HEALTH RESEARCH METHODOLOGY IN SPINE SURGERY

HEALTH RESEARCH METHODOLOGY IN SPINE SURGERY

By

NATHAN EVANIEW, B.Sc. (Hons), M.D.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment

of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2016)

Hamilton, Ontario (Health Research Methodology)

TITLE: Health Research Methodology in Spine Surgery

AUTHOR: Nathan Evaniew, B.Sc. (Hons), M.D.

SUPERVISOR: Distinguished Professor Gordon Guyatt

NUMBER OF PAGES: xvi, 161

ABSTRACT

Symptomatic spinal disorders affect a large proportion of the population and are associated with substantial morbidity, social burden, and economic impact. Spine surgery interventions can provide excellent results in carefully selected patients whose symptoms fail to improve with non-operative management, but an evidence-based approach is paramount to optimize outcomes and rigorous standards of health research methodology are critical to avoid misleading conclusions. This thesis aimed to investigate and apply modern innovations in health research methodology to the field of spine surgery. It consists of seven chapters divided between three sections: randomized controlled trials, observational studies, and systematic reviews and meta-analyses. By applying the findings of each chapter, clinicians, researchers, and other evidence users can advance the credibility of future research and enhance the care of patients with spinal disorders.

ACKNOWLEDGEMENTS

I am forever indebted to an outstanding group of mentors, collaborators, and supporters: Distinguished Prof. Gordon Guyatt, Prof. Mohit Bhandari, and Dr. Michelle Ghert for their extraordinary guidance and thoughtful supervision; Carly Files, Christopher Smith, Michael Walsh, PJ Devereaux, Alonso Carrasco-Labra, Kari Tikkinen, Yutong Fei, Vanessa Noonan, Nader Fallah, Carly Rivers, Marcel Dvorak, The Rick Hansen Spinal Cord Injury Registry Network, Tahira Devji, Brian Drew, Devin Peterson, Leon van der Watt, Ilyas Aleem, Emilie Belley-Côté, and Moin Khan, for their collaboration and contributions to each chapter; the Department of Surgery, Bradley Petrisor and the Division of Orthopaedics, Clive Kearon and the Clinical Investigator Program, Sheila Singh and the Surgeon Scientist Program, and the Health Research Methodology Program at McMaster University for providing the opportunity to embark on this journey; the Michael G. DeGroote fellowship awards program, the Canadian Institutes of Health Research Doctoral Research Awards program, the Hamilton Health Sciences New Investigator Fund, and Physicians' Services Incorporated for funding and research grants; my parents Allen and Susan, and the rest of my extended family for their ongoing support; and my loving wife Allison for her enthusiasm and companionship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACAS – Asymptomatic Carotid Artery Surgery

AIS – ASIA Impairment Scale

APPAC - The Appendicitis Acuta trial

ASIA – American Spinal Injury Association

BMP – Bone Morphogenic Protein

CABG – Coronary Artery Bypass Grafting

CAC – Central Adjudication Committee

CI – Confidence Interval

CORONARY – CABG Off or On Pump Revascularization Study

EMBASE – Excerpta Medica Database

FDA – United States Food and Drug Administration

GRADE – Grading of Recommendations Assessment, Development, and Evaluation

ICC – Intraclass Correlation Coefficient

ICD-10 - International Statistical Classification of Diseases and Related Health Problems, 10^{th} revision

IQR – Interquartile Range

ISNCSCI - International Standards for Neurological Classification of Spinal Cord Injury

LEMS – Lower Extremity Motor Scores

MD – Mean Difference

MEDLINE – Medical Literature Analysis and Retrieval System Online

MeSH – Medical Subject Headings

MINORS – Methodological Index for Non-Randomized Studies

MPS – Methylprednisolone

NASCIS-II – Second National Spinal Cord Injury Study

NR – Not Reported

OBS – Observational studies

OR – Odds Ratio

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROFHER – Proximal Fracture of the Humerus Evaluation by Randomization

RCT – Randomized Controlled Trial

RD – Risk Difference

RHSCIR – Rick Hansen Spinal Cord Injury Registry

RR – Risk Ratio; Relative Risk

SD – Standard Deviation

SF-12 – Short-form 12

SSI – Surgical Site Infection

STASCIS – Surgical Timing in Acute Spinal Cord Injury Study

TMS – Total Motor Score

TSCI – Traumatic Spinal Cord Injury

UEMS – Upper Extremity Motor Score

UTI – Urinary Tract Infection

DECLARATION OF ACADEMIC ACHIEVEMENT

This is a “sandwich” thesis that consists of seven chapters, six of which have been published in peer-reviewed medical journals (chapters 1, 3-7), and one of which has been submitted to a peer-reviewed medical journal but not yet accepted (chapter 2). I was the first author and main contributor for each chapter. My independent contributions involved leading each chapter’s conception and/or design; performing acquisition, analysis, and/or interpretation of data; drafting the manuscripts and revising them critically for important intellectual content; providing final approval of the versions published or submitted; and agreeing to be accountable for all aspects of each work. My co-authors also made important contributions to each chapter related to conception and/or design; acquisition, analysis, and/or interpretation of data; revising the manuscripts critically for important intellectual content; providing final approval of the versions published or submitted; and agreeing to be accountable for all aspects of each work. I completed all aspects of this thesis between July 2013 and November 2015.

INTRODUCTION

Symptomatic spinal disorders affect a large proportion of the population and are associated with substantial morbidity, social burden, and economic impact [1–5]. Spine surgery interventions can provide excellent results in carefully selected patients whose symptoms fail to improve with non-operative management, but an evidence-based approach is paramount to optimize outcomes [6–8]. Publication rates of clinical research in spine surgery have increased dramatically in recent decades [9–14] and this progress has led to important advances in knowledge, surgical skills, and patient care, but many important gaps remain. Rigorous standards of health research methodology are critical to avoid misleading conclusions and inform clinical practice [6,12,15,16].

This principal objective of this thesis was to investigate and apply modern innovations in health research methodology to the field of spine surgery. In doing so, it aimed to advance the credibility of future research, support the clinical practice of spine surgeons, and optimize the care of patients with spinal disorders. This thesis consists of seven chapters divided between three sections: randomized controlled trials, observational studies, and systematic reviews and meta-analyses.

Chapters 1 and 2 address health research methodology issues related to randomized controlled trials in spine surgery. Chapter 1 presents a systematic survey that determined the robustness of statistically significant results from RCTs of spine surgery interventions according to the Fragility Index, which is the minimum number of patients in a trial whose status would have to change from a nonevent to an event to change a statistically significant result to a non-significant result [17]. This chapter advances current knowledge by showing that statistically significant results in spine surgery RCTs are frequently fragile, that the addition of only a small number of outcome events can completely eliminate significance, and that Fragility Index values are often eclipsed by losses to follow-up. We conclude that readers should exercise caution when interpreting the findings from RCTs with low Fragility Index values.

Chapter 2 presents a Users' Guide to the interpretation and application of randomized controlled trials that address surgical and other non-pharmacological therapies [15]. Because surgical and other non-pharmacological therapies require clinicians to develop and maintain procedural expertise, and because blinding in randomized controlled trials of such therapies is often challenging, their critical appraisal raises unique issues. These issues have not been previously applied to a conceptual framework suitable for clinicians to understand. This Users' Guide, in

addressing these issues, discusses the importance of remote randomization systems, blinding, sham-controlled trials, split-body trials, expertise-based trials, and mechanistic versus practical trials.

Chapters 3 and 4 address health research methodology issues related to observational studies in spine surgery. Chapter 3 presents a propensity score-matched cohort study from a Canadian multi-center spinal cord injury registry to evaluate methylprednisolone for the treatment of patients with acute spinal cord injuries [18]. This chapter advances current knowledge by considering how patients' neurological levels of injury and the baseline severity of their impairments might be important prognostic factors and it discusses potential strengths and limitations of using propensity score matching in comparison to other approaches.

Chapter 4 presents a scoping review performed to identify knowledge gaps and direct future research in the complex field of spinal deformity surgery [13]. This chapter demonstrates the dominance of observational studies in spinal deformity research and highlights that although there exists a broad body of research to guide surgeons managing patients with scoliosis, higher-quality studies are necessary to specifically investigate surgical indications, surgical approaches, surgical techniques, and implant selection. Scoping reviews have not been implemented previously in the

spine surgery literature, so this chapter also discusses their utility and clarifies how they differ from systematic reviews and narrative reviews.

Chapters 5, 6, and 7 relate to systematic reviews and meta-analyses in spine surgery. Chapter 5 used a methodological perspective to show how systematic reviews and meta-analyses have become increasingly popular in spine surgery but are frequently conducted and presented with limited credibility [12]. We discuss that researchers can improve future meta-analyses through exhaustive literature searches, addressing possible explanations of heterogeneity, presenting results in a clinically useful manner, reproducibly selecting and assessing primary studies, addressing confidence in the pooled effect estimates, and adhering to reporting guidelines.

Chapters 6 and 7 apply concepts from chapter 5 in meta-analyses of methylprednisolone for the treatment of patients with acute spinal cord injuries and intrawound vancomycin to prevent post-operative surgical site infections, respectively [19,20]. Chapter 6 complements work from chapter 3 because it addresses the same clinical intervention and places our results in context of the totality of the literature.

Section I: Randomized Controlled Trials

Chapter 1

The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey

Reprinted from *The Spine Journal*; 15(10); Nathan Evaniew, Carly Files, Christopher Smith, Mohit Bhandari, Michelle Ghert, Michael Walsh, Philip J. Devereaux, and Gordon Guyatt; The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey; Pages 2188-97; Copyright 2015; with permission from Elsevier. doi:10.1016/j.spinee.2015.06.004.

ELSEVIER

CrossMark

The Spine Journal 15 (2015) 2188–2197

THE SPINE JOURNAL

Clinical Study

The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey

Nathan Evaniew, MD[a,b,*], Carly Files, BSc[a], Christopher Smith, MSc[a], Mohit Bhandari, MD, PhD, FRCSC[a,b], Michelle Ghert, MD, FRCSC[a], Michael Walsh, MD, MSc, PhD, FRCPC[b], Philip J. Devereaux, MD, PhD, FRCPC[b], Gordon Guyatt, MD, MSc, FRCPC[b]

[a]*Division of Orthopaedics, Department of Surgery, McMaster University, Hamilton, Ontario, Canada*
[b]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada*

Received 28 January 2015; revised 27 April 2015; accepted 1 June 2015

Abstract

BACKGROUND CONTEXT: Randomized controlled trials (RCTs) are the most trustworthy source for evaluating treatment effects, but RCTs of spine surgery interventions often produce discordant results. The Fragility Index is a novel metric to inform about the robustness of statistically significant results.

PURPOSE: The aim was to determine the robustness of statistically significant results from RCTs of spine surgery interventions.

STUDY DESIGN/SETTING: This was a systematic survey.

PATIENT SAMPLE: The sample included RCTs of spine surgery interventions.

OUTCOME MEASURES: The Fragility Index is the minimum number of patients in a trial whose status would have to change from a nonevent to an event to change a statistically significant result to a nonsignificant result. Events refer to the occurrence of any dichotomous outcome, such as successful fusion, incident fracture, adjacent segment degeneration, or achievement of a certain functional score. A small Fragility Index indicates that the statistical significance of a result hinges on only a few events, and a large Fragility Index increases one's confidence in the observed treatment effects.

METHODS: We systematically reviewed a database for evidence-based orthopedics and identified all the RCTs that reported at least one positive outcome (ie, $p<.05$). Two reviewers independently assessed eligibility and extracted data. We used the Fisher exact test to compute Fragility Index values and multivariable linear regression to evaluate potential associated factors.

RESULTS: We identified 40 eligible RCTs with a median sample size of 132 patients (interquartile range [IQR] 79–208) and a median total number of outcome events for the chosen outcome of 31 (IQR 13–63). The median Fragility Index was two (IQR 1–3), which means that adding two events to one of the trial's treatment arms eliminated its statistical significance. The Fragility Index was less than or equal to three events in 75% of the trials, and was less than or equal to the number of patients lost to follow-up in 65% of the trials. Fragility Index values correlated positively with total sample size ($r=0.35$; $p<.05$). When adjusted for losses to follow-up and risk of bias, increasing Fragility Index values were associated only with increasingly significant reported p values ($p<.01$).

CONCLUSIONS: Statistically significant results in spine surgery RCTs are frequently fragile. The addition of only a small number of outcome events can completely eliminate significance.

FDA device/drug status: Not applicable.

Author disclosures: ***NE:*** Nothing to disclose. ***CF:*** Employment: OrthoEvidence Inc.. ***CS:*** Other: OrthoEvidence (Employee). ***MB:*** Consultancy: Stryker (B). ***MG:*** Wright Medical (Consultant ad hoc—C). ***MW:*** Pfizer (D, Research grant); Isis Pharmaceuticals (Consulting, B). ***PJD:*** Nothing to disclose. ***GG:*** Nothing to disclose.

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

This study was not funded.

This study did not require local institutional ethics board approval.

* Corresponding author. Division of Orthopaedics, Department of Surgery, McMaster University, Hamilton, Ontario, Canada. Tel.: (905) 521-4322x44992; fax: (905) 523-8781.

E-mail address: nathan.evaniew@medportal.ca (N. Evaniew)

http://dx.doi.org/10.1016/j.spinee.2015.06.004

Surgeons, researchers, and other evidence users should exercise caution when interpreting the findings from RCTs with low Fragility Index values and applying these results to patient care.

Keywords: Spine surgery; Randomized controlled trials; Outcomes; Statistical significance; Fragility Index; Clinical epidemiology

Introduction

The most trustworthy results for addressing the impact of treatment effects and establishing causality come from rigorously conducted and adequately powered randomized controlled trials (RCTs), but RCTs of spine surgery interventions often provide discordant results [1]. Although each of study quality, sample size, and conflicts of interest have been previously explored as potential associated factors, little attention has focused on the importance of the number of outcome events in each arm [2–6]. Trials with small numbers of outcome events are at risk of producing implausibly large treatment effects, particularly when their sample sizes are also small [7,8].

The Fragility Index was recently developed as a novel metric to describe the robustness of statistically significant results, and it is intended to complement p values and 95% confidence intervals (CIs) [9]. The Fragility Index for a given study is defined as the minimum number of patients in the trial group with fewer outcome events whose status would have to change from a nonevent to an event to change a statistically significant result to a nonsignificant result. Events refer to the occurrence of any dichotomous outcome, such as successful fusion, incident fracture, adjacent segment degeneration, or achievement of a certain functional score. A small Fragility Index indicates that the statistical significance of a result hinges on only a few events, and a large Fragility Index increases one's confidence in the observed treatment effects.

For example, consider an RCT in which 100 patients with symptomatic spinal stenosis were randomized to either surgical treatment with a minimally invasive interspinous spacer or conventional open decompression [10]. In this trial, 13 patients in the interspinous spacer group underwent a subsequent reoperation in comparison to three patients in the conventional open decompression group. This difference was statistically significant (p=.04), but it would have been completely insignificant if just two additional patients in the conventional open decompression group had also undergone a reoperation (p=.07). Thus, the Fragility Index for this outcome is two events.

Investigators can apply the Fragility Index to any dichotomous outcome in a 1:1 parallel design RCT, and its application does not require specialized statistical expertise. In their review of 399 randomized trials from high-impact medical journals, Walsh et al. [9] reported a median Fragility Index of eight events; 24% percent of the included trials had a Fragility Index of three or less, and 53% had a Fragility Index less than the number of patients lost to follow-up.

Given that many trials in spine surgery are characterized by small sample sizes and few events [5,6,11], our primary objective was to determine the robustness of statistically significant results in RCTs of spine surgery interventions by systematically applying the Fragility Index. Our secondary objective was to identify potential factors associated with the Fragility Index.

Methods

Eligibility criteria

We performed a systematic survey of RCTs of spine surgery interventions published from January 2009 to September 2014. We included all trials that reported in their abstract at least one statistically significant dichotomous outcome (ie, p value <.05 under a null hypothesis that no difference existed or a 95% CI that excluded a null value), were randomized according to a 1:1 parallel two-arm design, and examined a preoperative, intraoperative, or postoperative intervention in patients undergoing spine surgery.

Identification of studies

We identified potential trials using a systematic database for evidence-based orthopedics [12]. The database search strategy includes search terms that identify relevant RCTs published in orthopedic surgery journals, neurosurgery journals, and general medical journals without any language restrictions (Appendix 1). The database search strategy is executed at the beginning of each month using MEDLINE, and its contents have been kept updated from January 1, 2009 onward. We queried the database on September 11, 2014, and two reviewers independently screened the titles and abstracts of all studies using piloted electronic forms. Reviewers resolved discrepancies through discussion of the rationale for their decisions.

Assessments of risk of bias and data extraction

Two reviewers independently assessed trial-level risk of bias and extracted study outcome data using piloted electronic forms. Risk of bias was assessed using the Cochrane Collaboration Risk of Bias tool [13], which includes items

EVIDENCE & METHODS

Context
The authors present their work regarding the fragility of recent randomized controlled trials (RCTs) in spine surgery. This study was performed as a systematic review with 40 eligible studies with median samples of 132 patients.

Contribution
The authors found that, on average, an addition of only two events to a trial's treatment arm would result in a change in statistical significance. The Fragility Index was less than or equal to three events in 75% of the RCTs, while it was less than or equal to the number of patients lost to follow-up in 65% of the included trials.

Implications
The results presented here are valuable in that they show why RCTs in spine surgery may not represent best available evidence in terms of translation and clinical efficacy. Readers should recognize that the findings presented in this work are limited to the 40 studies that the authors' considered, and certainly the study's findings have no implications for forthcoming RCTs that have yet to be published.

—The Editors

for each of: sequence generation; allocation concealment; blinding of surgeons, patients, and outcome assessors; losses to follow-up and missing data; selective outcome reporting; and other bias, such as expertise bias [5,14].

For each included RCT, we extracted outcome data for one statistically significant dichotomous outcome that was identified from the abstract. For trials that reported more than one eligible outcome in their abstracts, we chose the primary outcome whenever possible, and otherwise chose the most patient-important secondary outcome according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach for distinguishing between outcomes that are critical for decision making, important but not critical, or of low importance [15].

Extracted data for each outcome included: journal name; publication year; funding source; sample size for each arm; losses to follow-up for each arm; number of events for each arm; reported p value or 95% CI; statistical test used; and whether outcomes were primary or secondary, composite, adjusted, or time-to-event. We also recorded the 2013 Thomson Reuters Journal Impact Factor and the most recent Science Citation Index for each trial. The Science Citation Index is a metric of citation frequency that reflects the cumulative number of citations to source items indexed within the Web of Science Core Collection [16].

Application of the Fragility Index

We calculated the Fragility Index for each outcome using two-by-two contingency tables according to the method described by Walsh et al. [9]. We first recalculated the p value for each outcome using the two-sided Fisher exact test. We then iteratively added events to the group with the smaller number of events while subtracting nonevents to keep the total number of patients constant. The smallest number of additional events required to obtain a p value equal to or greater than .05 represented the Fragility Index.

Statistical analysis

We report discrete variables as counts or proportions, normally distributed continuous variables as means with standard deviations, and skewed continuous variables as medians with interquartile ranges (IQRs). We quantified interobserver agreement for the reviewers' assessments of study eligibility using the Cohen kappa coefficient and interpreted kappa values according to Landis and Koch as: 0, poor; 0.01 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect [17]. We used multivariable linear regression to evaluate characteristics associated with greater fragility, with a logarithmic transformation applied to p values. We evaluated direct correlations using the Pearson correlation coefficient. All tests of significance were two-tailed, and p values less than .05 were considered significant. All analyses were performed using Microsoft Excel (Microsoft, Santa Rosa CA, USA; 2011) and SPSS version 21 (IBM, Chicago, IL, USA; 2012).

Results

Study selection

Our database query yielded 2,335 potential studies. Screening of titles and abstracts and review of full texts led to the final inclusion of 40 RCTs of spine surgery interventions (Table 1; Fig. 1). Of the 2,295 excluded studies, 149 were not randomized according to a 1:1 parallel design, and 203 had statistically significant dichotomous outcomes but were not related to spine surgery. Agreement between the two reviewers for eligibility was substantial (kappa=0.70).

