals owned by the BPG represent the views and opinions of their authors, and not the views, opinions or policies of the BPG, except where otherwise explicitly indicated.

Biostatistical editing
Statistical review is performed after peer review. We invite an expert in Biomedical Statistics to evaluate the statistical method used in the paper, including *t*-test (group or paired comparisons), chi-squared test, Ridit, probit, logit, regression (linear, curvilinear, or stepwise), correlation, analysis of variance, analysis of covariance, *etc*. The reviewing points include: (1) Statistical methods should be described when they are used to verify the results; (2) Whether the statistical techniques are suitable or correct; (3) Only homogeneous data can be averaged. Standard deviations are preferred to standard errors. Give the number of observations and subjects (*n*). Losses in observations, such as drop-outs from the study should be reported; (4) Values such as ED50, LD50, IC50 should have their 95% confidence limits calculated and compared by weighted probit analysis (Bliss and Finney); and (5) The word 'significantly' should be replaced by its synonyms (if it indicates extent) or the *P* value (if it indicates statistical significance).

Conflict-of-interest statement
In the interests of transparency and to help reviewers assess any potential bias, *WJMA* requires authors of all papers to declare any competing commercial, personal, political, intellectual, or religious interests in relation to the submitted work. Referees are also asked to indicate any potential conflict they might have reviewing a particular paper. Before submitting, authors are suggested to read "Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Ethical Considerations in the Conduct and Reporting of Research: Conflicts of Interest" from International Committee of Medical Journal Editors (ICMJE), which is available at: http://www.icmje.org/ethical_4conflicts.html.

Sample wording: [Name of individual] has received fees for serving as a speaker, a consultant and an advisory board member for [names of organizations], and has received research funding from [names of organization]. [Name of individual] is an employee of [name of organization]. [Name of individual] owns stocks and shares in [name of organization]. [Name of individual] owns patent [patent identification and brief description].

Statement of informed consent
Manuscripts should contain a statement to the effect that all human studies have been reviewed by the appropriate ethics committee or it should be stated clearly in the text that all persons gave their informed consent prior to their inclusion in the study. Details that might disclose the identity of the subjects under study should be omitted. Authors should also draw attention to the Code of Ethics of the World Medical Association (Declaration of Helsinki, 1964, as revised in 2004).

Statement of human and animal rights
When reporting the results from experiments, authors should follow the highest standards and the trial should conform to Good Clinical Practice (for example, US Food and Drug Administration Good Clinical Practice in FDA-Regulated Clinical Trials; UK Medicines Research Council Guidelines for Good Clinical Practice in Clinical Trials) and/or the World Medical Association Declaration of Helsinki. Generally, we suggest authors follow the lead investigator's national standard. If doubt exists whether the research was conducted in accordance with the above standards, the authors must explain the rationale for their approach and demonstrate that the institutional review body explicitly approved the doubtful aspects of the study.

Before submitting, authors should make their study approved by the relevant research ethics committee or institutional review board. If human participants were involved, manuscripts must be accompanied by a statement that the experiments were undertaken with the understanding and appropriate informed consent of each. Any personal item or information will not be published without explicit consents from the involved patients. If experimental animals were used, the materials and methods (experimental procedures) section must clearly indicate that appropriate measures were taken to minimize pain or discomfort, and details of animal care should be provided.

SUBMISSION OF MANUSCRIPTS

Manuscripts should be typed in 1.5 line spacing and 12 pt. Book Antiqua with ample margins. Number all pages consecutively, and start each of the following sections on a new page: Title Page, Abstract, Introduction, Materials and Methods, Results, Discussion, Acknowledgements, References, Tables, Figures, and Figure Legends. Neither the editors nor the publisher are responsible for the opinions expressed by contributors. Manuscripts formally accepted for publication become the permanent property of BPG, and may not be reproduced by any means, in whole or in part, without the written permission of both the authors and the publisher. We reserve the right to copy-edit and put onto our website accepted manuscripts. Authors should follow the relevant guidelines for the care and use of laboratory animals of their institution or national animal welfare committee. For the sake of transparency in regard to the performance and reporting of clinical trials, we endorse the policy of the ICMJE to refuse to publish papers on clinical trial results if the trial was not recorded in a publicly-accessible registry at its outset. The only register now available, to our knowledge, is http://www.clinicaltrials.gov sponsored by the United States National Library of Medicine and we encourage all potential contributors to register with it. However, in the case that other registers become available you will be duly notified. A letter of recommendation from each author's organization should be provided with the contributed article to ensure the privacy and secrecy of research is protected.

Authors should retain one copy of the text, tables, photographs and illustrations because rejected manuscripts will not be returned to the author(s) and the editors will not be responsible for loss or damage to photographs and illustrations sustained during mailing.

Online submissions
Manuscripts should be submitted through the Online Submission System at: http://www.wjgnet.com/esps/. Authors are highly recommended to consult the ONLINE INSTRUCTIONS TO AUTHORS (http://www.wjgnet.com/2308-3840/g_info_20100722180909.htm) before attempting to submit online. For assistance, authors encountering problems with the Online Submission System may send an email describing the problem to bpgoffice@wjgnet.com, or by telephone: +86-10-85381892. If you submit your manuscript online, do not make a postal contribution. Repeated online submission for the same manuscript is strictly prohibited.

MANUSCRIPT PREPARATION

All contributions should be written in English. All articles must be submitted using word-processing software. All submissions must be typed in 1.5 line spacing and 12 pt. Book Antiqua with ample margins. Style should conform to our house format. Required information for each of the manuscript sections is as follows:

Title page

Title: Title should be less than 12 words.

Running title: A short running title of less than 6 words should be provided.

Authorship: Authorship credit should be in accordance with the standard proposed by ICMJE, based on (1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published. Authors should meet conditions 1, 2, and 3.

Institution: Author names should be given first, then the complete name of institution, city, province and postcode. For example, Xu-Chen Zhang, Li-Xin Mei, Department of Pathology, Chengde Medical College, Chengde 067000, Hebei Province, China. One author may be represented from two institutions, for example, George Sgourakis, Department of General, Visceral, and Transplantation Surgery, Essen 45122, Germany; George Sgourakis, 2nd Surgical Department, Korgialenio-Benakio Red Cross Hospital, Athens 15451, Greece

Author contributions: The format of this section should be: Author contributions: Wang CL and Liang L contributed equally to this work; Wang CL, Liang L, Fu JF, Zou CC, Hong F and Wu XM designed the research; Wang CL, Zou CC, Hong F and Wu XM performed the research; Xue JZ and Lu JR contributed new reagents/analytic tools; Wang CL, Liang L and Fu JF analyzed the data; and Wang CL, Liang L and Fu JF wrote the paper.

Supportive foundations: The complete name and number of supportive foundations should be provided, *e.g.*, Supported by National Natural Science Foundation of China, No. 30224801

Correspondence to: Only one corresponding address should be provided. Author names should be given first, then author title, affiliation, the complete name of institution, city, postcode, province, country, and email. All the letters in the email should be in lower case. A space interval should be inserted between country name and email address. For example, Montgomery Bissell, MD, Professor of Medicine, Chief, Liver Center, Gastroenterology Division, University of California, Box 0538, San Francisco, CA 94143, United States. montgomery.bissell@ucsf.edu

Telephone and fax: Telephone and fax should consist of +, country number, district number and telephone or fax number, *e.g.*, Telephone: +86-10-85381892 Fax: +86-10-85381893

Peer reviewers: All articles received are subject to peer review. Normally, three experts are invited for each article. Decision on acceptance is made only when at least two experts recommend publication of an article. All peer-reviewers are acknowledged on Express Submission and Peer-review System website.

Abstract

There are unstructured abstracts (no less than 200 words) and structured abstracts. The specific requirements for structured abstracts are as follows:

An informative, structured abstract should accompany each manuscript. Abstracts of original contributions should be structured into the following sections: AIM (no more than 20 words; Only the purpose of the study should be included. Please write the Aim in the form of "To investigate/study/…"), METHODS (no less than 140 words for Original Articles; and no less than 80 words for Brief Articles), RESULTS (no less than 150 words for Original Articles and no less than 120 words for Brief Articles; You should present *P* values where appropriate and must provide relevant data to illustrate how they were obtained, *e.g.*, 6.92 ± 3.86 *vs* 3.61 ± 1.67, $P < 0.001$), and CONCLUSION (no more than 26 words).