Characteristics of trials and outcomes

The median sample size of the included trials was 132 (IQR 79–208), and the median total number of losses to follow-up was 5 (IQR 0–26; ie, 3.8% of the patients were lost to follow-up across trials; Table 2). The median journal impact factor was 2.4 (IQR 2.4–2.5), and the median

Table 1
Included randomized controlled trials; n=40 trials

Authors	Year	Intervention	Outcome	Sample size	Total events	Fragility Index
Bai et al. [40]	2012	Electronic conductivity device for pedicle screw placement in scoliosis	Pedicle perforation	694	62	21
Berg et al. [41]	2009	Total disc replacement versus lumbar fusion	Outcome of totally pain-free	152	35	1
Bible et al. [42]	2013	Covered versus uncovered implant trays	Implant contamination	105	10	1
Blasco et al. [43]	2012	Vertebral augmentation for osteoporotic vertebral compression fractures	Fracture	125	25	0
Burkus et al. [44]	2010	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Neurologic improvement or maintenance	541	416	5
Chen et al. [45]	2013	Titanium versus polyetheretherketone cages in cervical spondylotic myelopathy	Outcome of "excellent or good"	80	39	0
Cheng et al. [46]	2009	Posterior lumbar interbody fusion in spondylolisthesis	Complications	138	6	1
Cheng et al. [47]	2011	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Dysphagia	83	8	0
Coric et al. [48]	2011	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Overall success (composite)	269	183	3
Dimar et al. [49]	2009	rhBMP-2 For instrumented posterolateral lumbar fusion	Fusion	463	337	2
Engquist et al. [50]	2013	Surgery versus nonsurgical treatment of cervical radiculopathy	Outcome of "better/much better"	68	47	1
Farrokhi et al. [51]	2011	Vertebral augmentation for osteoporotic vertebral compression fractures	Fracture	82	7	0
Gauger et al. [52]	2009	Epidural analgesia after pediatric posterior spinal fusion	Requirement of diazepam	38	24	2
Hart et al. [53]	2014	Allograft with bone marrow concentrate for posterolateral lumbar fusion	Fusion	80	48	7
He et al. [54]	2014	Pedicle screw techniques for posterolateral lumbar fusion	Adjacent segment degeneration	210	110	6
Hiller et al. [55]	2012	Acetaminophen after spine surgery in children and adolescents	Pain score≥6	36	20	0
Hurlbert et al. [56]	2013	rhBMP-2 For instrumented posterolateral lumbar fusion	Fusion	207	127	39
Jiya et al. [57]	2009	Nonresorbable versus resorbable fusion devices	Fusion	26	18	1
Kallmes et al. [58]	2009	Vertebral augmentation for osteoporotic vertebral compression fractures	Crossover between interventions	131	41	14
Klazen et al. [59]	2010	Vertebral augmentation for osteoporotic vertebral compression fractures	Further height loss	202	46	15
Korovessis et al. [60]	2013	Vertebral augmentation for osteoporotic vertebral compression fractures	Leakage	185	16	0
Löfgren et al. [61]	2010	Trabecular metal in anterior cervical fusion for degenerative disease	Fusion	80	63	2
Murrey et al. [62]	2009	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Secondary surgery	209	11	0
Nagahama et al. [63]	2011	Alendronate after posterior lumbar interbody fusion	Fusion	40	29	1
Nandyala et al. [64]	2014	Silicate-substituted calcium phosphate versus rhBMP-2	Fusion	52	41	1
O'Neill et al. [65]	2014	Bupivacaine after iliac crest bone graft harvest	Meeting expectations for surgery	40	28	2
Ohtori et al. [66]	2011	Local bone graft versus iliac crest bone graft	Presence of pain	82	6	2
Phillips et al. [67]	2013	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Overall success (composite)	416	240	1
Pitzen et al. [68]	2009	Cervical plating with dynamic versus rigid plates	Implant complications	132	4	1
Putzier et al. [69]	2009	Allogenic versus autologous cancellous bone graft	Fusion	44	13	1
Ringel et al. [70]	2012	Robot-assisted placement of lumbar and sacral pedicle screws	Good screw position	298	266	2
Roh et al. [71]	2014	Palonosetron versus ramosetron for preventing nausea and vomiting	Postoperative nausea and vomiting	196	115	3
Ruetten et al. [72]	2009	Full-endoscopic interlaminar approach for lumbar lateral recess stenosis	Complications	192	8	1
Sasso et al. [73]	2011	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Overall success (composite)	582	254	6
Sköld et al. [74]	2013	Total disc replacement versus lumbar fusion	Outcome of totally pain-free	152	41	5
Strömqvist et al. [10]	2013	Interspinous process spacer versus decompression for neurogenic claudication	Secondary surgery	100	16	2
Thalgott et al. [75]	2009	Frozen versus freeze-dried allograft in anterior lumbar interbody fusion	Pseudarthrosis	50	7	1
Wu et al. [76]	2010	Computer-assisted placement of thoracic pedicle screws	Pedicle perforation	176	18	3
Yang et al. [77]	2012	Vertebral augmentation and chemotherapy for multiple myeloma fractures	Overall response (composite)	76	44	0
Zigler et al. [78]	2014	Cervical disc arthroplasty versus anterior cervical discectomy and fusion	Secondary surgery	228	9	2

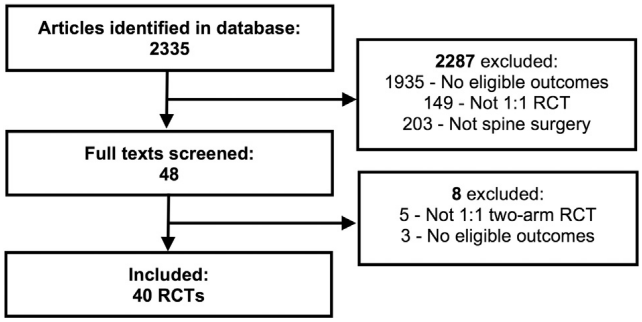


Fig. 1. Flow of articles through screening and reasons for exclusion. RCT, randomized controlled trial.

Science Citation Index was 9 (IQR 2.5–20). Among the included trials, sequence generation was at low risk of bias in 24 (60%), and allocation concealment was at low risk of bias in 15 (38%). Investigators blinded surgeons in 5 (13%), patients in 6 (15%), and outcome assessors in 17 (43%; Table 3).

Of the 40 outcomes chosen, 23 (58%) were primary and 17 (42%) were secondary. Reported p values for each outcome were less than .05 but greater than or equal to .01 for 26 (45%), less than .01 but greater than or equal to .001 for 9 (28%), and less than .001 for 5 (8%). Seven outcomes (18%) were composites, one was based on a time-to-event analysis, and none were adjusted. The median total number of events for each outcome was 31 (IQR 13–63).

Fragility Index

The median Fragility Index for the 40 evaluated outcomes was two events (IQR 1–3; Fig. 2), which means that adding two events to one of the trial's treatment arms eliminated its statistical significance. Eight outcomes (20%) had a Fragility Index of zero because they lost their statistical significance when we initially recalculated their p values using the two-sided Fisher exact test. All eight of these were originally analyzed using a chi-square test. The Fragility Index was less than or equal to three events in 30 (75%) studies and less than or equal to the total number of patients lost to follow-up for 26 (65%) outcomes.

Table 4 summarizes the median Fragility Index values according to subgroups based on outcome type, sample size, number of events, losses to follow-up, funding, and p values. Fragility Index values correlated positively with total

Table 2
Characteristics of included randomized controlled trials; n=40 trials

Characteristic	Number of studies (%)
Year of publication	
2009	11 (28)
2010	4 (10)
2011	6 (15)
2012	5 (13)
2013	8 (20)
2014	6 (15)
Journal	
Spine	15 (38)
European Spine Journal	7 (18)
Journal of Neurosurgery: Spine	5 (13)
Journal of Bone and Joint Surgery	2 (5)
Other	11 (28)
Industry funding	
Yes	12
No	17
Not reported	11

Table 3
Risk of bias of included randomized controlled trials; n=40 trials

Item	Yes (%)	No (%)	Unclear (%)
Was the allocation sequence adequately generated?	24 (60)	1 (3)	15 (38)
Was allocation adequately concealed?	15 (38)	3 (8)	22 (55)
Blinding surgeons: was knowledge of the allocated interventions adequately prevented?	5 (13)	33 (83)	2 (5)
Blinding patients: was knowledge of the allocated interventions adequately prevented?	6 (15)	20 (50)	14 (35)
Blinding outcome assessors: was knowledge of the allocated interventions adequately prevented?	17 (43)	12 (30)	11 (28)
Were losses to follow-up (missing outcome data) accounted for?	34 (85)	4 (10)	2 (5)
Are reports of the study free of suggestion of selective outcome reporting?	39 (98)	0	1 (3)
Was the study free of any other potential bias (such as expertise bias)?	28 (70)	2 (5)	10 (25)

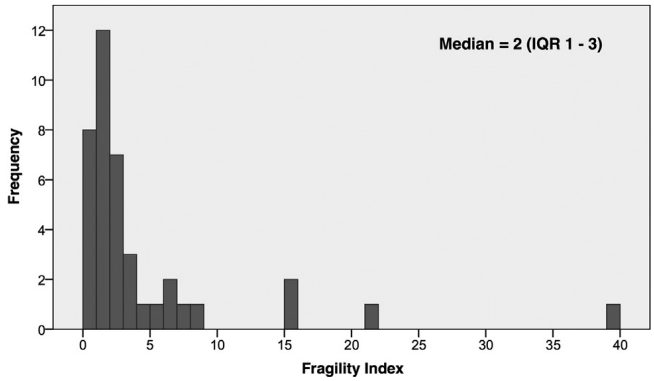


Fig. 2. Distribution of Fragility Index values from 40 trials. The median number of patients whose status would have to change from a nonevent to an event to change a statistically significant result to a nonsignificant result was 2 (IQR 1–3). IQR, interquartile range.

Table 4
Application of the Fragility Index across outcomes

Characteristic	Median Fragility Index (IQR)
All outcomes (n=40)	
Primary outcomes (n=23)	2 (1–6)
Secondary (n=17)	1 (0–2)
Composite (n=7)	2 (2–5)
Time-to-event (n=1)	1 (1–1)
Total sample size	
26 to 76 (n=10)	1 (1–2)
80 to 131 (n=10)	1 (0–2)
132 to 207 (n=10)	2 (1–5)
209 to 694 (n=10)	3 (2–6)
Total number of events	
4 to 12 (n=10)	1 (1–2)
13 to 29 (n=10)	1 (0–2)
32 to 62 (n=10)	3 (1–13)
63 to 337 (n=10)	3 (2–5)
Losses to follow-up	
<20% (n=15)	3 (1–6)
≥20% (n=25)	1 (1–2)
Industry funding	
Present (n=12)	2 (1–7)
Absent (n=28)	1 (1–3)
p Value	
<.05 to .01 (n=26)	1 (1–2)
<.01 to .001 (n=9)	6 (2–8)
<.001 (n=5)	7 (3–15)

IQR, interquartile range.

sample size (r=0.35; p<.05). However, when adjusted for losses to follow-up and risk of bias, increasing Fragility Index values were associated only with increasingly significant reported p values (p<.01; Table 5). The median Fragility Index of the five trials that reported p values less than .001 was seven events (IQR 3–15). Fragility Index values did not correlate significantly with journal impact factors, cumulative citation rates, or any of the individual component risk of bias items.

Discussion

Across 40 RCTs of spine surgery interventions, we found that adding as few as two outcome events completely eliminated the statistical significance of the trial. The number of events required to eliminate statistical significance was less than or equal to the total number of patients lost to follow-up for 65% of outcomes and was associated only with increasingly significant p values when adjusted for sample size, losses to follow-up, and risk of bias.

Table 5
Associations between characteristics of outcomes and Fragility Index. A logarithmic transformation was applied to the p values

Variable	Coefficient	95% CI—lower	95% CI—upper	Significance
Losses to follow-up	−0.02	−0.07	0.04	0.51
Risk of bias	0.75	−1.07	2.56	0.41
p Values	−2.05	−3.19	−.90	<.01
Sample size	0.01	0.00	0.03	0.07

CI, confidence interval.

Strengths and limitations

Eight outcomes had a Fragility Index of zero events because they lost their statistical significance when we recalculated their p values using the two-sided Fisher exact test. We recalculated p values using the two-sided Fisher exact test for all outcomes to allow standardized assessments of robustness independent of the statistical test chosen. The Fisher exact test is an alternative to the Pearson chi-square test for comparing proportions in a two-by-two contingency table and is more conservative than the Pearson chi-square test for most situations [18,19]. Moreover, the Fisher exact test is valid for all sample sizes, and it is the preferred approach when sample sizes are small or outcome events are uncommon (as was the case for the trials in our study).

We used an existing database of trials with a search strategy limited to MEDLINE. Hopewell et al.[20] showed that searching MEDLINE alone can miss up to 45% of eligible trials in systematic reviews, and some experts recommend the combination of MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials as a minimum search requirement for most clinical questions [13]. However, the database search strategy specifically targeted the four highest impact journals in spine research and the highest impact journals in orthopedic surgery and general medicine, and trials with larger sample sizes are most often published in higher impact journals [21]. Given that we demonstrated a correlation between sample size and Fragility Index, the potential exclusion of smaller trials from lower impact journals could have biased our results toward higher Fragility Index values.

Our small number of included trials may have resulted in underestimates of the potential associations between the Fragility Index and trial characteristics. However, this seems unlikely to be important given that our results are consistent with earlier findings in the medical literature [9].

Relation to prior work

This study is the first to report Fragility Index values from RCTs of spine surgery interventions. There are many challenges to the design and conduct of RCTs in spine surgery, and surgical trials in general have historically been characterized by small sample sizes with few outcome events [22]. We identified a median sample size of 132 patients and a median number of events per outcome of 31, both of which are substantially smaller than the median sample size of 682 patients and median number of events per outcome of 112 that Walsh et al. [9] reported in their analysis of 399 trials from high-impact medical journals. Accordingly, we found a median Fragility Index of just

two events (IQR 1–3), which is also lower than their reported median Fragility Index of eight events (3–18).

Other investigators have considered the robustness of statistical significance in RCTs using similar techniques. Concepts similar to the Fragility Index were previously published by Pocock in 1985 [23], Feinstein in 1990 [24], and Walter in 1991 [25]. More recently, Tornetta III et al. [26] incrementally added or subtracted events from one group in their review of 118 RCTs and 80 observational studies in the orthopedic fracture care literature. They found that a median of four event changes (range: 1–340) was required to lose significance, whereas a median of five (range: 1–40) was required to make previously nonsignificant results significant. We suggest that the Fragility Index as previously reported and defined is the preferable approach because it allows total sample size to remain constant, and it provides an intuitive metric than can be easily understood by researchers and clinicians [9].

Implications

Trials that evaluate spine surgery interventions are often interpreted using a hypothesis-testing framework in which the null hypothesis is that there is no difference between the interventions under investigation [1]. To demonstrate superiority for an intervention, investigators commonly attempt to disprove a null hypothesis using statistical tests. Type-I errors, or ''false-positives,'' occur when investigators incorrectly find a difference that does not really exist, and investigators must decide on the level of risk for making a Type-I error that they are willing to accept. This level of risk is conventionally set to be five times out of 100 (ie, p values $<.05$ are used to establish ''statistical significance'').

Limitations of hypothesis testing and p values include inabilities to inform the magnitude of treatment effects or identify the ranges of plausible true values consistent with the observed data [4,9,27]. The reporting of point estimates with 95% CIs can overcome some of these limitations, but even this approach does not fully address the concern that trials with smaller sample sizes and fewer outcome events may produce spurious inferences of benefit [9].

The robustness of many statistically significant results in RCTs of spine surgery interventions is limited by small sample sizes and few outcome events. Surgeons, researchers, and other evidence users should exercise caution when interpreting the findings from these trials and applying these results to patient care. Smaller p values are more likely to occur in larger studies with greater differences in rates of outcome events between groups, and they may be a marker of increased robustness, but our regression model demonstrated that a p value difference of several orders of magnitude is required to produce even a single unit change in Fragility Index values. Widespread adoption of the Fragility Index could complement p values and 95% CIs to allow easy identification of trials with less robust results.

Missing data threaten the validity of even the largest and most rigorously designed RCTs [28]. Akl et al. found a median percentage of losses to follow-up of 6% (IQR 2%–14%) in their review of 235 trials from high-impact journals, and higher losses have been observed in surgical trials with trauma patients [28–30]. The analysis of trials with missing data requires at least some non-verifiable statistical assumptions, and even plausible assumptions can change the interpretation of results [29–31]. Losses to follow-up greater than the Fragility Index further suggest the need for caution in making inferences from trials with positive results; that the total number of patients lost to follow-up exceeded the Fragility Index in 65% of outcomes indicates that this need for additional caution is frequent [28,32].

Further research is warranted to understand the potential clinical impact and practical utility of the Fragility Index. Surgeons can be confident that trials with increasing Fragility Index values are more robust than trials with lower values, but it remains unclear whether certain thresholds of acceptable robustness exist [33]. The process of translating perceptions about evidence from clinical trials to changes in clinical practice remains complex and poorly understood [34,35]. In this study, Fragility Index values were not associated with impact factors or cumulative citation rates.

Further research is also required to understand how sample size calculations for future trials might best incorporate Fragility Index estimations. Investigators should consider that both the number of participants and the number of events are relevant [7,36], but an investigators' ability to guard against a low Fragility Index in advance is limited. Even a well-powered trial, if results happen to be on the edge of statistical significance, will have a Fragility Index at or near 1. Thus, we see the concept of fragility as one that comes into play at the point of examining the results, rather than at the stage of sample size calculation. Other investigators have suggested that trials may require several thousand participants and several hundred events to adequately minimize the risk for spuriously positive or negative results [37].

Our finding of outcomes in which statistical significance was eliminated on the basis of selecting an alternative statistical test highlights the importance of conservative statistical approaches, specifying the selection of statistical tests a priori, and testing important statistical assumptions with planned sensitivity analyses [38,39].

Conclusions

The statistical significance of results from RCTs in spine surgery is frequently fragile because of relatively small sample sizes with few outcome events. The Fragility Index is a novel metric that informs the robustness of statistically significant results. Surgeons, researchers, and other

evidence users should exercise caution when interpreting the findings from trials with low Fragility Index values and applying these results to patient care. Widespread adoption of the Fragility Index could complement p values and 95% CIs to help identify trials with less robust results, and it could also help inform the potential importance of missing data.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.spinee.2015.06.004.

References

[1] Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. J Bone Joint Surg Am 2001;83-A:916–26.
[2] Bhandari M, Busse JW, Jackowski D, Montori VM, Schunemann H, Sprague S, et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ 2004;170:477–80.
[3] SPRINT InvestigatorsBhandari M, Tornetta P 3rd, Rampersad SA, Sprague S, Heels-Ansdell D, Sanders DW, et al. (Sample) size matters! an examination of sample size from the SPRINT trial study to prospectively evaluate reamed intramedullary nails in patients with tibial fractures. J Orthop Trauma 2013;27:183–8.
[4] Bhandari M, Montori VM, Schemitsch EH. The undue influence of significant p-values on the perceived importance of study results. Acta Orthop 2005;76:291–5.
[5] van Oldenrijk J, van Berkel Y, Kerkhoffs GM, Bhandari M, Poolman RW. Do authors report surgical expertise in open spine surgery related randomized controlled trials? A systematic review on quality of reporting. Spine 2013;38:857–64.
[6] Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. Spine 2004;29:1146–9.
[7] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 2010;303:1180–7.
[8] Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. JAMA 2005;294:2203–9.
[9] Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. J Clin Epidemiol 2014;67:622–8.
[10] Stromqvist BH, Berg S, Gerdhem P, Johnsson R, Moller A, Sahlstrand T, et al. X-stop versus decompressive surgery for lumbar neurogenic intermittent claudication: randomized controlled trial with 2-year follow-up. Spine 2013;38:1436–42.
[11] Bederman SS, Wright JG. Randomized trials in surgery: how far have we come? J Bone Joint Surg Am 2012;94:2–6.
[12] OE—OrthoEvidence [homepage on the Internet]. Available at: www.myorthoevidence.com. Accessed September 1, 2014.
[13] Higgins JP, Greene S, eds. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. The Cochrane Collaboration: 2011. Available at: www.cochrane-handbook.org. Accessed September 1, 2014.
[14] Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, et al. Need for expertise based randomised controlled trials. BMJ 2005;330:88.
[15] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395–400.
[16] Science citation index [homepage on the Internet]. Available at: http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=K. Accessed November 1, 2014.
[17] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.
[18] Kuhn JE, Greenfield ML, Wojtys EM. A statistics primer. Statistical tests for discrete data. Am J Sports Med 1997;25:585–6.
[19] Lydersen S, Pradhan V, Senchaudhuri P, Laake P. Choice of test for association in small sample unordered r x c tables. Stat Med 2007;26:4328–43.
[20] Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev 2007;MR000001.
[21] Bala MM, Akl EA, Sun X, Bassler D, Mertz D, Mejza F, et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. J Clin Epidemiol 2013;66:286–95.
[22] Farrokhyar F, Karanicolas PJ, Thoma A, Simunovic M, Bhandari M, Devereaux PJ, et al. Randomized controlled trials of surgical interventions. Ann Surg 2010;251:409–16.
[23] Pocock SJ. Current issues in the design and interpretation of clinical trials. Br Med J (Clin Res Ed) 1985;290:39–42.
[24] Feinstein AR. The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions. J Clin Epidemiol 1990;43:201–9.
[25] Walter SD. Statistical significance and fragility criteria for assessing a difference of two proportions. J Clin Epidemiol 1991;44:1373–8.
[26] Tornetta P III, Bhandari M, Parisien RL, Dashe J, Cronin P. In: Statistical significance in trauma research: too unstable to trust? 2014 Annual Meeting of the Orthopaedic Trauma Association: Tampa, Florida, 2014. p. 354.
[27] Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.
[28] Sprague S, Leece P, Bhandari M, Tornetta P 3rd, Schemitsch E, Swiontkowski MF, et al. Limiting loss to follow-up in a multicenter randomized trial in orthopedic surgery. Control Clin Trials 2003;24:719–25.
[29] Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ 2012;344:e2809.
[30] Zelle BA, Bhandari M, Sanchez AI, Probst C, Pape HC. Loss of follow-up in orthopaedic trauma: is 80% follow-up still acceptable? J Orthop Trauma 2013;27:177–81.
[31] Scharfstein DO, Hogan J, Herman A. On the prevention and analysis of missing data in randomized clinical trials: the state of the art. J Bone Joint Surg Am 2012;94:80–4.
[32] Kau M, Sprague S, Ignacy T, Thoma A, Bhandari M, Farrokhyar F. How to optimize participant retention and complete follow-up in surgical research. Can J Surg 2014;57:420–7.
[33] Khan H, Hussain N, Bhandari M. The influence of large clinical trials in orthopedic trauma: do they change practice? J Orthop Trauma 2013;27:e268–74.
[34] Straus SE, Tetroe J, Graham I. Defining knowledge translation. CMAJ 2009;181:165–8.
[35] Noonan VK, Wolfe DL, Thorogood NP, Park SE, Hsieh JT, Eng JJ, et al. Knowledge translation and implementation in spinal cord injury: a systematic review. Spinal Cord 2014;52:578–87.
[36] Devereaux PJ, Chan MT, Eisenach J, Schricker T, Sessler DI. The need for large clinical studies in perioperative medicine. Anesthesiology 2012;116:1169–75.
[37] Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409–22.