Key words

Please list 5-10 key words, selected mainly from *Index Medicus*, which reflect the content of the study.

Core tip

Please write a summary of less than 100 words to outline the most innovative and important arguments and core contents in your paper to attract readers.

Text

For articles of these sections, original articles and brief articles, the main text should be structured into the following sections: INTRODUCTION, MATERIALS AND METHODS, RESULTS and DISCUSSION, and should include appropriate Figures and Tables. Data should be presented in the main text or in Figures and Tables, but not in both.

Illustrations

Figures should be numbered as 1, 2, 3, *etc.*, and mentioned clearly in the main text. Provide a brief title for each figure on a separate page. Detailed legends should not be provided under the figures. This part should be added into the text where the figures are applicable. Keeping all elements compiled is necessary in line-art image. Scale bars should be used rather than magnification factors, with the length of the bar defined in the legend rather than on the bar itself. File names should identify the figure and panel. Avoid layering type directly over shaded or textured areas. Please use uniform legends for the same subjects. For example: Figure 1 Pathological changes in atrophic gastritis after treatment. A: ...; B: ...; C: ...; D: ...; E: ...; F: ...; G: …*etc.* It is our principle to publish high resolution-figures for the E-versions.

Tables

Three-line tables should be numbered 1, 2, 3, *etc.*, and mentioned clearly in the main text. Provide a brief title for each table. Detailed legends should not be included under tables, but rather added into the text where applicable. The information should complement, but not duplicate the text. Use one horizontal line under the title, a second under column heads, and a third below the Table, above any footnotes. Vertical and italic lines should be omitted.

Notes in tables and illustrations

Data that are not statistically significant should not be noted. $^{a}P < 0.05$, $^{b}P < 0.01$ should be noted ($P > 0.05$ should not be noted). If there are other series of *P* values, $^{c}P < 0.05$ and $^{d}P < 0.01$ are used. A third series of *P* values can be expressed as $^{e}P < 0.05$ and $^{f}P < 0.01$. Other notes in tables or under illustrations should be expressed as ^{1}F, ^{2}F, ^{3}F; or sometimes as other symbols with a superscript (Arabic numerals) in the upper left corner. In a multi-curve illustration, each curve should be labeled with ●, ○, ■, □, ▲, △, *etc.*, in a certain sequence.

Acknowledgments

Brief acknowledgments of persons who have made genuine contributions to the manuscript and who endorse the data and conclusions should be included. Authors are responsible for obtaining written permission to use any copyrighted text and/or illustrations.

REFERENCES

Coding system

The author should number the references in Arabic numerals according to the citation order in the text. Put reference numbers in square brackets in superscript at the end of citation content or after the cited author's name. For citation content which is part of the narration, the coding number and square brackets should be typeset normally. For example, "Crohn's disease (CD) is associated with increased intestinal permeability[1,2]". If references are cited directly

in the text, they should be put together within the text, for example, "From references[19,22-24], we know that..."

When the authors write the references, please ensure that the order in text is the same as in the references section, and also ensure the spelling accuracy of the first author's name. Do not list the same citation twice.

PMID and DOI

Pleased provide PubMed citation numbers to the reference list, *e.g.*, PMID and DOI, which can be found at http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed and http://www.crossref.org/SimpleTextQuery/, respectively. The numbers will be used in E-version of this journal.

Style for journal references

Authors: the name of the first author should be typed in bold-faced letters. The family name of all authors should be typed with the initial letter capitalized, followed by their abbreviated first and middle initials. (For example, Lian-Sheng Ma is abbreviated as Ma LS, Bo-Rong Pan as Pan BR). The title of the cited article and italicized journal title (journal title should be in its abbreviated form as shown in PubMed), publication date, volume number (in black), start page, and end page [PMID: 11819634 DOI: 10.3748/wjg.13.5396].

Style for book references

Authors: the name of the first author should be typed in bold-faced letters. The surname of all authors should be typed with the initial letter capitalized, followed by their abbreviated middle and first initials. (For example, Lian-Sheng Ma is abbreviated as Ma LS, Bo-Rong Pan as Pan BR) Book title. Publication number. Publication place: Publication press, Year: start page and end page.