[38] Mbuagbaw L, Morfaw F, Kunda JE, Mukonzo JK, Kastner J, Zhang S, et al. Tips for charting the course of a successful health research career. J Multidiscip Healthc 2013;6:163–8.

[39] Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013;13:92.

[40] Bai YS, Niu YF, Chen ZQ, Zhu XD, Gabriel LK, Wong HK, et al. Comparison of the pedicle screws placement between electronic conductivity device and normal pedicle finder in posterior surgery of scoliosis. J Spinal Disord Tech 2013;26:316–20.

[41] Berg S, Tullberg T, Branth B, Olerud C, Tropp H. Total disc replacement compared to lumbar fusion: a randomised controlled trial with 2-year follow-up. Eur Spine J 2009;18:1512–9.

[42] Bible JE, O'Neill KR, Crosby CG, Schoenecker JG, McGirt MJ, Devin CJ. Implant contamination during spine surgery. Spine J 2013;13:637–40.

[43] Blasco J, Martinez-Ferrer A, Macho J, San Roman L, Pomes J, Carrasco J, et al. Effect of vertebroplasty on pain relief, quality of life, and the incidence of new vertebral fractures: a 12-month randomized follow-up, controlled trial. J Bone Miner Res 2012;27:1159–66.

[44] Burkus JK, Haid RW, Traynelis VC, Mummaneni PV. Long-term clinical and radiographic outcomes of cervical disc replacement with the prestige disc: results from a prospective randomized controlled clinical trial. J Neurosurg Spine 2010;13:308–18.

[45] Chen Y, Wang X, Lu X, Yang L, Yang H, Yuan W, et al. Comparison of titanium and polyetheretherketone (PEEK) cages in the surgical treatment of multilevel cervical spondylotic myelopathy: a prospective, randomized, control study with over 7-year follow-up. Eur Spine J 2013;22:1539–46.

[46] Cheng L, Nie L, Zhang L. Posterior lumbar interbody fusion versus posterolateral fusion in spondylolisthesis: a prospective controlled study in the Han nationality. Int Orthop 2009;33:1043–7.

[47] Cheng L, Nie L, Li M, Huo Y, Pan X. Superiority of the bryan ((R)) disc prosthesis for cervical myelopathy: a randomized study with 3-year follow up. Clin Orthop Relat Res 2011;469:3408–14.

[48] Coric D, Nunley PD, Guyer RD, Musante D, Carmody CN, Gordon CR, et al. Prospective, randomized, multicenter study of cervical arthroplasty: 269 patients from the Kineflex|C artificial disc investigational device exemption study with a minimum 2-year follow-up: clinical article. J Neurosurg Spine 2011;15:348–58.

[49] Dimar JR 2nd, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY. Clinical and radiographic analysis of an optimized rhBMP-2 formulation as an autograft replacement in posterolateral lumbar spine arthrodesis. J Bone Joint Surg Am 2009;91:1377–86.

[50] Engquist M, Lofgren H, Oberg B, Holtz A, Peolsson A, Soderlund A, et al. Surgery versus nonsurgical treatment of cervical radiculopathy: a prospective, randomized study comparing surgery plus physiotherapy with physiotherapy alone with a 2-year follow-up. Spine 2013;38:1715–22.

[51] Farrokhi MR, Alibai E, Maghami Z. Randomized controlled trial of percutaneous vertebroplasty versus optimal medical management for the relief of pain and disability in acute osteoporotic vertebral compression fractures. J Neurosurg Spine 2011;14:561–9.

[52] Gauger VT, Voepel-Lewis TD, Burke CN, Kostrzewa AJ, Caird MS, Wagner DS, et al. Epidural analgesia compared with intravenous analgesia after pediatric posterior spinal fusion. J Pediatr Orthop 2009;29:588–93.

[53] Hart R, Komzak M, Okal F, Nahlik D, Jajtner P, Puskeiler M. Allograft alone versus allograft with bone marrow concentrate for the healing of the instrumented posterolateral lumbar fusion. Spine J 2014;14:1318–24.

[54] He B, Yan L, Guo H, Liu T, Wang X, Hao D. The difference in superior adjacent segment pathology after lumbar posterolateral fusion by using 2 different pedicle screw insertion techniques in 9-year minimum follow-up. Spine 2014;39:1093–8.

[55] Hiller A, Helenius I, Nurmi E, Neuvonen PJ, Kaukonen M, Hartikainen T, et al. Acetaminophen improves analgesia but does not reduce opioid requirement after major spine surgery in children and adolescents. Spine 2012;37:E1225–31.

[56] Hurlbert RJ, Alexander D, Bailey S, Mahood J, Abraham E, McBroom R, et al. rhBMP-2 for posterolateral instrumented lumbar fusion: a multicenter prospective randomized controlled trial. Spine 2013;38:2139–48.

[57] Jiya T, Smit T, Deddens J, Mullender M. Posterior lumbar interbody fusion using nonresorbable poly-ether-ether-ketone versus resorbable poly-L-lactide-co-D,L-lactide fusion devices: a prospective, randomized study to assess fusion and clinical outcome. Spine 2009;34:233–7.

[58] Kallmes DF, Comstock BA, Heagerty PJ, Turner JA, Wilson DJ, Diamond TH, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. N Engl J Med 2009;361:569–79.

[59] Klazen CA, Venmans A, de Vries J, van Rooij WJ, Jansen FH, Blonk MC, et al. Percutaneous vertebroplasty is not a risk factor for new osteoporotic compression fractures: results from VERTOS II. AJNR Am J Neuroradiol 2010;31:1447–50.

[60] Korovessis P, Vardakastanis K, Repantis T, Vitsas V. Balloon kyphoplasty versus KIVA vertebral augmentation–comparison of 2 techniques for osteoporotic vertebral body fractures: a prospective randomized study. Spine 2013;38:292–9.

[61] Lofgren H, Engquist M, Hoffmann P, Sigstedt B, Vavruch L. Clinical and radiological evaluation of trabecular metal and the Smith-Robinson technique in anterior cervical fusion for degenerative disease: a prospective, randomized, controlled study with 2-year follow-up. Eur Spine J 2010;19:464–73.

[62] Murrey D, Janssen M, Delamarter R, Goldstein J, Zigler J, Tay B, et al. Results of the prospective, randomized, controlled multicenter food and drug administration investigational device exemption study of the ProDisc-C total disc replacement versus anterior discectomy and fusion for the treatment of 1-level symptomatic cervical disc disease. Spine J 2009;9:275–86.

[63] Nagahama K, Kanayama M, Togawa D, Hashimoto T, Minami A. Does alendronate disturb the healing process of posterior lumbar interbody fusion? A prospective randomized trial. J Neurosurg Spine 2011;14:500–7.

[64] Nandyala SV, Marquez-Lara A, Fineberg SJ, Pelton M, Singh K. Prospective, randomized, controlled trial of silicate-substituted calcium phosphate versus rhBMP-2 in a minimally invasive transforaminal lumbar interbody fusion. Spine 2014;39:185–91.

[65] O'Neill KR, Lockney DT, Bible JE, Crosby CG, Devin CJ. Bupivacaine for pain reduction after iliac crest bone graft harvest. Orthopedics 2014;37:e428–34.

[66] Ohtori S, Suzuki M, Koshi T, Takaso M, Yamashita M, Yamauchi K, et al. Single-level instrumented posterolateral fusion of the lumbar spine with a local bone graft versus an iliac crest bone graft: a prospective, randomized study with a 2-year follow-up. Eur Spine J 2011;20:635–9.

[67] Phillips FM, Lee JY, Geisler FH, Cappuccino A, Chaput CD, DeVine JG, et al. A prospective, randomized, controlled clinical investigation comparing PCM cervical disc arthroplasty with anterior cervical discectomy and fusion. 2-year results from the US FDA IDE clinical trial. Spine 2013;38:E907–18.

[68] Pitzen TR, Chrobok J, Stulik J, Ruffing S, Drumm J, Sova L, et al. Implant complications, fusion, loss of lordosis, and outcome after anterior cervical plating with dynamic or rigid plates: two-year results of a multi-centric, randomized, controlled study. Spine 2009;34:641–6.

[69] Putzier M, Strube P, Funk JF, Gross C, Monig HJ, Perka C, et al. Allogeneic versus autologous cancellous bone in lumbar segmental spondylodesis: a randomized prospective study. Eur Spine J 2009;18:687–95.

[70] Ringel F, Stuer C, Reinke A, Preuss A, Behr M, Auer F, et al. Accuracy of robot-assisted placement of lumbar and sacral pedicle

Appendix 1. Database search strategy for MEDLINE (Jan 1, 2009 to Sept 11, 2014).

1. **Algorithm for relevant specialty and subspecialty journals:**

   1980:2012 [dp] AND (random* OR meta-analys* OR metaanalys*) AND ("Arthroscopy" [ta] OR "BMC Musculoskelet Disord"[ta] OR "Eur Spine J" [ta] OR "Foot Ankle Int" [ta] OR "Hip Int" [ta] OR "J Am Acad Orthop Surg" [ta] OR "J Arthroplasty" [ta] OR "J Bone Joint Surg Am" [ta] OR "J Bone Joint Surg br" [ta] OR "J Hand Surg Am" [ta] OR "J Orthop Res" [ta] OR "J Orthop Sports Phys Ther" [ta] OR "J Pediatr Orthop B" [ta] OR "J Pediatr Orthop" [ta] OR "J Shoulder Elbow Surg" [ta] OR "J Trauma" [ta] OR "Knee" [ta] OR "Osteoarthritis Cartilage" [ta] OR "Phys Ther" [ta] OR "Spine (Phila Pa 1976)" [ta] OR "Spine J" [ta] OR "Acta Orthop" [ta] OR "Acta Orthop Scand" [ta] OR "Gait Posture" [ta] OR "injury" [ta] OR "J Orthop Trauma" [ta] OR "Clin Orthop Relat Res" [ta] OR "Int Orthop" [ta] OR "J Neurosurg Spine" [ta] OR "Am J Sports Med" [ta] OR "Knee Surg Sports Traumatol Arthrosc" OR"Am J Knee Surg" [ta] OR "J Knee Surg" [ta] OR "J Hand Surg Br" [ta])

2. **Algorithm for relevant general medical journals:**

   1980:2012 [dp] AND (random* OR meta-analys* OR metaanalys*) AND ("Ann Intern Med" [ta] OR " BMJ" [ta] OR "JAMA" [ta] OR "Lancet" [ta] OR "N Engl J Med" [ta] OR "Arch Intern Med" [ta] OR "CMAJ" [ta] OR "Cochrane Database Syst Rev"[ta])

3. **Algorithm for relevant supplementary keywords:**

   (1980:2012 [dp] AND (randomi* OR meta-analys* OR metaanalys*) AND english [la] AND ((fracture* OR arthroplast* OR arthriti* OR joint OR orthopaed* OR orthoped*) NOT rheumat*)) NOT ("Arthroscopy" [ta] OR "BMC Musculoskelet Disord"[ta] OR "Eur Spine J" [ta] OR "Foot Ankle Int" [ta] OR "Hip Int" [ta] OR "J Am Acad Orthop Surg" [ta] OR "J Arthroplasty" [ta] OR "J Bone Joint Surg Am" [ta] OR "J Bone Joint Surg br" [ta] OR "J Hand Surg Am" [ta] OR "J Orthop Res" [ta] OR "J Orthop Sports Phys Ther" [ta] OR "J Pediatr Orthop B" [ta] OR "J Pediatr Orthop" [ta] OR "J Shoulder Elbow Surg" [ta] OR "J Trauma" [ta] OR "Knee" [ta] OR "Osteoarthritis Cartilage" [ta] OR "Phys Ther" [ta] OR "Spine (Phila Pa 1976)" [ta] OR "Spine J" [ta] OR "Acta Orthop" [ta] OR "Acta Orthop Scand" [ta] OR "Gait Posture" [ta] OR "injury" [ta] OR "J Orthop Trauma" [ta] OR "Clin Orthop Relat Res" [ta] OR "Int Orthop" [ta] OR "J Neurosurg Spine" [ta] OR "Am J Sports Med" [ta] OR "Knee Surg Sports Traumatol Arthrosc" OR "Am J Knee Surg" [ta] OR"J Knee Surg" [ta] OR "J Hand Surg Br" [ta] OR"Ann Intern Med" [ta] OR " BMJ" [ta] OR "JAMA" [ta] OR

"Lancet" [ta] OR "N Engl J Med" [ta] OR "Arch Intern Med" [ta] OR "CMAJ" [ta] OR "Cochrane Database Syst Rev"[ta])

4. Algorithm for relevant non-surgical specialty and subspecialty journals

((("J Physiother"[Journal] OR "Ann Phys Rehabil Med"[Journal] OR "PM R"[Journal] OR "Eur J Phys Rehabil Med"[Journal] OR "J Pediatr Rehabil Med"[Journal] OR "Dev Neurorehabil"[Journal] OR "Disabil Rehabil Assist Technol"[Journal] OR "J Neurol Phys Ther"[Journal] OR "J Rehabil Med Suppl"[Journal] OR "J Geriatr Phys Ther"[Journal] OR "J Rehabil Med"[Journal] OR "Phys Ther Sport"[Journal] OR "Neurorehabil Neural Repair"[Journal] OR "Ortop Traumatol Rehabil"[Journal] OR "Pediatr Rehabil"[Journal] OR "J Bodyw Mov Ther"[Journal] OR "Physiother Res Int"[Journal] OR "Man Ther"[Journal] OR "Occup Ther Int"[Journal] OR "Scand J Occup Ther"[Journal] OR "Disabil Rehabil"[Journal] OR "J Sport Rehabil"[Journal] OR "J Back Musculoskelet Rehabil"[Journal] OR "J Occup Rehabil"[Journal] OR "Phys Med Rehabil Clin N Am"[Journal] OR "Physiother Theory Pract"[Journal] OR "Pediatr Phys Ther"[Journal] OR "Am J Phys Med Rehabil"[Journal] OR "Clin Rehabil"[Journal] OR "Adv Clin Rehabil"[Journal] OR "Rehab Manag"[Journal] OR "J Hand Ther"[Journal] OR "J Rehabil Res Dev Clin Suppl"[Journal] OR "J Rehabil Res Dev"[Journal] OR "Phys Occup Ther Pediatr"[Journal] OR "J Orthop Sports Phys Ther"[Journal] OR "Int J Rehabil Res"[Journal] OR "Int Rehabil Med"[Journal] OR "Rheumatol Rehabil"[Journal] OR "Prog Phys Ther"[Journal] OR "Rheumatol Phys Med"[Journal] OR "Scand J Rehabil Med"[Journal] OR "Am Correct Ther J"[Journal] OR "Phys Ther"[Journal] OR "Aust Occup Ther J"[Journal] OR "Aust J Physiother"[Journal] OR "Arch Phys Med Rehabil"[Journal] OR "Am J Phys Med"[Journal] OR "Ann Phys Med"[Journal] OR "Physiotherapy"[Journal] OR "Am J Occup Ther"[Journal] OR "J Assoc Phys Ment Rehabil"[Journal] OR "J Rehabil"[Journal] OR "Can J Occup Ther"[Journal])) AND ((random* OR meta-analys* OR metaanalys*)))

Chapter 2

How to use a randomized controlled trial addressing a surgical or other non-pharmacological therapy: Users' Guides to the medical literature

Submitted to *JAMA* June 21, 2015;

transferred to *JAMA Surgery* Oct 29, 2015.

How to use a randomized controlled trial addressing a surgical or other non-pharmacological therapy: Users' Guides to the medical literature

Nathan Evaniew MD[1,2]
Alonso Carrasco-Labra DDs MSc[1,3]
PJ Devereaux MD PhD FRCPC[1,4]
Kari A O Tikkinen MD PhD[5]
Yutong Fei MD PhD[1,6]
Mohit Bhandari MD PhD FRCSC[1,2]
Gordon Guyatt MD MSc FRCPC[1,8]

[1]Department of Clinical Epidemiology and Biostatistics
McMaster University, Hamilton, ON, Canada

[2] Division of Orthopaedics, Department of Surgery
McMaster University, Hamilton, ON, Canada

[3]Evidence-Based Dentistry Unit, Faculty of Dentistry,
Universidad de Chile, Santiago, Chile

[4]Population Health Research Institute,
Hamilton, ON, Canada

[5]Departments of Urology and Public Health,
University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[6]Centre for Evidence-Based Chinese Medicine,
Beijing University of Chinese Medicine, Beijing, China

[8]Department of Medicine
McMaster University, Hamilton, Ontario, Canada

ABSTRACT

Background:
Because surgical and other non-pharmacological therapies require clinicians to develop and maintain procedural expertise, and because blinding in randomized controlled trials of such therapies is often challenging, their critical appraisal raises unique issues.

Findings:
Risk of bias of trials of non-pharmacologic therapies increases if investigators fail to rigorously conceal allocation and, where possible, to ensure blinding of those involved in the trial. Variability in clinicians' expertise can also increase bias and lead to important limitations in applicability. This Users' Guide, in addressing these issues, reviews the use of remote randomization systems, blinding, sham-controlled trials, split-body trials, expertise-based trials, and mechanistic versus practical trials.

Conclusion and relevance:
Consideration of risk of bias and applicability issues will allow clinicians to make optimal use of trials addressing surgical and other non-pharmacological therapies.

CLINICAL SCENARIO

You are an orthopaedic surgeon seeing a 65-year old woman with a displaced fracture of her right proximal humerus. Her injury occurred earlier in the day, when she tripped on a staircase and fell onto an outstretched hand. She did not suffer any other associated injuries, does not smoke, and has no serious medical comorbidities that would preclude surgery. Her fracture involves the surgical neck and greater tuberosity of her dominant proximal humerus, and each fragment is displaced approximately one centimeter.

You have treated many patients like this non-surgically, but the patient tells you she has a friend who suffered a similar fracture and did well after an operation. The patient's question brings to mind a notification you recently received on your mobile device for the Proximal Fracture of the Humerus Evaluation by Randomization (PROFHER) trial; you let the patient know you will get back to her very shortly after reviewing the latest evidence on the matter.[1] As you pull up the report, you wonder if there are any special issues to which you should attend when reviewing a randomized controlled trial (RCT) addressing a surgical or other non-pharmacological therapy.

INTRODUCTION

Although research to evaluate surgical and other non-pharmacological therapies has historically been dominated by non-randomized observational studies, increasing awareness that randomization reduces bias by ensuring similar distributions of prognostic factors in the intervention and control groups has led to a marked increase in the conduct of RCTs of such therapies.[2–4] This positive development raises new issues: trials of surgical and other non-pharmacological therapies present special challenges in understanding and applying their results to patient management.[5–7]

Many of these challenges arise because - in contrast to pharmacological treatments – interventions such as surgical procedures, interventional cardiology or radiology procedures, cognitive behavioral therapy, rehabilitation therapy, occupational therapy, chiropractic, and acupuncture rely on procedural expertise. In comparing alternative non-pharmacologic interventions, variability in procedural expertise between the intervention and control groups can therefore influence outcomes and result in spurious inferences.[8–10] Critical appraisal of RCTs of non-pharmacological interventions also requires addressing the frequent lack of blinding, which may be unavoidable, and even if possible is typically much more challenging than in RCTs of drug therapies.[11,12]

This User's Guide to the Medical Literature presents a practical approach to assessing RCTs addressing surgical or other non-pharmacological therapies using the three-step approach of our Users' Guides: assessment of risk of bias, results, and application to patient care (**Box 1**). Most of these issues apply to all RCTs; this article highlights those that are specific to surgical or other non-pharmacological interventions (**Table 1**).

HOW SERIOUS IS THE RISK OF BIAS?

Previous Users' Guides have considered whether allocation was concealed, patients were similar with respect to known prognostic factors, blinding was implemented, follow-up was complete, patients were analyzed in the groups to which they were randomized, and whether investigators avoided early stopping for benefit.[4,13–16] For RCTs of surgical or other non-pharmacological therapies, allocation concealment and blinding each warrant further discussion, and the expertise with which the study interventions were administered presents a unique issue requiring consideration.

Did the intervention and control groups start with the same prognosis?

Randomization, if successful, creates groups with a similar likelihood of experiencing the outcomes of interest (i.e. a similar prognosis). *Allocation*

concealment describes the extent to which individuals responsible for enrolling patients were unaware of and could not influence the study arm to which the randomization schedule assigned patients.[11] Concealment refers not to the process of creating the randomization schedule, nor to the methods of blinding used to maintain prognostic balance as a study progresses, but rather to safeguarding the implementation of the randomization process. Trials with inadequate methods of allocation concealment may overestimate treatment effects, and trials of surgical procedures frequently implement methods that are less secure.[6, 20]

To understand concealment, consider an RCT in which patients with appendicitis were randomly allocated to receive open or laparoscopic appendectomies, and the laparoscopic procedure but not the open procedure required the attending surgeons' presence in the operating room.[17] Resident physicians, responsible for recruiting and enrolling patients, obtained the treatment assignments from sealed envelopes. The residents were typically eager to perform procedures independently and more often familiar and confident with open rather than laparoscopic appendectomy. In this trial, experienced residents did not require supervision for open procedures, but all residents required supervision for laparoscopic procedures.

When patients required surgery during the night, residents who were reluctant to call in their attending staff held up the envelopes to the light until they found one that contained an open procedure. In the morning, the next eligible patients were allocated according to the envelopes that were passed over (D. Wall, written communication, June 9, 2000). If the patients who presented overnight were sicker or the care they received without the attending surgeons' presence was inferior, the lack of concealment would have biased the results in favour of the laparoscopic procedure.[4,18]

This example illustrates the vulnerability of RCTs of surgical or other non-pharmacological therapies to lack of concealment.[3,4,6,20,21] In a typical blinded pharmacological trial, study drugs are packaged and labelled before they are sent to each participating center, independent of the research coordinators who enrol patients. This arrangement prevents the coordinators from knowing which medication the next patient will receive, and thus secures the randomization sequence. In order to corrupt the sequence in a pharmacological trial, the coordinators would have to obtain the central randomization sequence and unblind the packaged medications. As the appendectomy example illustrates, circumventing randomization is far easier in non-pharmacological trials that implement envelopes (or other even less secure methods) for concealment.

Investigators of non-pharmacologic RCTs can ensure allocation concealment by using remote web-based and 24-hour central telephone randomization services that require individuals who are enrolling patients to contact an independent source.[11] Each contact is logged, and treatment assignments for each patient are provided only after eligibility has been confirmed. Although sealed, opaque, and sequentially numbered envelopes are preferable to envelopes that are unsealed, translucent, and not numbered, they remain vulnerable to tampering and are therefore less secure than remote randomization systems.

Randomization may fail to do its job of creating prognostic balance through bad luck (chance), or through failure to conceal allocation. Either way, one can appraise the success of randomization by examining the distributions of the baseline characteristics in the intervention and control groups, usually presented in the first table of results. Clinicians can be reassured when the known prognostic factors are similar – and be legitimately concerned if they are not.

Was prognostic balance maintained as the study progressed?

Blinding is the process of withholding information about treatment assignments from groups of individuals who could introduce bias if they gained this information after patients were randomized.[22] Investigators can

blind pharmacological therapies easily using placebo medications; placebos are, however, often not possible for surgical or other non-pharmacological therapies, leaving blinding challenging to implement. Lack of blinding is a particularly serious concern when, as is often the case in non-pharmacological RCTs, investigators focus on outcomes that are subjective (such as pain, function, quality of life, and satisfaction).[3,12]

There are six groups of individuals who should ideally be blinded in RCTs: participants, healthcare providers, data collectors, outcomes assessors, data analysts, and the investigators responsible for interpreting the results (**Table 2**).[13,23] In trials of surgical or other non-pharmacological therapies, it is sometimes possible to blind the participants and some health care personnel, but almost never possible to blind the surgeons or other clinicians who administer the treatments. With careful planning it may be possible to blind the data collectors, is usually possible to blind outcome assessors, and always possible to blind data analysts and those interpreting the results.[12,24]

For example, consider the ReCharge trial, in which 239 participants were randomly allocated to undergo surgical implantation of an active vagal nerve block device or a sham device to determine the effect of reversible intermittent intra-abdominal vagal nerve blockade on morbid obesity.[25] The

surgeons could not be blinded to the procedures that they were performing and the patients might have become unblinded by asking their surgeons which operation they had, so interaction between the surgeons and participants was limited post-operatively and blinded staff conducted patient follow-up.

Trials of non-pharmacological procedures can also implement independent blinded Central Adjudication Committees (CACs) to evaluate outcomes.[26] CACs typically consist of three or more experts who, blinded to allocation, assess anonymized outcome data according to pre-defined criteria.[27]

Participants in RCTs of surgical or other non-pharmacological therapies can sometimes be blinded using standardized wound coverings, digitally altered radiographs, or split-body designs (**Supplement 1**). This is much easier, however, when the comparisons are between alternative but similar interventions (such as one surgical approach versus another, or one physiotherapy approach versus another) than when the control is a non-surgical intervention or standard of care.[28–31]

For outcomes that are subjective, placebo effects from surgical interventions or other non-pharmacological interventions (such as, for

instance, acupuncture) are often substantial.[32–35] Moreover, rituals, stressors, and environmental cues associated with admission, preparation, anesthesia, and recovery can heighten placebo effects associated with surgery because they may lead patients to have greater expectations for benefit.[36]

Sham-controlled RCTs can control for the placebo effect of surgery by ensuring that neither intervention or control patients know whether they have undergone the surgical procedure.[36] Sham surgical procedures might seem to expose participants to unreasonable harms without promise of a direct benefit, but participants who receive sham treatments may not only experience substantial placebo effects but also receive additional monitoring, imaging, or clinic visits beyond standard practice.[31] Sham-controlled RCTs are most effectively implemented by ensuring the relevant clinical community feels unsure about the relative effects of intervention and control, minimizing risk, obtaining informed consent, and avoiding ongoing active deception.[35–38]

In the ReCharge trial, participants assigned to the sham vagal nerve block device experienced similar operations as those in the active device treatment group because both procedures required similar standard laparoscopic techniques and general anaesthesia, both involved

identification and dissection of the anterior and posterior vagus nerves, and both involved implantation of a similarly-sized subcutaneous device on the thoracic side wall.[25] Participants in the active device group experienced significantly greater weight loss than participants in the sham device group, but the sham device group nonetheless experienced three-fold greater weight loss than was predicted, which suggests the influence of substantial placebo effects.

In summary, in non-pharmacological RCTs, data analysts and those interpreting results can always be blinded, outcome assessors can usually be blinded, and patients and data collectors can sometimes be blinded, particularly if investigators have used sham procedures. When blinding is not possible or not undertaken, excluding placebo effects as an explanation of apparently positive results may not be possible.

The extent to which concern about placebo effects undermines trial credibility is a matter of judgment and will differ according to circumstances. For instance, if the outcome is mortality, placebo effects are much less of a concern than if the outcome is subjective symptoms.

Were the groups prognostically balanced at the study's completion?

As is the case for RCTs of pharmacological interventions, one can be increasingly confident that the treatment groups were prognostically balanced at the completion of a trial when follow-up was complete, patients were analyzed in the groups to which they were randomized, and the trial was not stopped early because of apparently large treatment effects.[15,16,39]

Were the interventions administered with similar expertise?

RCTs of surgical or other non-pharmacological therapies are at unique risk for additional bias due to potential differences in the expertise with which the interventions were administered. *Differential expertise bias* occurs when expertise is systematically greater in one treatment over the other.[8,40–42] Not only may clinicians be more experienced and skilled in performing an experimental or control intervention, but their conscious or subconscious investment in the superiority of one procedure over the other may lead to differential administration of effective co-interventions (i.e. administration of antibiotics, wound care, or early mobilization).

For example, consider an RCT in which 206 participants with ventral incisional hernias were randomly allocated to undergo laparoscopic or open hernia repair.[43] This trial involved multiple surgeons from 10 participating hospitals, each of whom was considered “experienced and

dedicated" and performed both procedures. If the surgeons in this trial had greater skill or investment in performing open rather than laparoscopic incisional hernia repair (perhaps as a result of their training and experience), we would expect differential expertise bias to favour the open procedure. The results showed longer operative times, higher rates of perioperative complications, and higher rates of recurrence in the laparoscopic group. Whether the results would have been similar if potential bias due to differential expertise had been addressed (for instance, by randomizing patients to surgeons who did only open hernia repairs or those who did only laparoscopic hernia repairs, what we call an *expertise-based* RCT design[8]) remains uncertain.

Indeed, use of such an expertise-based design in which patients are randomized to surgeons experienced and invested in the experimental treatment or to surgeons experienced and invested in the control treatment is the best way to guard against differential expertise bias. For example, consider The Coronary Artery Bypass Grafting (CABG) Off or On Pump Revascularization Study (CORONARY) in which 4752 participants were randomly allocated according to an expertise-based design between a group of surgeons who preferred a novel beating-heart technique (off-pump CABG) or a group of surgeons who preferred a conventional cardiopulmonary bypass technique (on-pump CABG).[44,45] To minimize

possible confounding due to a learning curve, the trial included only surgeons who had completed at least 100 cases with their preferred procedure. To minimize the potential for differential local co-interventions, such as post-operative wound care or rehabilitation protocols, each participating centre had both off-pump and on-pump surgeons available.

RCTs that compare surgical interventions to non-surgical procedural interventions, such as trials of surgery versus physiotherapy, are expertise-based by definition because the surgeons do not administer the physiotherapy and the physiotherapists do not administer the surgery. Trials that compare surgical procedures against pharmacological interventions are also expertise-based by definition because the administration of pharmacological interventions does not rely on procedural expertise.

Box 2 presents the conclusions of the risk of bias assessment for the PROFHER trial.

WHAT ARE THE RESULTS?

In discussing interpretation of results, previous Users' Guides have discussed composite endpoints, and non-inferiority trials.[13,46–49] These issues are equally relevant to RCTs of surgical or other non-

pharmacological therapies, and we will not address them in this article. We will, however, review the most basic considerations: measures of effect, and measures of precision.

When considering the magnitude of a dichotomous (yes or no, e.g. dead or alive) treatment effect, relative measures of association can be misleading.[50] For instance, a relative risk reduction of 50% sounds impressive. Indeed, it would be impressive if it reflected a reduction in death or surgical complications from 40% to 20% (a risk difference or absolute risk reduction of 20%). It might also, however, represent a reduction from 2% to 1% (a 1% risk difference) that may, in context, be trivial. Therefore, clinicians should also look for risk differences or its inverse, the number needed to treat (100/20 or 5 when the risk difference is 20%; 100/1 or 100 when the risk difference is 1%) to understand the magnitude of effect.[13,51]

For continuous outcomes, clinicians must decide whether observed treatment effects are likely to be sufficiently large to justify changes in management given any foreseeable harms or costs.[49,52] When outcomes are measured using instruments unfamiliar to clinicians (e.g. score on a quality of life instrument) such decisions may be extremely challenging. Knowing the minimally important difference (the smallest difference that

patients would consider important) is likely to be very helpful in interpreting the results.[13]

When considering precision, clinicians should look for *confidence intervals* – 95% confidence intervals are typically reported. Clinicians should look to the upper and lower boundaries of confidence intervals to discern the largest and smallest possible treatment effects that, given the results, remain plausible.[13,51]

Box 3 presents considerations in understanding results from the PROFHER trial.

HOW CAN I APPLY THE RESULTS TO PATIENT CARE?

In order to apply the results from an RCT to patient care, one must consider the extent to which a trial is *applicable*. The setting of the trial, the methods of selecting the trial participants, the characteristics of the participants, differences between the trial protocol and routine practice, the chosen outcome measures, the duration of follow-up, and the observed adverse effects can all influence applicability.[53,54] A given trial could be perfectly applicable for some clinicians and their particular patient encounter, yet of limited applicability for others.

Clinicians can be confident applying trial results when their patients are similar to study patients in factors such as age or severity of illness.[13,14] Because true subgroup effects (that is, modification of effect by factors such as age, sex, comorbidity, or illness severity) are rare, the best approach to interpretability is to consider whether there are any compelling reasons not to apply the results to one's own patients: most often, the answer will be no.[55,56] In addressing applicability, clinicians should also address the extent to which all patient-important outcomes were considered and whether the observed treatment benefits were likely to be worth any potential harms or costs (**Box 1**).[57]

Were the study interventions similar to interventions in my setting?

For RCTs of surgical or other non-pharmacological therapies, applicability also depends on the extent to which the study interventions are similar to interventions in one's own setting. Variability in individual clinicians' expertise can affect their ability to achieve results from trials in their own practice. For instance, consider a trial demonstrating that a novel surgical procedure conducted by surgeons already experienced in the new procedure appears superior to the standard. For a surgeon who has never used anything but the standard, training (potentially extensive) would be required before we could be confident that surgeon could achieve results demonstrated in the trial.

Consider the Asymptomatic Carotid Artery Surgery (ACAS) trial, in which 1662 participants with asymptomatic carotid artery stenosis were randomly allocated to carotid endarterectomy surgery or optimal medical risk factor management. This trial found that carotid endarterectomy surgery significantly reduced the overall risk of stroke or death from 11.0% to 5.1%.[58] However, this trial selected a group of surgeons who had established very low peri-operative complication rates, lower than that found elsewhere in the literature (and likely lower than what one might anticipate in many routine clinical practices).[53,59,60] This difference in peri-operative complications is great enough that it would be likely to reduce, or even reverse, any possible benefit.[53]

This example highlights how a trial can be useful for addressing one question, but not at all useful for addressing an apparently similar, but actually fundamentally different question. One may ask: what is the effect of a procedure, delivered by those most expert, in comparison to the competing management strategy? Alternatively, one may ask: what is the effect of the procedure when undertaken by the level of expertise one might expect in most communities?

An RCT addressing the first question might be considered a *mechanistic* or *explanatory* trial that addresses the impact of an intervention administered under ideal testing circumstances.[10] An RCT addressing the second question might be considered a *practical* or *pragmatic* trial that bears directly on health care decisions in practice. Whether a trial is mechanistic or practical may also depend on your perspective: if you are in a community of surgeons with exceptional expertise, the first and not the second trial may be practical from your point of view.

The Appendicitis Acuta (APPAC) trial illustrates the mechanistic versus practical perspective.[61,62] In this trial, 530 participants with uncomplicated acute appendicitis were randomly allocated to receive either early appendectomy or a standardized regimen of antibiotic treatment and followed for one year. Of 257 participants in the antibiotic treatment group, 15 underwent appendectomy during their initial hospitalization because the surgeons suspected progressive infection, perforation, or peritonitis.

From a mechanistic perspective that seeks the impact of surgery versus medical management, this trial might be considered problematic because some patients allocated to antibiotics underwent appendectomy. If, however, the real-world options are immediate surgery versus a wait-and-see approach in which one holds off surgery indefinitely if patients

improve, but only temporarily if they do not, this trial is eminently practical (and thus extremely helpful).

Clinicians considering the applicability of the APPAC trial must also consider whether the non-operative antibiotic regimen (three days of intravenous ertapenem followed by seven days of oral levofloxacin and metronidazole) was similar to the regimen that they would use in their own practice and whether all patient-important outcomes were considered, including long-term complications of surgery.

The mechanistic versus practical perspective is particularly important when interpreting trials of *behavioral interventions*, which include activities, techniques, or strategies that target biological, cognitive, emotional, interpersonal, social, or environmental mediators.[63] Trials of behavioral interventions frequently implement complex strategies to ensure optimal administration of the interventions (*treatment fidelity*), such as video or audio-recording treatment sessions, verbatim treatment protocols, regular conferences to review study cases, and continued re-training.[64]

For example, consider an RCT in which 135 participants with chronic migraine headaches were randomly allocated to receive cognitive behavioral therapy or headache education in addition to standardized

medications.[65] Before starting the trial, an expert clinical psychologist with expertise in pain management trained the clinicians who administered the interventions and directed them to use structured treatment manuals. During the trial, clinicians' treatment sessions were recorded using audiotapes and reviewed quarterly by an independent evaluator with a checklist of required topics. If the clinicians' treatment accuracy fell below 80%, study personnel provided formal re-training.

At the end of this trial, both cognitive behavioral therapy and headache education had been delivered with high treatment fidelity (mean accuracy scores of 94.5% and 96.9%, respectively). One may see this as impressive and admirable, but without training by an expert psychologist, structured treatment manuals, and recording of audiotapes with subsequent feedback, will it reflect administration of the intervention in clinical practice? If the answer – as we suspect is the case – is very likely to be that it will not, the applicability of the trial is seriously jeopardized.

Clinicians must therefore consider whether the interventions delivered in the trial were similar to those of their own setting. If resource-intensive strategies do indeed substantially increase treatment fidelity, and if those strategies will not be in place in the community, then the intervention tested will certainly differ from that implemented in regular clinical practice.

Clinicians must then wait for a practical rather than a mechanistic trial for a definitive demonstration of the impact of the intervention in the real world.

Box 4 presents issues of applicability and the resolution of our clinical scenario.

CONCLUSIONS

Surgical and other non-pharmacological therapies require clinicians to develop and maintain procedural expertise. RCTs of surgical and other non-pharmacological therapies present unique methodological concerns in part related to the possibility of differential expertise, and the relative expertise of trial practitioners versus those in your community. Further, failure to apply rigorous methods to achieve allocation concealment and innovative approaches to implement blinding may introduce serious risk of bias. Consideration of these issues will allow clinicians to make optimal use of trials addressing surgical and other non-pharmacological therapies. Clinicians should proceed with increasing confidence when interpreting trials with less risk of bias and greater applicability, and increasing caution for trials with greater risk of bias and questionable applicability.

ACKNOWLEDGEMENTS

Author Contributions:

Dr. Evaniew and Dr. Guyatt had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Financial Disclosures:

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported related to this work. This study was not funded. Dr. Bhandari received funding in part from a Canadian Institutes of Health Research Canada Research Chair. Dr. Evaniew received funding in part from a Canadian Institutes of Health Research Doctoral Research Award. Dr. Tikkinen received funding from the Academy of Finland (#276046), Hospital District of Helsinki and Uusimaa, Jane and Aatos Erkko Foundation, and Sigrid Jusélius Foundation.