Format

Journals

English journal article (list all authors and include the PMID where applicable)

1 **Jung EM**, Clevert DA, Schreyer AG, Schmitt S, Rennert J, Kubale R, Feuerbach S, Jung F. Evaluation of quantitative contrast harmonic imaging to assess malignancy of liver tumors: A prospective controlled two-center study. *World J Gastroenterol* 2007; **13**: 6356-6364 [PMID: 18081224 DOI: 10.3748/wjg.13.6356]

Chinese journal article (list all authors and include the PMID where applicable)

2 **Lin GZ**, Wang XZ, Wang P, Lin J, Yang FD. Immunologic effect of Jianpi Yishen decoction in treatment of Pixu-diarrhoea. *Shijie Huaren Xiaohua Zazhi* 1999; **7**: 285-287

In press

3 **Tian D**, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in Arabidopsis. *Proc Natl Acad Sci USA* 2006; In press

Organization as author

4 **Diabetes Prevention Program Research Group**. Hypertension, insulin, and proinsulin in participants with impaired glucose tolerance. *Hypertension* 2002; **40**: 679-686 [PMID: 12411462 PMCID:2516377 DOI:10.1161/01.HYP.0000035706.28494.09]

Both personal authors and an organization as author

5 **Vallancien G**, Emberton M, Harving N, van Moorselaar RJ; Alf-One Study Group. Sexual dysfunction in 1, 274 European men suffering from lower urinary tract symptoms. *J Urol* 2003; **169**: 2257-2261 [PMID: 12771764 DOI:10.1097/01.ju.0000067940.76090.73]

No author given

6 21st century heart solution may have a sting in the tail. *BMJ* 2002; **325**: 184 [PMID: 12142303 DOI:10.1136/bmj.325.7357.184]

Volume with supplement

7 **Geraud G**, Spierings EL, Keywood C. Tolerability and safety of frovatriptan with short- and long-term use for treatment of migraine and in comparison with sumatriptan. *Headache* 2002; **42** Suppl 2: S93-99 [PMID: 12028325 DOI:10.1046/j.1526-4610.42.s2.7.x]

Issue with no volume

8 **Banit DM**, Kaufer H, Hartford JM. Intraoperative frozen section analysis in revision total joint arthroplasty. *Clin Orthop Relat Res* 2002; (**401**): 230-238 [PMID: 12151900 DOI:10.1097/00003086-200208000-00026]

No volume or issue

9 Outreach: Bringing HIV-positive individuals into care. *HRSA Careaction* 2002; 1-6 [PMID: 12154804]

Books

Personal author(s)

10 **Sherlock S**, Dooley J. Diseases of the liver and billiary system. 9th ed. Oxford: Blackwell Sci Pub, 1993: 258-296

Chapter in a book (list all authors)

11 **Lam SK**. Academic investigator's perspectives of medical treatment for peptic ulcer. In: Swabb EA, Azabo S. Ulcer disease: investigation and basis for therapy. New York: Marcel Dekker, 1991: 431-450

Author(s) and editor(s)

12 **Breedlove GK**, Schorfheide AM. Adolescent pregnancy. 2nd ed. Wieczorek RR, editor. White Plains (NY): March of Dimes Education Services, 2001: 20-34

Conference proceedings

13 **Harnden P**, Joffe JK, Jones WG, editors. Germ cell tumours V. Proceedings of the 5th Germ cell tumours Conference; 2001 Sep 13-15; Leeds, UK. New York: Springer, 2002: 30-56

Conference paper

14 **Christensen S**, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, Lutton E, Miller J, Ryan C, Tettamanzi AG, editors. Genetic programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3-5; Kinsdale, Ireland. Berlin: Springer, 2002: 182-191

Electronic journal (list all authors)

15 Morse SS. Factors in the emergence of infectious diseases. Emerg Infect Dis serial online, 1995-01-03, cited 1996-06-05; 1(1): 24 screens. Available from: URL: http://www.cdc.gov/ncidod/eid/index.htm

Patent (list all authors)

16 **Pagedas AC**, inventor; Ancel Surgical R&D Inc., assignee. Flexible endoscopic grasping and cutting device and positioning tool assembly. United States patent US 20020103498. 2002 Aug 1

Statistical data

Write as mean ± SD or mean ± SE.