REFERENCES

1. Rangan A, Handoll H, Brealey S, et al. Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA*. 2015;313(10):1037-1047.
2. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *JAMA*. 2003;290(8):1062-1070.
3. Farrokhyar F, Karanicolas PJ, Thoma A, et al. Randomized controlled trials of surgical interventions. *Ann Surg*. 2010;251(3):409-416.
4. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am*. 2001;83-A(6):916-926.
5. Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. *J Bone Joint Surg Am*. 2007;89(9):1970-1978.
6. Nagendran M, Harding D, Teo W, et al. Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study. *BMJ Open*. 2013;3(12):e003898.
7. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, CONSORT Group. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008;148(4):295-309.
8. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. *BMJ*. 2005;330(7482):88.
9. Walter SD, Ismaila AS, Devereaux PJ, SPRINT Study Investigators. Statistical issues in the design and analysis of expertise-based randomized clinical trials. *Stat Med*. 2008;27(30):6583-6596.
10. Karanicolas PJ, Montori VM, Devereaux PJ, Schünemann H, Guyatt GH. A new 'mechanistic-practical" framework for designing and interpreting randomized trials. *J Clin Epidemiol*. 2009;62(5):479-484.
11. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614-618.
12. Karanicolas PJ, Bhandari M, Taromi B, et al. Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. *J Bone Joint Surg Am*. 2008;90(5):1026-1033.
13. Guyatt G, Rennie D, Meade MO, Cook DJ. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3E*. Vol 3 edition. New York: McGraw-Hill Professional; 2014.
14. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are

the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 1993;270(21):2598-2601.
15. Montori VM, Devereaux PJ, Adhikari NKJ, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203-2209.
16. Bassler D, Briel M, Montori VM, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303(12):1180-1187.
17. Hansen JB, Smithers BM, Schache D, Wall DR, Miller BJ, Menzies BL. Laparoscopic versus open appendectomy: prospective randomized trial. *World J Surg*. 1996;20(1):17-20; discussion 21.
18. Phatak UR, Chan WM, Lew DF, et al. Is nighttime the right time? Risk of complications after laparoscopic cholecystectomy at night. *J Am Coll Surg*. 2014;219(4):718-724.
19. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-412.
20. Adie S, Harris IA, Naylor JM, Mittal R. CONSORT compliance in surgical randomized trials: are we there yet? A systematic review. *Ann Surg*. 2013;258(6):872-878.
21. Kim KH, Kang JW, Lee MS, Lee J-D. Assessment of the quality of reporting in randomised controlled trials of acupuncture in the Korean literature using the CONSORT statement and STRICTA guidelines. *BMJ Open*. 2014;4(7):e005068.
22. Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285(15):2000-2003.
23. Devereaux PJ, Bhandari M, Montori VM, Manns BJ, Ghali WA, Guyatt GH. Double blind, you are the weakest link--good-bye! *ACP J Club*. 2002;136(1):A11.
24. Järvinen TLN, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *J Clin Epidemiol*. 2014;67(7):769-772.
25. Ikramuddin S, Blackstone RP, Brancatisano A, et al. Effect of reversible intermittent intra-abdominal vagal nerve blockade on morbid obesity: The recharge randomized clinical trial. *JAMA*. 2014;312(9):915-922.
26. Vannabouathong C, Saccone M, Sprague S, Schemitsch EH, Bhandari M. Adjudicating outcomes: fundamentals. *J Bone Joint Surg Am*. 2012;94 Suppl 1:70-74.
27. SPRINT Investigators, Simunovic N, Walter S, et al. Outcomes assessment in the SPRINT multicenter tibial fracture trial:

Adjudication committee size has trivial effect on trial results. *J Clin Epidemiol*. 2011;64(9):1023-1033.
28. Stensbirk F, Thorborg K, Konradsen L, Jørgensen U, Hölmich P. Iliotibial band autograft versus bone-patella-tendon-bone autograft, a possible alternative for ACL reconstruction: a 15-year prospective randomized controlled trial. *Knee Surg Sports Traumatol Arthrosc Off J ESSKA*. 2014;22(9):2094-2101.
29. Karanicolas PJ, Bhandari M, Walter SD, Heels-Ansdell D, Guyatt GH, Collaboration for Outcomes Assessment in Surgical Trials (COAST) Musculoskeletal Group. Radiographs of hip fractures were digitally altered to mask surgeons to the type of implant without compromising the reliability of quality ratings or making the rating process more difficult. *J Clin Epidemiol*. 2009;62(2):214-223.e1.
30. Mantovani E, Arduino PG, Schierano G, et al. A Split-Mouth Randomized Clinical Trial to Evaluate the Performance of Piezosurgery Compared With Traditional Technique in Lower Wisdom Tooth Removal. *J Oral Maxillofac Surg*. 2014;72(10):1890-1897.
31. Flum DR. Interpreting surgical trials with subjective outcomes: avoiding UnSPORTsmanlike conduct. *JAMA*. 2006;296(20):2483-2485.
32. Koog YH, Lee JS, Wi H. Clinically meaningful nocebo effect occurs in acupuncture treatment: a systematic review. *J Clin Epidemiol*. 2014;67(8):858-869.
33. Paterson C, Dieppe P. Characteristic and incidental (placebo) effects in complex interventions such as acupuncture. *BMJ*. 2005;330(7501):1202-1205.
34. Teixeira MZ, Guedes CHFF, Barreto PV, Martins MA. The placebo effect and homeopathy. *Homeopathy J Fac Homeopathy*. 2010;99(2):119-129.
35. Wartolowska K, Judge A, Hopewell S, et al. Use of placebo controls in the evaluation of surgery: systematic review. *BMJ*. 2014;348:g3253.
36. Dowrick AS, Bhandari M. Ethical issues in the design of randomized trials: to sham or not to sham. *J Bone Joint Surg Am*. 2012;94 Suppl 1:7-10.
37. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med*. 2009;361(6):569-579.
38. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med*. 2009;361(6):557-568.
39. Akl EA, Briel M, You JJ, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ*. 2012;344:e2809.

40. Scholtes VA, Nijman TH, van Beers L, Devereaux PJ, Poolman RW. Emerging designs in orthopaedics: expertise-based randomized controlled trials. *J Bone Joint Surg Am*. 2012;94 Suppl 1:24-28.
41. Bednarska E, Bryant D, Devereaux PJ, Expertise-Based Working Group. Orthopaedic surgeons prefer to participate in expertise-based randomized trials. *Clin Orthop*. 2008;466(7):1734-1744.
42. Maruthappu M, Gilbert BJ, El-Harasis MA, et al. The influence of volume and experience on individual surgical performance: a systematic review. *Ann Surg*. 2015;261(4):642-647.
43. Eker HH, Hansson BE, Buunen M, et al. Laparoscopic vs open incisional hernia repair: A randomized clinical trial. *JAMA Surg*. 2013;148(3):259-263. doi:10.1001/jamasurg.2013.1466.
44. Lamy A, Devereaux PJ, Prabhakaran D, et al. Effects of off-pump and on-pump coronary-artery bypass grafting at 1 year. *N Engl J Med*. 2013;368(13):1179-1188.
45. Garg AX, Devereaux PJ, Yusuf S, et al. Kidney function after off-pump or on-pump coronary artery bypass graft surgery: a randomized clinical trial. *JAMA*. 2014;311(21):2191-2198.
46. Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. How to use a noninferiority trial: users' guides to the medical literature. *JAMA*. 2012;308(24):2605-2611.
47. Montori VM, Permanyer-Miralda G, Ferreira-González I, et al. Validity of composite end points in clinical trials. *BMJ*. 2005;330(7491):594-596.
48. Busse JW, Guyatt GH. Copresentation of relative and absolute effects is essential to promote optimal interpretability of treatment effects. *J Clin Epidemiol*. 2015;68(3):355-356.
49. Schünemann HJ, Guyatt GH. Commentary--goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res*. 2005;40(2):593-597.
50. Murad MH, Montori VM, Ioannidis JP, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA*. 2014;312(2):171-179.
51. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA*. 1994;271(1):59-63.
52. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.
53. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005;365(9453):82-93.

54. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006;1(1):e9.
55. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the "number needed to treat"? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol*. 2002;31(1):72-76.
56. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 2002;21(11):1575-1600.
57. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311(4):405-411.
58. Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. *JAMA*. 1995;273(18):1421-1428.
59. Moore WS, Vescera CL, Robertson JT, Baker WH, Howard VJ, Toole JF. Selection process for surgeons in the Asymptomatic Carotid Atherosclerosis Study. *Stroke J Cereb Circ*. 1991;22(11):1353-1357.
60. Moore WS, Young B, Baker WH, et al. Surgical results: a justification of the surgeon selection process for the ACAS trial. The ACAS Investigators. *J Vasc Surg*. 1996;23(2):323-328.
61. Salminen P, Paajanen H, Rautio T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: The appac randomized clinical trial. *JAMA*. 2015;313(23):2340-2348.
62. Livingston E, Vons C. Treating appendicitis without surgery. *JAMA*. 2015;313(23):2327-2328.
63. Pincus H, England M. Improving the quality of psychosocial interventions for mental and substance use disorders: A report from the iom. *JAMA*. July 2015.
64. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol Off J Div Health Psychol Am Psychol Assoc*. 2004;23(5):443-451.
65. Powers SW, Kashikar-Zuck SM, Allen JR, et al. Cognitive behavioral therapy plus amitriptyline for chronic migraine in children and adolescents: a randomized clinical trial. *JAMA*. 2013;310(24):2622-2630.
66. Zhang G, Chen J, Chai W, Liu M, Wang Y. Comparison between computer-assisted-navigation and conventional total knee arthroplasties in patients undergoing simultaneous bilateral procedures: a randomized clinical trial. *J Bone Joint Surg Am*. 2011;93(13):1190-1196.
67. Aversano T, Aversano LT, Passamani E, et al. Thrombolytic therapy vs primary percutaneous coronary intervention for myocardial infarction in patients presenting to hospitals without on-site cardiac

surgery: A randomized controlled trial. *JAMA*. 2002;287(15):1943-1951.
68. Yamanaka T, Kawai Y, Miyoshi T, et al. Remote ischemic preconditioning reduces contrast-induced acute kidney injury in patients with ST-elevation myocardial infarction: a randomized controlled trial. *Int J Cardiol*. 2015;178:136-141.
69. Serruys PW, Morice M-C, Kappetein AP, et al. Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *N Engl J Med*. 2009;360(10):961-972.
70. Zaidat OO, Fitzsimmons B-F, Woodward BK, et al. Effect of a balloon-expandable intracranial stent vs medical therapy on risk of stroke in patients with symptomatic intracranial stenosis: the VISSIT randomized clinical trial. *JAMA*. 2015;313(12):1240-1248.
71. Siddiqui AH, Zivadinov R, Benedict RHB, et al. Prospective randomized trial of venous angioplasty in MS (PREMiSe). *Neurology*. 2014;83(5):441-449.
72. Fiatarone Singh MA, Gates N, Saigal N, et al. The Study of Mental and Resistance Training (SMART) study—resistance training and/or cognitive training in mild cognitive impairment: a randomized, double-blind, double-sham controlled trial. *J Am Med Dir Assoc*. 2014;15(12):873-880.
73. Simpson HB, Foa EB, Liebowitz MR, et al. Cognitive-behavioral therapy vs risperidone for augmenting serotonin reuptake inhibitors in obsessive-compulsive disorder: a randomized clinical trial. *JAMA Psychiatry*. 2013;70(11):1190-1199.
74. Walsh TS, Salisbury LG, Merriweather JL, et al. Increased hospital-based physical rehabilitation and information provision after intensive care unit discharge: The recover randomized clinical trial. *JAMA Intern Med*. 2015;175(6):901-910.
75. Latham NK, Harris B, Bean JF, et al. Effect of a home-based exercise program on functional recovery following rehabilitation after hip fracture: A randomized clinical trial. *JAMA*. 2014;311(7):700-708.
76. Shakeri H, Keshavarz R, Arab AM, Ebrahimi I. Clinical effectiveness of kinesiological taping on pain and pain-free shoulder range of motion in patients with shoulder impingement syndrome: a randomized, double blinded, placebo-controlled trial. *Int J Sports Phys Ther*. 2013;8(6):800-810.
77. Thomas S, Thomas PW, Kersten P, et al. A pragmatic parallel arm multi-centre randomised controlled trial to assess the effectiveness and cost-effectiveness of a group-based fatigue management programme (FACETS) for people with multiple sclerosis. *J Neurol Neurosurg Psychiatry*. 2013;84(10):1092-1099.

78. Sackley CM, Walker MF, Burton CR, et al. An occupational therapy intervention for residents with stroke related disabilities in UK care homes (OTCH): cluster randomised controlled trial. *The BMJ*. 2015;350:h468.
79. Page SJ, Levine P, Leonard A. Mental Practice in Chronic Stroke Results of a Randomized, Placebo-Controlled Trial. *Stroke*. 2007;38(4):1293-1297.
80. Linde K, Streng A, Jürgens S, et al. Acupuncture for patients with migraine: A randomized controlled trial. *JAMA*. 2005;293(17):2118-2125.
81. Johnston BC, da Costa BR, Devereaux PJ, Akl EA, Busse JW, Expertise-Based RCT Working Group. The use of expertise-based randomized controlled trials to assess spinal manipulation and acupuncture for low back pain: a systematic review. *Spine*. 2008;33(8):914-918.
82. Smaïl-Faugeron V, Fron-Chabouis H, Courson F, Durieux P. Comparison of intervention effects in split-mouth and parallel-arm randomized controlled trials: a meta-epidemiological study. *BMC Med Res Methodol*. 2014;14(1):64.
83. Hopewell S, Dutton S, Yu L-M, Chan A-W, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340.
84. Mills EJ, Chan A-W, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials*. 2009;10:27. doi:10.1186/1745-6215-10-27.
85. Kim YS, Lee HJ, Cho SH, Lee JD, Kim HS. Early postoperative treatment of thyroidectomy scars using botulinum toxin: a split-scar, double-blind randomized controlled trial. *Wound Repair Regen Off Publ Wound Heal Soc Eur Tissue Repair Soc*. 2014;22(5):605-612.
86. Gurung A, Karki DB, Shrestha S, Rijal AP. Visual outcome of conventional extracapsular cataract extraction with posterior chamber intraocular lens implantation versus manual small-incision cataract surgery. *Nepal J Ophthalmol Biannu Peer-Rev Acad J Nepal Ophthalmic Soc NEPJOPH*. 2009;1(1):13-19.

Box 1: Users' Guides approach to a randomized controlled trial addressing a surgical or other non-pharmacological therapy

HOW SERIOUS IS THE RISK OF BIAS?

Did the intervention and control groups start with the same prognosis?

- Was allocation concealed? *
- Were patients in the treatment and control groups similar with respect to known prognostic factors?

Was prognostic balance maintained as the study progressed?

- Were participants, healthcare providers, data collectors, outcomes assessors, data analysts, and/or those interpreting results blinded? *

Were the groups prognostically balanced at the study's completion?

- Was follow-up complete?
- Were patients analyzed in the groups to which they were randomized?
- Was the trial stopped early for benefit?

*Were the interventions administered with similar expertise? **

WHAT ARE THE RESULTS?

How large was the treatment effect?

How precise was the estimate of the treatment effect?

HOW CAN I APPLY THE RESULTS TO PATIENT CARE?

Were the study patients similar to my patient?

*Were the study interventions similar to interventions in my setting? **

Were all patient-important outcomes considered?

Are the likely treatment benefits worth the potential harms and costs?

**Includes issues specific to trials of surgical or other non-pharmacological therapies.*

Box 2: Using the Guide - How serious is the risk of bias?

In the PROPHER trial, 250 participants with displaced proximal humerus fractures were randomly allocated to either surgical or non-surgical treatment.[1] A computer program generated the randomization sequence and allocation concealment was achieved using independent remote randomization with randomly varied blocking. The participants in the intervention and control groups were similar with respect to their known prognostic factors except that there were more smokers in the non-surgical group (32% versus 19%).

The participants, healthcare providers, data collectors, outcome assessors, and data analysts were not blinded in this trial, but complete follow-up data were available for 92% of the participants, the participants were analyzed in the groups to which they were randomized, and the trial was not stopped early for benefit.

Those allocated to surgery received internal fixation or humeral head replacement according to the preferences and familiarity of the participating surgeons and then received supervised post-operative physiotherapy in inpatient, outpatient, or community settings. Those allocated to non-surgical care received a sling or hanging bandage for as long as necessary followed by supervised physiotherapy.

You conclude that that this trial is at low risk of bias with two exceptions: the results could be biased if either surgery or physiotherapy has a strong placebo effect on subjective outcomes, or if those who evaluated the outcomes were biased toward one treatment or the other. Still, the trial is sufficiently credible that you continue on to the results.

Box 3: Using the Guide - What are the results?

The primary analysis of PROPHER compared Oxford Shoulder Scores between the surgical and non-surgical groups over two years of follow-up. The Oxford Shoulder Score is a continuous outcome that measures shoulder-related function, and five points was considered to be a minimally important difference. Secondary outcomes included Short-Form 12 (SF-12) scores, complications, and late interventions. Investigators conducted sensitivity analyses to evaluate the impact of missing data, to control for potential confounding due to the differing proportion of smokers in each group, and to explore possible clustering across the participating centers.

There were no significant differences between the groups for Oxford Shoulder Scores and the 95% confidence intervals (CIs) excluded a 5-point difference over two years (mean difference of 0.75 points in favor of the surgical group; 95% CI −1.33 to 2.84; p=0.48). There were also no significant differences in in SF-12 scores (mean difference of 1.77 points in favor of the surgical group; 95% CI -0.84 to 4.30; p=0.18), or rates of shoulder-related complications (30 in the surgical group versus 23 in the non-surgical group; risk difference 6%, 95% CI -5% to 15%; p=0.28). The rates of secondary surgery (11 versus 11) and serious adverse events (28 versus 28) were identical. Unadjusted and adjusted analyses yielded similar results.

Box 4: Using the Guide - How can I apply the results to patient care?

You review the tables presenting baseline characteristics, and note that the patients in PROFHER were very similar to your patient with respect to age, sex, injury mechanism, and fracture pattern. You therefore find no compelling reason to doubt applicability to your patient.

The report states that most of the surgeries (83%) involved open reduction and internal fixation with locking plates, and that the operations were performed by attending surgeons (82%), supervised senior residents (12%), or independent fellows and senior residents who had immediate access to an attending surgeon if required (6%). Based on your training and experience, you are confident that you would offer a similar procedure and perform it with similar competence.

On the other hand, the report also states that the physiotherapists provided a mean of 10 one-on-one sessions to each group, with a focus on restoring function and encouragement for additional home exercises. The proportions of participants who received education, exercises, stretching, soft-tissue techniques, and other modalities were similar between groups, as were the numbers of patients who performed additional home exercises (109 in surgical group versus 103 in non-surgical group). You are familiar with the practice of the physiotherapists in

your community and you are aware that they typically provide similar regimens to those in the trial for the surgical and non-surgical groups, so you believe that their care is unlikely to differ in an important way.

You conclude that this trial provides practical information about the effect of surgical versus non-surgical treatment applicable to your patient encounter. You return to your patient and discuss the results. The patient asks about anaesthetic and surgical complications, and you acknowledge that while rare, these are a possibility. After considering the lack of benefit with surgery and the potential risks for morbidity or serious harms, your patient – despite the experience of her friend - decides to proceed with non-surgical management. You provide her with a referral to an experienced physiotherapist and you make an appointment to see her for follow-up.

Table 1. Feasibility of methodological considerations for trials of surgical or other non-pharmacological therapies

Examples	Feasibility of methodological considerations					
	Remote randomization	Blinding*	Sham-control design	Split-body design	Expertise-based design	Mechanistic versus practical interpretation
Surgery	Yes[1]	Yes[44]	Yes[25]	Yes[66]	Yes[44]	Yes[61]
Interventional cardiology	Yes[67]	Yes[67]	Yes[68]	N/A	Yes[69]	Yes[67]
Interventional radiology	Yes[70]	Yes[70]	Yes[71]	Yes**	Yes[70]	Yes[70]
Cognitive behavioral therapy	Yes[65]	Yes[65]	Yes[72]	N/A	Yes[73]	Yes[65]
Rehabilitation therapy	Yes[74]	Yes[75]	Yes[76]	Yes**	Yes[75]	Yes[75]
Occupational therapy	Yes[77]	Yes[78]	Yes[79]	N/A	Yes[78]	Yes[77]
Acupuncture	Yes[80]	Yes[80]	Yes[80]	Yes**	Yes[81]	Yes[80]

N/A = not applicable
*In the absence of a sham-control design, blinding of outcome assessors, data analysts, and/or those responsible for interpreting results
**Split-body design would be feasible for interventions that can be randomized between two homologous body parts within the same person

Table 2. Groups of individuals that can potentially be blinded in a randomized controlled trial.[13,23]

Group	Aim(s) to blind
Participants	To minimize bias due to placebo effects, differential use of effective co-interventions, influence on patient reports, selective reporting of symptoms, or differential loss to follow-up.
Healthcare providers	To minimize bias due to differential administration of the study interventions or effective co-interventions.
Data collectors	To minimize bias due to encouragement during performance testing, or variable frequency, timing, and recording during data collection.
Outcomes assessors	To minimize bias due to decisions about whether participants have had an outcome of interest, or differential interpretations of subjective outcomes.
Data analysts	To minimize bias during decision-making about analytic strategies, patient withdrawals, post hoc analyses, selection of time points, and other statistical issues.
Those interpreting results	To minimize bias during the presentation and interpretation of results.

SUPPLEMENT

Supplement 1. Split-body randomized controlled trials

Patients who require surgical or other non-pharmacological therapies on paired body parts present a unique opportunity to minimize bias. In *split-body trials*, two homologous body parts within the same person are randomly allocated to receive the experimental or the control interventions.[82–84] Patients in these trials act as their own controls, which facilitates that each treatment group starts with the same prognosis. To the extent that the initially apparent effects of the interventions are similar, split body designs can also facilitate blinding of participants and outcome assessors. Investigators have successfully used split-body trials to maintain prognostic balance for many oral, ocular, and extremity surgery interventions.[30,82,85,86]

For example, consider an RCT in which investigators compared computer-navigated total knee arthroplasty to non-navigated knee arthroplasty in 64 knees from 32 participants with bilateral osteoarthritis.[66] Each participant underwent simultaneous bilateral knee surgery using the computer-navigated technique on one side and the conventional technique on the other. The participants and outcome assessors were easily blinded because the wounds, dressings, and radiographic appearances of the implants were indistinguishable.