Statistical expression

Express *t* test as *t* (in italics), *F* test as *F* (in italics), chi square test as χ^2 (in Greek), related coefficient as *r* (in italics), degree of freedom as υ (in Greek), sample number as *n* (in italics), and probability as *P* (in italics).

Units

Use SI units. For example: body mass, m (B) = 78 kg; blood pressure, p (B) = 16.2/12.3 kPa; incubation time, t (incubation) = 96 h, blood glucose concentration, c (glucose) 6.4 ± 2.1 mmol/L; blood CEA mass concentration, p (CEA) = 8.6 24.5 μg/L; CO_2 volume fraction, 50 mL/L CO_2, not 5% CO_2; likewise for 40 g/L formaldehyde, not 10% formalin; and mass fraction, 8 ng/g, *etc.* Arabic numerals such as 23, 243, 641 should be read 23 243 641.

The format for how to accurately write common units and quantums can be found at: http://www.wjgnet.com/2308-3840/g_info_20100725073806.htm.

Abbreviations

Standard abbreviations should be defined in the abstract and on first mention in the text. In general, terms should not be abbreviated unless they are used repeatedly and the abbreviation is helpful to the reader. Permissible abbreviations are listed in Units, Symbols and Abbreviations: A Guide for Biological and Medical Editors and Authors (Ed. Baron DN, 1988) published by The Royal Society of Medicine, London. Certain commonly used abbreviations, such as

DNA, RNA, HIV, LD50, PCR, HBV, ECG, WBC, RBC, CT, ESR, CSF, IgG, ELISA, PBS, ATP, EDTA, mAb, can be used directly without further explanation.

Italics

Quantities: *t* time or temperature, *c* concentration, *A* area, *l* length, *m* mass, *V* volume.
Genotypes: *gyrA*, *arg* 1, *c myc*, *c fos*, *etc.*
Restriction enzymes: *Eco*RI, *Hin*dI, *Bam*HI, *Kbo* I, *Kpn* I, *etc.*
Biology: *H. pylori*, *E coli*, *etc.*

Examples for paper writing

All types of articles' writing style and requirement will be found in the link: http://www.wjgnet.com/esps/NavigationInfo.aspx?id=15

RESUBMISSION OF THE REVISED MANUSCRIPTS

Authors must revise their manuscript carefully according to the revision policies of BPG. The revised version, along with the signed copyright transfer agreement, responses to the reviewers, and English language Grade A certificate (for non-native speakers of English), should be submitted to the online system *via* the link contained in the e-mail sent by the editor. If you have any questions about the revision, please send e-mail to esps@wjgnet.com.

Language evaluation

The language of a manuscript will be graded before it is sent for revision. (1) Grade A: priority publishing; (2) Grade B: minor language polishing; (3) Grade C: a great deal of language polishing needed; and (4) Grade D: rejected. Revised articles should reach Grade A.

Copyright assignment form

Please download a Copyright assignment form from http://www.wjgnet.com/2308-3840/g_info_20100725073726.htm.

Responses to reviewers

Please revise your article according to the comments/suggestions provided by the reviewers. The format for responses to the reviewers' comments can be found at: http://www.wjgnet.com/2308-3840/g_info_20100725073445.htm.

Proof of financial support

For papers supported by a foundation, authors should provide a copy of the approval document and serial number of the foundation.

Statement about anonymous publication of the peer reviewers' comments

In order to increase the quality of peer review, push authors to carefully revise their manuscripts based on the peer reviewers' comments, and promote academic interactions among peer reviewers, authors and readers, we decide to anonymously publish the reviewers' comments and author's responses at the same time the manuscript is published online.

PUBLICATION FEE

WJMA is an international, peer-reviewed, OA online journal. Articles published by this journal are distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium and format, provided the original work is properly cited. The use is non-commercial and is otherwise in compliance with the license. Authors of accepted articles must pay a publication fee. Publication fee: 698 USD per article. All invited articles are published free of charge.

Published by **Baishideng Publishing Group Inc**
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com

© 2014 Baishideng Publishing Group Inc. All rights reserved.