Section II: Observational Studies

Chapter 3

Methylprednisolone for the treatment of patients with acute spinal cord injuries: a propensity score-matched cohort study from a Canadian multi-center spinal cord injury registry

Reprinted from *Journal of Neurotrauma*; 32(21); Nathan Evaniew, Vanessa K. Noonan, Nader Fallah, Brian K. Kwon, Carly S. Rivers, Henry Ahn, Christopher S. Bailey, Sean D. Christie, Daryl R. Fourney, R. John Hurlbert, A. G. Linassi, Michael G. Fehlings, Marcel F. Dvorak, and The RHSCIR Network; Methylprednisolone for the Treatment of Patients with Acute Spinal Cord Injuries: A Propensity Score-Matched Cohort Study from a Canadian Multi-Center Spinal Cord Injury Registry; Pages 1674-83; Copyright 2015; Published by Mary Ann Liebert, Inc. doi:10.1089/neu.2015.3963.

JOURNAL OF NEUROTRAUMA 32:1674–1683 (November 1, 2015)
DOI: 10.1089/neu.2015.3963

Methylprednisolone for the Treatment of Patients with Acute Spinal Cord Injuries: A Propensity Score-Matched Cohort Study from a Canadian Multi-Center Spinal Cord Injury Registry

Nathan Evaniew,[1] Vanessa K. Noonan,[2,3] Nader Fallah,[2,3] Brian K. Kwon,[3] Carly S. Rivers,[2] Henry Ahn,[4,5] Christopher S. Bailey,[6] Sean D. Christie,[7] Daryl R. Fourney,[8] R. John Hurlbert,[9] A.G. Linassi,[10] Michael G. Fehlings,[5,11] Marcel F. Dvorak,[3] and The RHSCIR Network

Abstract

In prior analyses of the effectiveness of methylprednisolone for the treatment of patients with acute traumatic spinal cord injuries (TSCIs), the prognostic importance of patients' neurological levels of injury and their baseline severity of impairment has not been considered. Our objective was to determine whether methylprednisolone improved motor recovery among participants in the Rick Hansen Spinal Cord Injury Registry (RHSCIR).

We identified RHSCIR participants who received methylprednisolone according to the Second National Spinal Cord Injury Study (NASCIS-II) protocol and used propensity score matching to account for age, sex, time of neurological exam, varying neurological level of injury, and baseline severity of neurological impairment. We compared changes in total, upper extremity, and lower extremity motor scores using the Wilcoxon signed-rank test and performed sensitivity analyses using negative binomial regression.

Forty-six patients received methylprednisolone and 1555 received no steroid treatment. There were no significant differences between matched participants for each of total (13.7 vs. 14.1, respectively; $p = 0.43$), upper extremity (7.3 vs. 6.4; $p = 0.38$), and lower extremity (6.5 vs. 7.7; $p = 0.40$) motor recovery. This result was confirmed using a multivariate model and, as predicted, only cervical (C1–T1) rather than thoracolumbar (T2–L3) injury levels ($p < 0.01$) and reduced baseline injury severity (American Spinal Injury Association [ASIA] Impairment Scale grades; $p < 0.01$) were associated with greater motor score recovery. There was no in-hospital mortality in either group; however, the NASCIS-II methylprednisolone group had a significantly higher rate of total complications (61% vs. 36%; $p = 0.02$)

NASCIS-II methylprednisolone did not improve motor score recovery in RHSCIR patients with acute TSCIs in either the cervical or thoracic spine when the influence of anatomical level and severity of injury were included in the analysis. There was a significantly higher rate of total complications in the NASCIS-II methylprednisolone group. These findings support guideline recommendations against routine administration of methylprednisolone in acute TSCI.

Key words: methylprednisolone; motor score; neurological recovery; propensity scored-matched; spinal cord injury

[1]Division of Orthopaedics, McMaster University, Hamilton, Ontario, Canada.
[2]Rick Hansen Institute, [3]University of British Columbia, Vancouver, Canada.
[4]St. Michael's Hospital, Toronto, Ontario, Canada.
[5]University of Toronto Spine Program, [11]Division of Neurosurgery, University of Toronto, Ontario, Canada.
[6]Division of Orthopaedic Surgery, Western University; London Health Services Centre, London, Ontario, Canada.
[7]Division of Neurosurgery–Halifax Infirmary, Dalhousie University, Halifax, Nova Scotia, Canada.
[8]Department of Surgery, [10]Department of Physical Medicine and Rehabilitation, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
[9]University of Calgary Spine Program, Calgary, Alberta, Canada.

Introduction

Traumatic spinal cord injuries (TSCIs) affect up to 500,000 people worldwide each year, and their high morbidity is associated with substantial individual and societal burden and socioeconomic impact.[1,2] Patients with TSCIs often experience devastating neurological impairments, and they frequently require complex long-term multidisciplinary care.[3,4] Total health care costs related to TSCIs exceed $10 billion annually in the United States alone, and lifetime per person direct and indirect costs can exceed $3 million.[5,6] TSCIs most commonly affect young males and result from road traffic accidents, but recent reports also highlight their increasing incidence in older adults as a result of low-energy falls.[2,7–9]

The identification of novel interventions to reduce the morbidity of TSCIs is an urgent ongoing research priority.[3,10] Methylprednisolone is a corticosteroid that was proposed to inhibit the inflammatory cascades contributing to secondary spinal cord damage after TSCIs, but its clinical utility remains controversial.[11,12] Considerable debate has centered on the validity of results from the landmark Second National Spinal Cord Injury Study (NASCIS-II), which was published in 1990.[11,13,14] In NASCIS-II, 487 patients with acute TSCIs were randomized to an initial bolus of 30 mg/kg of methylprednisolone followed by an infusion of 5.4 mg/kg per h for 23 h versus either naloxone or placebo.

The primary analysis among the 487 patients enrolled within 12 h in NASCIS-II failed to demonstrate a significant neurological benefit in the 162 patients randomized to methylprednisolone. However, a secondary analysis of 65 of these patients who received methylprednisolone within 8 h of injury suggested that this subgroup experienced improved neurological recovery at 6 months.[13,15] Critics of NASCIS-II highlight the limited credibility of subgroup testing, the potential importance of losses to follow-up, the small magnitude of observed treatment effects, and the arbitrary nature of an 8-h threshold.[14,16–19] Advocates discuss a lack of otherwise high-quality evidence and cite indirect support elsewhere in the literature.[15,20]

The use of methylprednisolone has decreased dramatically in many centers, but some clinicians still report a belief in its efficacy or concerns about medical-legal pressure.[21–25] Potential harms include increased risks for respiratory, urinary tract, and wound infections, hyperglycemia, gastrointestinal hemorrhage, steroid-induced myopathy, and all-cause mortality.[17,26,27] Early critical reviews of the NASCIS studies recommended that methylprednisolone administration not be considered a ''standard of care'' for acute TSCI, but rather, a treatment option. More recently, the 2013 ''Guidelines for the Management of Acute Cervical Spine and Spinal Cord Injuries'' recommended against the routine administration of methylprednisolone for the treatment of acute TSCIs.[28–30]

Recent evidence from the Rick Hansen Spinal Cord Injury Registry (RHSCIR) suggests that the prognostic importance of patients' neurological level of injury in combination with the baseline severity of their neurological impairments may have been previously overlooked.[3] Controlling for the joint distribution of these two variables in TSCI research might increase the likelihood of detecting true treatment effects while simultaneously avoiding spurious or misleading results.[31] In this study, our primary objective was to determine whether the NASCIS-II regimen of methylprednisolone started within 8 h of injury improved motor recovery in comparison with no steroid treatment among RHSCIR patients with acute TSCIs. Our secondary objectives were to consider the effect of patients' neurological level of injury and the baseline severity of their neurological impairments on motor recovery, and to compare rates of complications between groups.

Methods

Study design

We performed a propensity score-matched cohort study using patient data that were prospectively collected in RHSCIR. RHSCIR is an ongoing multi-center observational study of patients with acute TSCIs who are admitted to major trauma centers and accompanying rehabilitation centers in Canada.[32] There are currently 31 participating study sites in the RHSCIR network, which are located across 16 cities from 9 out of 10 Canadian provinces. This article's primary objective was specified *a priori* during the development of RHSCIR, along with several other research objectives.[32] Each participating site obtained local Research Ethics Board or Institutional Review Board approval prior to enrolling patients and collecting data.

Participants

Patients were eligible for this study if they were 18 years of age or older and they presented to a participating site following an acute TSCI. Patients with non-traumatic etiologies of SCI such as infection, neoplasm, iatrogenic, or acute vascular causes were ineligible, but no exclusions were made on the basis of age, sex, medical co-morbidities, associated injuries, or planned treatment. According to the RHSCIR protocol, approximately 265 data elements were collected during participants' pre-hospital, acute, and rehabilitation phases of care. Further descriptions of the RHSCIR data elements, procedures, governance structure, and patient privacy and confidentiality framework are available elsewhere.[3,32,33]

We used the RHSCIR database to identify all patients from May 2004 to March 2014 who received either the NASCIS-II regimen of methylprednisolone started within 8 h of their acute injury or no steroid treatment. Patients who received regimens of methylprednisolone other than NASCIS-II, patients who received steroids other than methylprednisolone, and patients whose steroid status was indeterminate were excluded. Patients who received the NASCIS-II regimen followed by an additional 24 h of methylprednisolone were included.[15]

The indications for NASCIS-II methylprednisolone were not standardized across the participating sites, and patients could have received NASCIS-II methylprednisolone at RHSCIR acute care sites or at non-participating community hospitals prior to being transferred to an RHSCIR acute care site.

Data sources

Motor function scores were measured by trained physicians, nurse practitioners, or physiotherapists according to the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI).[34] ISNCSCI total motor scores (TMS) can range from 0 (absent motor function) to 100 (intact motor function) and comprise component upper extremity motor scores (UEMS; range 0–50), and lower extremity motor scores (LEMS; range 0–50). We considered patients' baseline motor scores to be those obtained on their admission to acute care and we considered patients' final motor scores to be those obtained at the time of their discharge to the community from acute care or inpatient rehabilitation.[31] Each ISNCSCI record was processed through a customized electronic algorithm that maintained consistency and high quality.[32]

We also retrieved the following variables from the RHSCIR database for each patient: age, sex, Body Mass Index, Glasgow Coma Scale and Injury Severity Score at admission, injury mechanism, Charlson Comorbidity Index,[35] whether or not patients underwent surgery, and RHSCIR study site. These data elements were collected by trained research personnel and entered into

standardized local RHSCIR databases before being exported to the RHSCIR national office for centralized quality checks.[32] Missing or ambiguous data were reconciled with local research coordinators, hospital health records, and medical chart abstraction whenever possible.

We collected rates of in-hospital mortality, urinary tract infections (UTIs), pneumonias, decubitus ulcers, deep vein thrombosis or pulmonary embolism, surgical site infections, and sepsis using *International Classification of Diseases*, 10th Revision (ICD-10) codes from the Canadian Institute for Health Information's Discharge Abstract Database.[36]

Statistical analysis

We used 1:1 propensity score matching based on logistic regression to match patients who received NASCIS-II methylprednisolone with controls who received no steroid treatment. To control for potential confounding, we matched according to varying neurological level of injury (cervical: C1–T1, or thoracic: T2–L3) and baseline severity of neurological impairments (ISNCSCI ASIA Impairment Scale A, B, C, or D), as well as age, sex, and time from injury to first neurological examination (<72 h, 72 h to one week, greater than one week, or unknown).[3,37–39]

Jitter plots and propensity histograms were used to verify the distribution of propensity scores in each group. Sensitivity analysis were performed to control for any residual imbalance by (i) comparing the matched groups while adjusting for the matched variables using negative binomial regression; and (ii) comparing the NASCIS-II methylprednisolone group against the full cohort of unmatched potential controls while adjusting for the same variables and RHSCIR site using negative binomial regression.[40] Goodness of fit was confirmed using the Akaike information criterion and the Bayesian information criterion.

Discrete variables are reported as counts or proportions, normally distributed continuous variables as means with standard deviations (SD), and skewed continuous variables as medians with interquartile ranges (IQR). We used parametric tests for data with normal distributions and non-parametric tests for data without normal distributions.[3,31] We compared unmatched groups with the independent samples *t* test using Levene's test to assess the equality of variance or the Mann-Whitney U test, and matched groups with the paired *t* test or the Wilcoxon signed-rank test. We used Pearson's χ^2 or Fisher's exact test for categorical data depending upon the number of the sample in each cell. Direct correlations were evaluated using Pearson's correlation coefficient.

Participants with missing data were excluded from each analysis and imputations were not performed.[18,41] Extreme outliers were removed from each group when comparing lengths of stay. All tests of significance were two-tailed and *p* values of less than 0.05 were considered significant. All analyses were performed using R 3.1 (CRAN: the Comprehensive R Archive Network at http://cran.r-project.org/), Excel 2011 (Microsoft Corp., Redmond, WA), and IBM SPSS Version 22, 2012 (SPSS Inc., Chicago IL).

Results

Participants

There were 2009 patients with acute TSCIs who consented to RHSCIR enrollment and were discharged to the community from acute care or inpatient rehabilitation (Fig. 1). Of these, we excluded 318 because their steroid administration status was indeterminate,

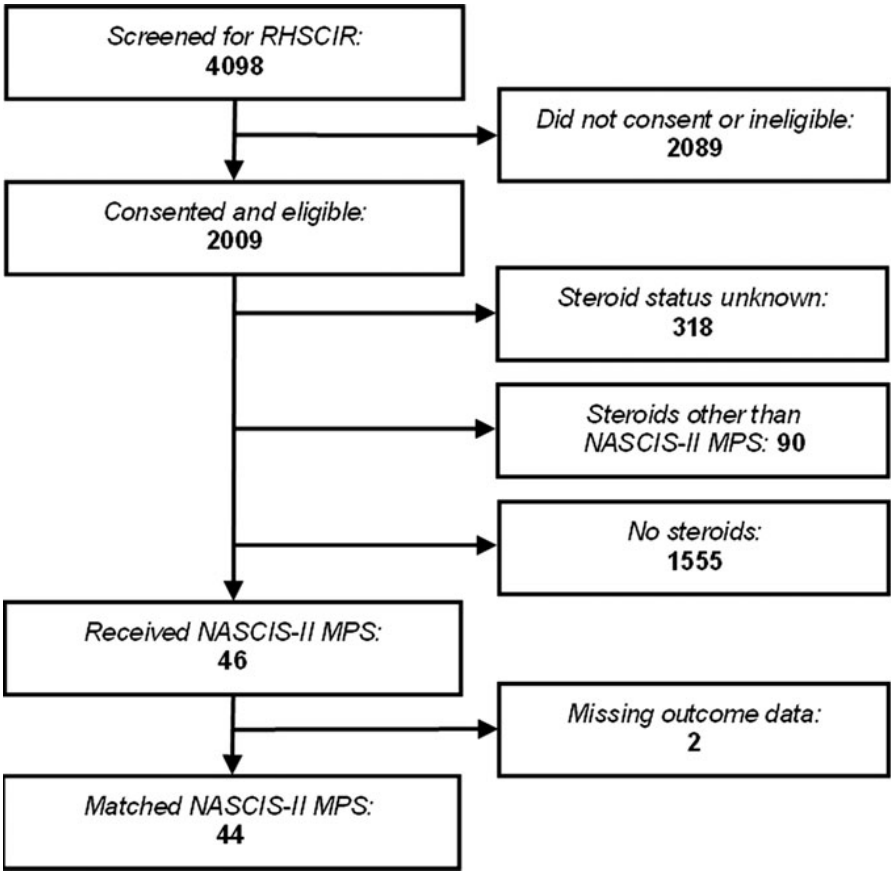


FIG. 1. Flow of participants in the RHSCIR and selection of patients for propensity score matching. MPS, methylprednisolone; RHSCIR, Rick Hansen Spinal Cord Injury Registry.

72 because they received dexamethasone, 5 because they received non-NASCIS-II methylprednisolone, and 14 because they received steroid regimens that were not further specified. In total, 46 consecutive patients were included who received the NASCIS-II regimen of methylprednisolone within 8 h of their acute injury, 5 of whom received the NASCIS-II regimen followed by an additional 24 h of methylprednisolone. There were 1555 included patients who received no steroid treatment.

Of the 46 patients who received NASCIS-II methylprednisolone, 20 were enrolled between 2004 and 2006, 25 between 2007 and 2010, and one was enrolled between 2011 and March 2014. NASCIS-II methylprednisolone was initiated at least once at 7 of the 18 acute care RHSCIR sites, but 25 of the 46 patients who received NASCIS-II methylprednisolone did so at a non-RHSCIR community hospital prior to being transferred to a RHSCIR site. These patients received their NASCIS-II methylprednisolone prior to their baseline neurological examinations, which were performed upon arrival at the RHSCIR site.

Baseline characteristics

There were no significant baseline differences between the group of patients who received NASCIS-II methylprednisolone ($n=46$) and the cohort of potential controls who received no steroid treatment ($n=1555$) except that those who received NASCIS II methylprednisolone had a significantly longer time from injury to first ISNCSCI examination (median 72 vs. 56 h, $p=0.01$; see Table 1).

Propensity score matching

Two of the 46 patients who received NASCIS-II methylprednisolone were excluded from the matched analysis because they had incomplete motor score outcome data. The remaining 44 were matched in a 1:1 ratio with controls who received no steroid treatment. The propensity score distributions within each group were similar (Fig. 2), and there were no significant differences in the proportions of patients with each combination of neurological level (cervical/thoracic) and ASIA Impairment Scale (A, B, C, or D), or any of the other baseline characteristics (Table 2). The median interval from injury to baseline neurological exam was 44 h (IQR 152) in the matched NASCIS-II methylprednisolone group and 31 h (IQR 170) in the matched no steroids group ($p=0.47$), whereas the median interval from injury to final neurological exam was 127 days (IQR 142) in the matched NASCIS-II methylprednisolone group and 117 days (IQR 138) in the matched no steroids group ($p=0.78$). Surgery was performed in 91% of the matched NASCIS-II methylprednisolone group and 82% of the matched no steroids group ($p=0.29$).

Motor score recovery

There were no significant differences in motor recovery between the matched NASCIS-II methylprednisolone group and the matched no steroids group for each of TMS ($p=0.43$), UEMS ($p=0.38$), and LEMS ($p=0.40$; see Fig. 3). Patients in the matched NASCIS-II methylprednisolone group experienced a mean TMS recovery of 13.7 points (SD 15.6), compared with 14.1 points (SD 21.6) for patients in the matched no steroids group. The mean UEMS recovery was 7.3 points (SD 8.4) in the matched NASCIS-II methylprednisolone group and 6.4 points (SD 12) in the matched no steroids group, and the mean LEMS recovery was 6.5 points (SD 10.7) in the matched NASCIS-II methylprednisolone group and 7.7 points (SD 12.5) in the matched no steroids group.

Table 1. Characteristics of Patients Who Received NASCIS-II Methylprednisolone or No Steroid Treatment

Characteristic	*NASCIS-II MPS* (n=*46*)	*No steroids* (n=*1555*)	P *value*
Age: mean (SD)	45.9 (16.6)	45.0 (18.6)	0.82
Male sex: *n* (%)	38 (82.6)	1211 (77.9)	0.45
Injury to first neurological exam, hours: median (IQR)	72 (154)	56 (172)	0.01[a]
Injury to final neurological exam, days: median (IQR)	142 (96)	124 (100)	0.17[b]
ASIA Impairment Scale: *n*			
A	21	536	0.82
B	6	152	
C	8	264	
D	11	384	
Neurological level: *n*			
Cervical	32	796	0.39
Thoracic	14	458	
Neurological level and ASIA Impairment Scale: *n*			
Cervical			0.97[b]
A	12	260	
B	4	87	
C	6	173	
D	10	253	
Thoracic			
A	9	243	
B	2	56	
C	2	72	
D	1	76	
High-energy: *n* (%)			
High	21 (45.7)	791 (50.9)	0.16
Low	25 (54.3)	685 (44.1)	
Unknown	0	79 (5.1)	
Treated with surgery: *n* (%)			
Yes	42 (91.3)	1247 (80.2)	0.13
No	4 (8.7)	230 (14.8)	
Unknown	0	78 (5)	
Injury to time of surgery, h: median (IQR)	31.5 (39.75)	29 (41)	0.95[c]
Glasgow Coma Scale: mean (SD)	14.3 (2.7)	14.1 (5.9)	0.08
Body Mass Index: mean (SD)	26.7 (5.6)	26.1 (5.7)	0.47
Injury Severity Score: mean (SD)	25.2 (12.1)	27.2 (12)	0.31
Charlson Comorbidity Index: mean (SD)	0.19 (0.46)	0.2 (0.63)	0.71

[a]P value reported using categorical value.
[b]Mann Whitney U test was used.
[c]Patients treated without surgery were excluded.
ASIA, American Spinal Injury Association; IQR, interquartile range; MPS, methylprednisolone; NASCIS-II, Second National Spinal Cord Injury Study SD, standard deviation.

There was also no significant difference in motor recovery when we performed sensitivity analyses to compare the matched groups while adjusting for the matched variables using negative binomial regression (Table 3), or when we compared the NASCIS-II methylprednisolone group against the full cohort of unmatched potential controls ($n=1555$) while adjusting for the same variables and RHSCIR site (Table 4). When analyzing cervical and thoracic injuries separately, the methylprednisolone group and the matched groups had near identical mean motor score recovery. Using the Mann-Whitney U test to compare cervical patients treated with methylprednisolone versus matched patients and thoracic patients

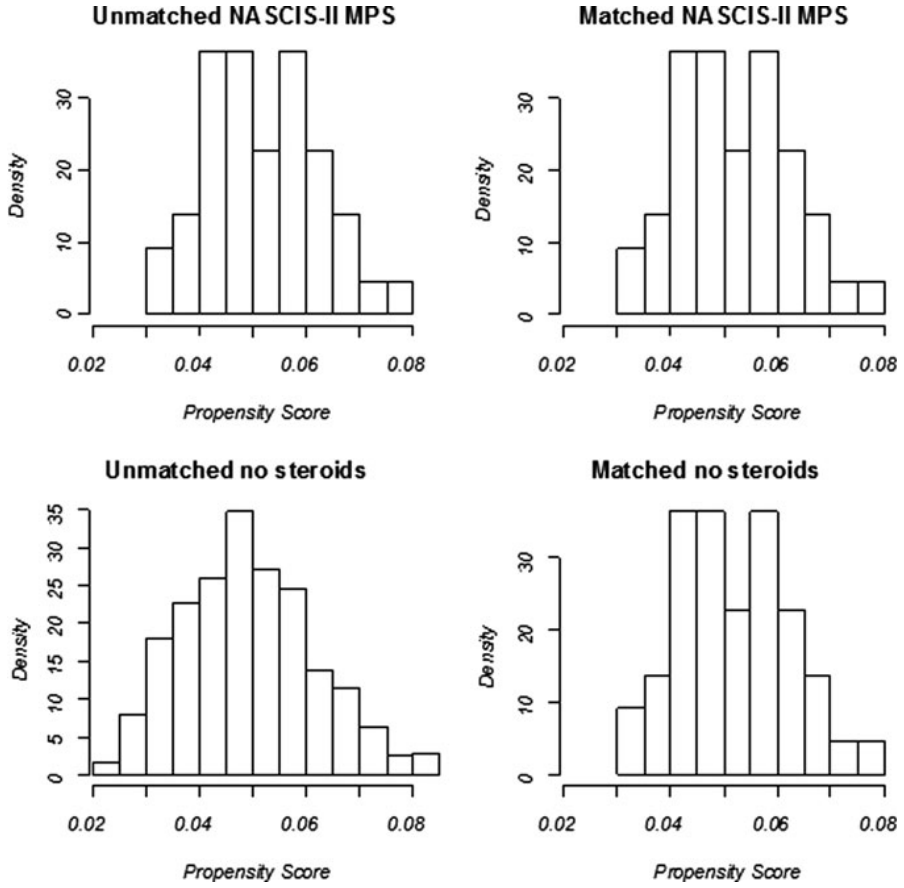


FIG. 2. Propensity histograms show the distributions of propensity scores among unmatched and matched patients who received NASCIS-II methylprednisolone or no steroids. MPS, methylprednisolone; NASCIS-II, Second National Spinal Cord Injury Study.

treated with methylprednisolone versus matched patients revealed no significant differences ($p = 0.65$ for cervical, $p = 0.69$ for thoracic).

In the analysis of the full cohort of unmatched potential controls, cervical rather than thoracic injury levels ($p < 0.01$) and reduced baseline injury severity (ASIA Impairment Scale A, B, C, or D; $p < 0.01$) were each significantly associated with greater TMS recovery.

Complications and length of stay

The most common complications in either matched group were urinary tract infections, decubitus ulcers, and pneumonias. None of the patients in either group experienced in-hospital mortality and there were no surgical site infections. The NASCIS-II methylprednisolone group had a significantly higher rate of total complications (61% vs. 36%; $p = 0.02$), but there were not significant differences in the rates of specific complications between groups (Table 5).

Patients in the NASCIS-II methylprednisolone group experienced a significantly shorter mean length of stay in acute care (34.4 days vs. 48.4 days; $p = 0.02$), but there were no significant differences in the lengths of stay at inpatient rehabilitation (106.7 vs. 117.9 days; $p = 0.45$) or the total lengths of stay, which is a combination of the acute care and inpatient rehabilitation lengths (mean 143.6 days vs. 152.9 days; $p = 0.28$).

Discussion

Using data prospectively collected in the RHSCIR, we performed a propensity-matched cohort study and found that the NASCIS-II regimen of methylprednisolone started within 8 h of injury did not improve motor recovery in comparison with no steroid treatment in patients with acute cervical and thoracic TSCIs. In a sensitivity analysis, cervical rather than thoracic injury level and reduced baseline injury severity were each associated with greater recovery. The NASCIS-II methylprednisolone group did not demonstrate a difference in motor recovery in cervical or thoracic patients when analyzed separately, but the methylprednisolone patients had a higher rate of total complications. There were no differences between groups for the rates of individual complications or for total length of stay.

Strengths and limitations

RHSCIR is part of the Translational Research Program of the Rick Hansen Institute, and it was created with the explicit purpose of facilitating clinical research to improve patient outcomes. Each data element was developed according to *a priori* research objectives and was standardized to optimize quality and accuracy[32]; ISNCSCI motor scores for this study were collected by trained clinical research staff and were verified using a customized electronic algorithm.[42] Administration of the NASCIS-II bolus and infusion of methylprednisolone were confirmed to begin within 8 h of patients' injuries, as per this protocol.

The timing of ISNCSCI examinations was not standardized, and differences in timing could have introduced bias in the results. Early baseline examinations risk confounding due to spinal shock, and delayed baseline examinations risk missing early recovery. For

Table 2. Propensity Score Matching of Patients Who Received NASCIS-II Methylprednisolone with Controls Who Received No Steroid Treatment

Characteristic	*NASCIS-II MPS* (n=*44*)	*No steroids* (n=*44*)	P *value*
Matched variables			
Age: mean (SD)	45.4 (16.2)	45.5 (16.6)	0.97
Male sex: *n* (%)	36 (81.8)	41 (93.2)	0.2
Injury to first neurological exam, hours: median (IQR)	44 (152)[a]	31 (170)	0.47
ASIA Impairment Scale: *n*			
A	21	19	0.90
B	6	5	
C	7	7	
D	10	13	
Neurological level: *n*			
Cervical	31	33	0.63
Thoracic	13	11	
Neurological level and ASIA Impairment Scale: *n*			
Cervical			0.99[b]
A	12	11	
B	4	4	
C	6	6	
D	9	12	
Thoracic			
A	9	8	
B	2	1	
C	1	1	
D	1	1	
Unmatched variables			
High-energy: *n* (%)			
High	19 (43.2)	20 (45.5)	0.51[b]
Low	25 (56.8)	22 (50)	
Unknown	0	2 (4.5)	
Treated with surgery: *n* (%)			
Yes	40 (90.9)	36 (81.8)	0.29[b]
No	4 (9.1)	6 (13.6)	
Unknown	0	2 (4.5)	
Injury to time of surgery, h: median (IQR)	33 (41)	33 (26.25)	0.96[c]
Glasgow Coma Scale: mean (SD)	14.4 (1.7)	14.3 (2.7)	0.84
Body Mass Index: mean (SD)	26.9 (7.2)	26.7 (5.7)	0.86
Injury Severity Score: mean (SD)	25.2 (12.1)	25.5 (10.2)	0.83
Charlson Comorbidity Index: mean (SD)	0.19 (0.46)	0.3 (0.79)	0.93
Injury to final neurological exam, days: median (IQR)	127 (142)	117 (138)	0.78

[a]Three observations were excluded as outliers (time > 512 h) for this variable only.

[b]*P* values reported are based on Fisher's exact test by applying Monte Carlo estimation.

[c]Patients treated without surgery were excluded.

ASIA, American Spinal Injury Association; IQR, interquartile range; MPS, methylprednisolone; SD, standard deviation.

example, the median time from injury to baseline examination was longer in the methylprednisolone group, and those patients who received methylprednisolone prior to their baseline examinations could have experienced some neurological recovery that was not captured. Nonetheless, Marino and colleagues showed that delays in baseline examinations are of minimal importance as long as they are conducted within 7 days.[43] Neurological improvement may continue up to or beyond one year,[44] but Pollard and Apple reported that more than 70% of neurological recovery occurs before discharge from rehabilitation.[45]

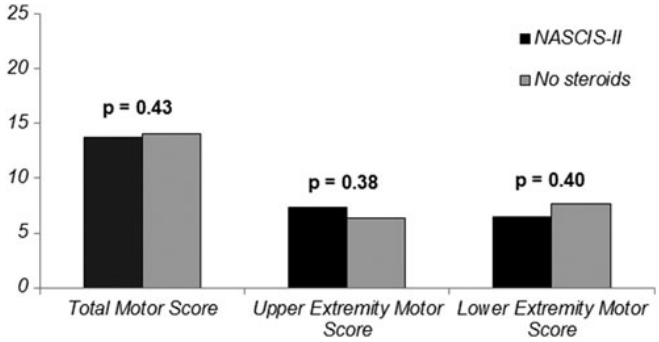


FIG. 3. Early motor recovery for patients who received NASCIS-II methylprednisolone(MPS; $n=44$) compared with matched controls who received no steroids ($n=44$). *P* values are from Wilcoxon signed-ranks tests. NASCIS-II, Second National Spinal Cord Injury Study.

We identified only 46 patients who received the NASCIS-II regimen of methylprednisolone within 8 h of their injuries since 2005. This sample size is small and may limit confidence in our results, particularly the rates of complications. However, it is unlikely to reflect selection bias because RHSCIR includes all of the specialized acute care spine centers in Canada and methylprednisolone use is known to have sharply declined.[21,23] For comparison, it is worthwhile to note that the analyses of the NASCIS II motor score improvements reported in a 2012 Cochrane Review rely on only 65 patients who received the NASCIS II protocol within 8 h.[15] Our finding that the frequency of NASCIS-II methylprednisolone administration has decreased over time suggests that the NASCIS-II protocol has fallen into widespread disfavor in Canada.

Table 3. Negative Binomial Regression Model of Total Motor Score Recovery on Age, Sex, Body Mass Index, ASIA Impairment Scale, and Level of Injury among Matched Patients Who Received NASCIS-II Methylprednisolone (*N*=44) or No Steroid Treatment (*N*=44)

Variable	*Coefficient*	*95% CI: lower*	*95% CI: upper*	P *value*
Age	−0.01	−0.03	−0.00	0.08
Male sex	−0.04	−0.81	0.80	0.99
BMI	−0.04	−0.08	−0.01	0.11
ASIA Impairment Scale A	−1.59	−2.31	−0.87	**<0.01**
ASIA Impairment Scale B	0.69	−0.10	1.48	0.09
ASIA Impairment Scale C	1.17	0.46	1.88	**<0.01**
ASIA Impairment Scale D[a]	-	-	-	-
Cervical	1.08	0.52	1.65	**<0.01**
Thoracic[a]	-	-	-	-
NASCIS-II MPS	0.04	−0.44	0.52	0.87
No steroids[a]	-	-	-	-

[a]Reference value.

AIS, American Spinal Injury Association; BMI, Body Mass Index; CI, confidence interval; MPS, methylprednisolone; NASCIS-II, Second National Spinal Cord Injury Study.

Bolded values were statistically significant.

Table 4. Negative Binomial Regression Model of Total Motor Score Recovery on Age, Sex, ASIA Impairment Scale, Level of Injury, and Site among Unmatched Patients Who Received NASCIS-II Methylprednisolone ($N=44$) or No Steroid Treatment ($N=1555$)

Variable	*Coefficient*	*95% CI: lower*	*95% CI: upper*	P *value*
Age	0.00	0.00	0.01	0.05
Male sex	0.02	−0.18	−0.22	0.87
BMI	−0.00	−0.01	0.01	0.77
ASIA Impairment Scale A	−0.45	−0.68	−0.22	**<0.01**
ASIA Impairment Scale B	0.82	0.56	1.08	**<0.01**
ASIA Impairment Scale C	1.13	0.96	1.3	**<0.01**
ASIA Impairment Scale D[a]	-	-	-	-
Cervical	0.96	0.77	1.15	**<0.01**
Thoracic[a]				
NASCIS-II MPS	−0.13	−0.43	0.16	0.38
No steroids[a]	-	-	-	-

[a]Reference value.

AIS, American Spinal Injury Association; BMI, Body Mass Index; CI, confidence interval; MPS, methylprednisolone; NASCIS-II, Second National Spinal Cord Injury Study.

Bolded values were statistically significant.

We excluded 318 patients whose steroid administration status was indeterminate because many of these patients received various steroid preparations peri-operatively for off-label neuro-protective indications, and we chose not to impute missing data in order to avoid introducing extra variability.[46] We also excluded patients who received steroid regimens other than NASCIS-II methylprednisolone in order to minimize confounding.[47]

Propensity score matching is an analytical technique that pairs treated and untreated patients on the basis of their conditional probability of receiving an intervention according to a set of observed co-variates.[37,38] Propensity score matching is more efficient than conventional multivariable regression when there are large differences in important prognostic characteristics between treatment groups, but its validity depends on the appropriate selection of covariates, matching techniques, and methods of final data analysis.[39]

Our propensity scores controlled for patients' neurological levels of injury and the baseline severity of their impairments, but our small sample precluded further differentiation according to high (C1–C4) versus low (C5–T1) cervical injuries or thoracic (T2–T10) versus thoracolumbar (T11–L2)injuries.[3] We were also unable to control for potential clustering due to local co-interventions at each RHSCIR site because more than half of the patients who received NASCIS-II methylprednisolone did so before arriving at a RHSCIR site. Propensity score matching cannot adjust for unknown confounders.[48]

Table 5. Rates of In-hospital Complications for Patients Who Received NASCIS-II Methylprednisolone Compared with Matched Controls Who Received No Steroids

Outcome	*NASCIS-II MPS* (n=*44*)	*No steroids* (n=*44*)	P *value*
Mortality	0	0	-
Urinary tract infection	11	9	0.61
Decubitus ulcer	6	2	0.27
Pneumonia	7	4	0.52
Deep vein thrombosis/ pulmonary embolism	2	0	0.49
Surgical site infection	0	0	-
Sepsis	1	1	-
Total	27	16	0.02

MPS, methylprednisolone; NASCIS-II, Second National Spinal Cord Injury Study.

Our approach to collecting complications data according to ICD-10 codes from a national database is known to be at risk for under-reporting, and ICD-10 codes may have been applied differently across the sites. Street and associates showed that nearly twice as many adverse events per person can be identified by prospectively applying the Spine Adverse Events Severity System.[36] Our use of a composite endpoint for total complications was justified because the component endpoints are likely to be of similar importance to patients, occurred with similar frequency, and are likely to share similar underlying biological plausibility.[49,50]

The time from injury to first neurological examination was significantly longer in the group of patients who received NASCIS-II methylprednisolone in comparison with the larger cohort of potential controls who received no steroids, which may suggest that the patients who received NASCIS-II methylprednisolone had greater injury severity. However, we used propensity score matching and negative binomial regression to control for this potential confounder and the times from injury to first neurological examination were not significantly different between the matched group of patients who received no steroids. There were also no significant differences between the matched groups for Injury Severity Score, Glasgow Coma Scale, ASIA Impairment Scale, or neurological level of injury.

We prospectively verified whether the patients who received NASCIS II methylprednisolone did so within 8 h of their injuries, but it is possible that the effect of NASCIS-II methylprednisolone might further vary according to whether patients received it earlier or later within 8 h of their injuries. In NASCIS-III, the 24-h regimen of methylprednisolone begun within the first 3 h after injury was not as effective if its initiation was delayed until between 3 and 8 h.[26] Our study was not designed to investigate this issue, however, and we did not collect exact timing data to explore it.

Surgical timing may be an important modifiable determinant of the outcomes in the management of patients with TSCIs. Decompression prior to 24 h was associated with improved neurological outcomes among RHSCIR patients with ASIA B, C, or D cervical, thoracic, or thoracolumbar injuries,[31] and it was also associated with improved outcomes in the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS).[51] A multivariate analysis of the STASCIS data suggested that methylprednisolone could have a synergistic effect with early decompression, and the incidence of wound infections among patients who received NASCIS-II methylprednisolone was lower in STASCIS than in the NASCIS-II trial.[52] However, STASCIS included only patients with cervical SCIs, who were more likely to undergo anterior surgery rather than posterior surgery, which may explain the reduced infection rates.[20] It is unlikely that surgical timing was a confounder in our study because the difference in the timing of surgery between the matched groups was not significant.

Relation to previous literature

Our results support a considerable body of literature that fails to demonstrate a benefit attributable to methylprednisolone for neurological functional recovery in patients with acute TSCIs, and our study is the first to adjust for patients' neurological level of injury

and the baseline severity of their impairments.[16,28,53,54] The original NASCIS-I trial found no significant differences in motor recovery at 6 months among 330 patients who were randomized to high- versus low-dose 10-day regimens of methylprednisolone,[55] the primary analysis of NASCIS-II found no significant differences in motor recovery at 6 months among 487 patients who were randomized to a 24-h regimen of methylprednisolone versus either naloxone or placebo,[13] and the primary analysis of NASCIS-III found no significant differences in motor recovery at 6 months among 499 patients who were randomized to 24 h or methylprednisolone, 48 h of methylprednisolone, or tirilazad.[26] A secondary analysis of 65 NASCIS-II patients who received methylprednisolone within 8 h of injury found that this subgroup experienced significantly improved sensory and motor recovery at 6 months.[13]

More recently, Chikuda and colleagues compared methylprednisolone against no steroid treatment in a propensity-matched analysis of their nationwide administrative database in Japan.[56] They matched 824 pairs of patients with cervical SCIs and found significantly higher rates of major complications including respiratory complications, urinary tract infections, sepsis, gastrointestinal bleedings, and pulmonary emboli in patients who received high doses of methylprednisolone, as well as longer lengths of stay. Their study did not specify whether lengths of stay included inpatient rehabilitation, did not include motor scores, did not control for levels or injury or severity of impairment, and did not verify that all patients received the NASCIS-II regimen within 8 h of their injuries. Three other small randomized trials and several earlier observational studies have been previously reviewed.[16,28,53,54]

Implications

Evidence-based medicine describes the careful integration of patient preferences and clinician expertise with the best available external evidence to facilitate decision-making, and clinicians, researchers, and other evidence users should consider the totality of relevant evidence before applying results to patient care.[57] Meta-analyses are systematic reviews in which the results from similar studies are combined using statistical tests to produce pooled treatment effects, and they are powerful tools that can synthesize conflicting literature and evaluate bias. However, they require high methodological credibility in order to avoid misleading conclusions.[58]

Bracken and Botelho and colleagues have each reported on meta-analyses that evaluate the effect of methylprednisolone against placebo in patients with TSCIs, but the conclusions from these studies are conflicting and each is limited by poor methodological credibility.[15,59] Neither ensured that the selection of studies was reproducible, neither explored possible explanations for between-studies differences in results, and neither study addressed the overall quality of the evidence or confidence in the pooled estimates.[58] According to the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach, confidence in pooled effect estimates depends on study design, risk of bias, imprecision, inconsistency, indirectness, publication bias, and other factors.[60] An updated independent meta-analysis could help resolve any ongoing controversy, and the open release of individual participant data for this purpose would allow adjustments for the prognostic importance of patients' neurological levels of injury and the baseline severity of their impairments.[61–63]

The clinical validation of novel interventions to treat patients with acute TSCIs remains an urgent ongoing research priority.[64,65] Randomized controlled trials are the most rigorous clinical research studies for investigating treatment effects and establishing causality, but their design and conduct for interventions in patients with acute TSCIs is challenging. The number of patients who might be eligible for enrollment at individual institutions is surprisingly small, and complex stratification is required to account for variability in baseline prognostic factors.[3,10] Multi-center trials can achieve sufficient power, but they require extensive coordination, collaboration, and resources.[66] Large observational studies can overcome some of these challenges, but they must be also appropriately designed and implemented in order to minimize bias.[48]

The Joint Section on Spine and Peripheral Nerves of the American Association of Neurological Surgeons and Congress of Neurological Surgeons recommended against the routine administration of methylprednisolone for the treatment of acute TSCIs in 2013.[28] Their guidelines highlight that methylprednisolone is not approved by the Food and Drug Administration for use in TSCIs, there is no Class I or Class II medical evidence supporting clinical benefit, and there is Class I, II, and III evidence suggesting harmful side effects including death. The Canadian Neurosurgical Society, the Canadian Spine Society, and the Canadian Association of Emergency Physicians have previously contributed to position statements recognizing insufficient evidence to support the use of high-dose methylprednisolone in acute TSCIs.[21,29]

Conclusions

NASCIS-II methylprednisolone started within 8 h of injury did not improve motor score recovery in RHSCIR patients with acute cervical or thoracic TSCIs. These findings support guideline recommendations against its routine administration, and validate trends toward decreasing utilization. Clinicians, researchers, and other evidence users should consider these results in the context of a considerable body of evidence, and should recognize that patients' neurological levels of injury and the baseline severity of their impairments are important prognostic factors that warrant further consideration.

Acknowledgments

The authors would like to thank the RHSCIR network and all the participating local RHSCIR sites: GF Strong Rehabilitation Centre, Vancouver General Hospital, Foothills Hospital, Glenrose Rehabilitation Hospital, Royal Alexandra Hospital, University of Alberta Hospital, Royal University Hospital, Saskatoon City Hospital, Winnipeg Health Sciences Centre, Toronto Western Hospital, Toronto Rehabilitation Institute, St. Michael's Hospital, Sunnybrook Health Sciences Centre, Hamilton General Hospital, Hamilton Health Sciences Regional Rehabilitation Centre, Victoria Hospital (London), University Hospital (London), Parkwood Hospital (London), The Ottawa Hospital Rehabilitation Centre, The Ottawa Hospital Civic Campus, Hopital de l'Enfant Jesus, Institut de readaptation en deficience physique de Quebec, Centre de readaptation Lucie-Bruneau, Institut de readaptation Gingras-Lindsay-de-Montreal, Hopital du SacreCœur de Montreal, Nova Scotia Rehabilitation Centre, QEII Health Sciences Centre, Saint John Regional Hospital, Stan Cassidy Centre for Rehabilitation, St. John's Health Sciences Centre, and L.A. Miller Rehabilitation Centre.

The Rick Hansen Spinal Cord Injury Registry and this work are supported by funding from the Rick Hansen Institute, Health Canada, Western Economic Diversification Canada, and the Governments of Alberta, British Columbia, Manitoba, and Ontario.

Author Disclosure Statement

No competing financial interests exist.

References

1. WHO. Spinal Cord Injury, Fact Sheet. Available at 2013.http://www.who.int/mediacentre/factsheets/fs384/en/. Accessed _____.
2. Singh, A., Tetreault, L., Kalsi-Ryan, S., Nouri, A., Fehlings, M.G. (2014). Global prevalence and incidence of traumatic spinal cord injury. Clin. Epidemiol. 6, 309–331.
3. Dvorak, M.F., Noonan, V.K., Fallah, N., Fisher, C.G., Rivers, C.S., Ahn, H., Tsai, E.C., Linassi, A.G., Christie, S.D., Attabib, N., Hurlbert, R.J., Fourney, D.R., Johnson, M.G., Fehlings, M.G., Drew, B., Bailey, C.S., Paquet, J., Parent, S., Townson, A., Ho, C., Craven, B.C., Gagnon, D., Tsui, D., Fox, R., Mac-Thiong, J.M., Kwon, B. (2014). Minimizing errors in acute traumatic spinal cord injury trials by acknowledging the heterogeneity of spinal cord anatomy and injury severity: an observational Canadian cohort analysis. J. Neurotrauma 31, 1540–1547.
4. Noonan, V.K., Fallah, N., Park, S.E., Dumont, F.S., Leblond, J., Cobb, J., Noreau, L. (2014). Health care utilization in persons with traumatic spinal cord injury: the importance of multimorbidity and the impact on patient outcomes. Top Spinal Cord Inj. Rehabil. 20, 289–301.
5. DeVivo, M.J. (1997). Causes and costs of spinal cord injury in the United States. Spinal Cord 35, 809–813.
6. Krueger, H., Noonan, V.K., Trenaman, L.M., Joshi, P., Rivers, C.S. (2013). The economic burden of traumatic spinal cord injury in Canada. Chronic Inj. Can. 33, 113–122.
7. Noonan, V.K., Fingas, M., Farry, A., Baxter, D., Singh, A., Fehlings, M.G., Dvorak, M.F. (2012). Incidence and prevalence of spinal cord injury in Canada: a national perspective. Neuroepidemiology 38, 219–226.
8. Selvarajah, S., Hammond, E.R., Haider, A.H., Abularrage, C.J., Becker, D., Dhiman, N., Hyder, O., Gupta, D., Black, J.H. 3rd, Schneider E.B. (2014). The burden of acute traumatic spinal cord injury among adults in the United States: an update. J. Neurotrauma 31, 228–238.
9. Wyndaele, M., Wyndaele, J.J. (2006). Incidence, prevalence and epidemiology of spinal cord injury: what learns a worldwide literature survey? Spinal Cord 44, 523–529.
10. Lee, R.S., Noonan, V.K., Batke, J., Ghag, A., Paquette, S.J., Boyd, M.C., Fisher, C.G., Street, J., Dvorak, M.F., Kwon, B.K. (2012). Feasibility of patient recruitment into clinical trials of experimental treatments for acute spinal cord injury. J. Clin. Neurosci 19, 1338–1343.
11. Kwon, B.K., Tetzlaff, W., Grauer, J.N., Beiner, J., Vaccaro, A.R. (2004). Pathophysiology and pharmacologic treatment of acute spinal cord injury. Spine J. 4, 451–464.
12. Lenzer, J. (2013). Why we can't trust clinical guidelines. BMJ. 346, f3830.
13. Bracken, M.B., Shepard, M.J., Collins, W.F., Holford, T.R., Young, W., Baskin, D.S., Eisenberg, H.M., Flamm, E., Leo-Summers, L., Maroon, J., Marshall, L.F., Perot, P.L. Jr., Piepmeier, J., Sonntag, V.K.H., Wagner, F.C., Wilberger, J.E., Winn, H.R. (1990). A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury. Results of the Second National Acute Spinal Cord Injury Study. N. Engl. J. Med. 322, 1405–1411.
14. Coleman, W.P., Benzel, D., Cahill, D.W., Ducker, T., Geisler, F., Green, B., Gropper, M.R., Goffin, J., Madsen, P.W. 3rd, Maiman, D.J., Ondra, S.L., Rosner, M., Sasso, R.C., Trost, G.R., Zeidman, S. (2000). A critical appraisal of the reporting of the National Acute Spinal Cord Injury Studies (II and III) of methylprednisolone in acute spinal cord injury. J. Spinal Disord. 13, 185–199.
15. Bracken, M.B. (2012). Steroids for acute spinal cord injury. Cochrane Database Syst. Rev. 1, Cd001046.
16. Hurlbert, R.J. (2000). Methylprednisolone for acute spinal cord injury: an inappropriate standard of care. J. Neurosurg. 93, 1–7.
17. Hurlbert, R.J. (2014). Methylprednisolone for the treatment of acute spinal cord injury: point. Neurosurgery 61 Suppl 1, 32–35.
18. Akl, E.A., Briel, M., You, J.J., Sun, X., Johnston, B.C., Busse, J.W., Mulla, S., Lamontagne, F., Bassler, D., Vera, C., Alshurafa, M., Katsios, C.M., Zhou, Q., Cukierman-Yaffe, T., Gangii, A., Mills, E.J., Walter, S.D., Cook, D.J., Schünemann, H.J., Altman, D.G., Guyatt, G.H. (2012). Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ. 344, e2809.
19. Sun, X., Briel, M., Busse, J.W., You, J.J., Akl, E.A., Mejza, F., Bala, M.M., Bassler, D., Mertz, D., Diaz-Granados, N., Vandvik, P.O., Malaga, G., Srinathan, S.K., Dahm, P., Johnston, B.C., Alonso-Coello, P., Hassouneh, B., Walter, S.D., Heels-Ansdell, D., Bhatnagar, N., Altman, D.G., Guyatt, G.H. (2012). Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 344, e1553.
20. Fehlings, M.G., Wilson, J.R., Cho, N. (2014). Methylprednisolone for the treatment of acute spinal cord injury: counterpoint. Neurosurgery 61 Suppl 1, 36–42.
21. Hurlbert, R.J., Moulton, R. (2002). Why do you prescribe methylprednisolone for acute spinal cord injury? A Canadian perspective and a position statement. Can. J. Neurol. Sci. 29, 236–239.
22. Schroeder, G.D., Kwon, B.K., Eck, J.C., Savage, J.W., Hsu, W.K., Patel, A.A. (2014). Survey of cervical spine research society members on the use of high-dose steroids for acute spinal cord injuries. Spine (Phila. Pa. 1976) 39, 971–977.
23. Hurlbert, R.J., Hamilton, M.G. (2008). Methylprednisolone for acute spinal cord injury: 5-year practice reversal. Can. J. Neurol. Sci. 35, 41–45.
24. Miekisiak, G., Kloc, W., Janusz, W., Kaczmarczyk, J., Latka, D., Zarzycki, D. (2014). Current use of methylprednisolone for acute spinal cord injury in Poland: survey study. Eur. J. Orthop. Surg. Traumatol. 24, Suppl 1, S269–S273.
25. Druschel, C., Schaser, K.D., Schwab, J.M. (2013). Current practice of methylprednisolone administration for acute spinal cord injury in Germany: a national survey. Spine (Phila. Pa. 1976) 38, E669–E677.
26. Bracken, M.B., Shepard, M.J., Holford, T.R., Leo-Summers, L., Aldrich, E.F., Fazl, M., Fehlings, M., Herr, D.L., Hitchon, P.W., Marshall, L.F., Nockels, R.P., Pascale, V., Perot, P.L. Jr., Piepmeier, J., Sonntag, V.K., Wagner, F., Wilberger, J.E., Winn, H.R., Young, W. (1997). Administration of methylprednisolone for 24 or 48 hours or tirilazad mesylate for 48 hours in the treatment of acute spinal cord injury. Results of the Third National Acute Spinal Cord Injury Randomized Controlled Trial. National Acute Spinal Cord Injury Study. JAMA. 277, 1597–1604.
27. Edwards, P., Arango, M., Balica, L., Cottingham, R., El-Sayed, H., Farrell, B., Fernandes, J., Gogichaisvili, T., Golden, N., Hartzenberg, B., Husain, M., Ulloa, M.I., Jerbi, Z., Khamis, H., Komolafe, E., Laloë, V., Lomas, G., Ludwig, S., Mazairac, G., Muñoz Sanchéz, Mde L., Nasi, L., Olldashi, F., Plunkett, P., Roberts, I., Sandercock, P., Shakur, H., Soler, C., Stocker, R., Svoboda, P., Trenkler, S., Venkataramana, N.K., Wasserberg, J., Yates, D., Yutthakasemsunt, S., CRASH trial collaborators. (2005). Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. Lancet 365, 1957–1959.
28. Walters, B.C., Hadley, M.N., Hurlbert,R.J., Aarabi, B., Dhall, S.S., Gelb, D.E., Harrigan, M.R., Rozelle, C.J., Ryken, T.C., Theodore, N.; American Association of Neurological Surgeons; Congress of Neurological Surgeons. (2013). Guidelines for the management of acute cervical spine and spinal cord injuries: 2013 update. Neurosurgery 60, Suppl 1, 82–91.
29. Hugenholtz, H., Cass, D.E., Dvorak, M.F., Fewer, D.H., Fox, R.J., Izukawa, D.M., Lexchin, J., Tuli, S., Bharatwal, N., Short, C. (2002). High-dose methylprednisolone for acute closed spinal cord injury—only a treatment option. Can. J. Neurol. Sci. 29, 227–235.
30. Pharmacological therapy after acute cervical spinal cord injury. (2002). Neurosurgery 50, Suppl. 3, S63–S72.
31. Dvorak, M.F., Noonan, V.K., Fallah, N., Fisher, C.G., Finkelstein, J., Kwon, B.K., Rivers, C.S., Ahn, H., Paquet, J., Tsai, E.C., Townson, A., Attabib, N., Bailey, C.S., Christie, S.D., Drew, B., Fourney, D.R., Fox, R., Hurlbert, R.J., Johnson, M.G., Linassi, A.G., Parent, S., Fehlings, M.G. (2015). The influence of time from injury to surgery on motor recovery and length of hospital stay in acute traumatic spinal cord injury: an observational Canadian cohort study. J. Neurotrauma 32, 654–654.
32. Noonan, V.K., Kwon, B.K., Soril, L., Fehlings, M.G., Hurlbert, R.J., Townson, A., Johnson, M., Dvorak, M.F.; RHSCIR Network.(2012). The Rick Hansen Spinal Cord Injury Registry (RHSCIR): a national patient-registry. Spinal Cord 50, 22–27.
33. Noonan, V.K., Thorogood, N.P., Joshi, P.B., Fehlings, M.G., Craven, B.C., Linassi, G., Fourney, D.R., Kwon, B.K., Bailey, C.S., Tsai, E.C., Drew, B.M., Ahn, H., Tsui, D., Dvorak, M.F. (2013). Meeting the privacy requirements for the development of a multi-centre patient registry in Canada: the Rick Hansen Spinal Cord Injury Registry. Healthc. Policy 8, 87–99.
34. Kirshblum, S., Waring, W. (2014). Updates for the International Standards for Neurological Classification of Spinal Cord Injury. Phys. Med. Rehabil. Clin. N. Am. 25, 505–517, vii.

35. Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J. Chronic. Dis. 40, 373–383.
36. Street, J.T., Thorogood, N.P., Cheung, A., Noonan, V.K., Chen, J., Fisher, C.G., Dvorak, M.F. (2013). Use of the Spine Adverse Events Severity System (SAVES) in patients with traumatic spinal cord injury. A comparison with institutional ICD-10 coding for the identification of acute care adverse events. Spinal Cord 51, 472–476.
37. Biondi-Zoccai, G., Romagnoli, E., Agostoni, P., Capodanno, D., Castagno, D., D'Ascenzo, F., Sangiorgi, G., Modena, M.G. (2011). Are propensity scores really superior to standard multivariable analysis? Contemp. Clin. Trials 32, 731–740.
38. Garrido, M.M., Kelley, A.S., Paris, J., Roza, K., Meier, D.E., Morrison, R.S., Aldridge, M.D. (2014). Methods for constructing and assessing propensity scores. Health Serv. Res. 49, 1701–1720.
39. Streiner, D.L., Norman, G.R. (2012). The pros and cons of propensity scores. Chest 142, 1380–1382.
40. Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Borg Debono, V., Dillenburg, R., Fruci, V., Bawor. M., Lee, J., Wells, G., Goldsmith, C.H. (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC. Med. Res. Methodol. 13, 92.
41. Scharfstein, D.O., Hogan, J., Herman, A. (2012). On the prevention and analysis of missing data in randomized clinical trials: the state of the art. J. Bone Jt. Surg. Am. 94, Suppl 1, 80–84.
42. Schuld, C., Wiese, J., Franz, S., Putz, C., Stierle, I., Smoor, I.,Weidner, N.; EMSCI Study Group, Rupp, R. (2013). Effect of formal training in scaling, scoring and classification of the International Standards for Neurological Classification of Spinal Cord Injury. Spinal Cord 51, 282–288.
43. Marino, R.J., Burns, S., Graves, D.E., Leiby, B.E., Kirshblum, S., Lammertse, D.P. (2011). Upper- and lower-extremity motor recovery after traumatic cervical spinal cord injury: an update from the national spinal cord injury database. Arch. Phys. Med. Rehabil. 92, 369–375.
44. Kirshblum, S., Millis, S., McKinley, W., Tulsky, D. (2004). Late neurologic recovery after traumatic spinal cord injury. Arch. Phys. Med. Rehabil. 85, 1811–1817.
45. Pollard, M.E., Apple, D.F. (2003). Factors associated with improved neurologic outcomes in patients with incomplete tetraplegia. Spine (Phila. Pa. 1976) 28, 33–39.
46. Groenwold, R.H., Donders, A.R., Roes, K.C., Harrell, F.E., Moons, K.G. (2012). Dealing with missing outcome data in randomized trials and observational studies. Am. J. Epidemiol. 175, 210–217.
47. Kiwerski, J.E. (1993). Application of dexamethasone in the treatment of acute spinal cord injury. Injury 24, 457–460.
48. Hoppe, D.J., Schemitsch, E.H., Morshed, S., Tornetta, P., Bhandari, M. (2009). Hierarchy of evidence: where observational studies fit in and why we need them. J. Bone Jt. Surg. Am. 91, Suppl 3, 2–9.
49. Montori, V.M., Permanyer-Miralda, G., Ferreira-González, I., Busse, J.W., Pacheco-Huergo,V., Bryant, D., Alonso, J., Akl, E.A., Domingo-Salvany, A., Mills, E., Wu, P., Schünemann, H.J., Jaeschke, R., Guyatt, G.H. (2005). Validity of composite end points in clinical trials. BMJ. 330, 594–596.
50. Busse, J.W., Bhandari, M., Ferreira-González, I., Montori, V.M., Guyatt, G.H.(2012). Use andinterpretation of composite end points in orthopaedic trials. J. Bone Jt. Surg. Am. 94, Suppl 1, 65–69.
51. Fehlings, M.G., Vaccaro, A., Wilson, J.R., Singh, A., Cadotte, D., Harrop, J.S., Aarabi, B., Shaffrey, C., Dvorak, M,. Fisher, C., Arnold, P., Massicotte, E.M., Lewis, S., Rampersaud, R. (2012). Early versus delayed decompression for traumatic cervical spinal cord injury: results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). PLoS. One 7, e32037.
52. Wilson, J.R., Arnold, P.M., Singh, A., Kalsi-Ryan, S., Fehlings, M.G. (2012). Clinical prediction model for acute inpatient complications after traumatic cervical spinal cordinjury: a subanalysis from the Surgical Timing in Acute Spinal Cord Injury Study. J. Neurosurg. Spine 17, Suppl 1, 46–51.
53. Hurlbert, R.J., Hadley, M.N., Walters, B.C., Aarabi, B., Dhall, S.S., Gelb, D.E., Rozzelle, C.J., Ryken, T.C., Theodore, N. (2013). Pharmacological therapy for acute spinal cord injury. Neurosurgery 72, Suppl 2, 93–105.
54. Hurlbert, R.J. (2001). The role of steroids in acute spinal cord injury: an evidence-based analysis. Spine (Phila. Pa. 1976) 26, S39–S46.
55. Bracken, M.B., Collins, W.F., Freeman, D.F., Shepard, M.J., Wagner, F.W., Silten, R.M., Hellenbrand, K.G., Ransohoff, J., Hunt, W.E., Perot, P.L. Jr., Grossman, R.G., Green, B.A., Eisenberg, H.M., Rifkinson, N., Goodman, J.H., Meagher, J.N., Fischer, B., Clifton, G.L., Flamm, E.S., Rawe, S.E. (1984). Efficacy of methylprednisolone in acute spinal cord injury. JAMA. 251, 45–52.
56. Chikuda, H., Yasunaga, H., Takeshita, K., Horiguchi, H., Kawaguchi, H., Ohe, K., Fushimi, K., Tanaka, S. (2014). Mortality and morbidity after high-dose methylprednisolone treatment in patients with acute cervical spinal cord injury: a propensity-matched analysis using a nationwide administrative database. Emerg. Med. J. 31, 201–206.
57. Fisher, C.G., Wood, K.B. (2007). Introduction to and techniques of evidence-based medicine. Spine (Phila. Pa. 1976) 32, S66–S72.
58. Murad, M.H., Montori, V.M., Ioannidis, J.P., Jaeschke, R., Devereaux, P.J., Prasad, K., Neumann, I., Carrasco-Labra, A., Agoritsas, T., Hatala, R., Meade, M.O., Wyer, P., Cook, D.J., Guyatt, G. (2014). How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA. 312, 171–179.
59. Botelho, R.V., Daniel, J.W., Boulosa, J.L., Colli, B.O., Farias, Rde. L, Moraes, O.J., Pimenta, W.E. Jr, Ribeiro, C.H., Ribeiro, F.R., Taricco, M.A., Carvalho, M.V., Bernardo, W.M. (2009). Effectiveness of methylprednisolone in the acute phase of spinal cord injuries–a systematic review of randomized controlled trials. Rev. Assoc. Med. Bras. 55, 729–737.
60. Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E., Falck-Ytter, Y., Meerpohl, J., Norris, S., Guyatt, G.H. (2011). GRADE guidelines: 3. Rating the quality of evidence. J. Clin. Epidemiol. 64, 401–406.
61. Simmonds, M.C., Brown, J.V., Heirs, M.K., Higgins, J.P., Mannion, R.J., Rodgers, M.A., Stewart, L.A. (2013). Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. Ann. Intern. Med. 158, 877–889.
62. Fu, R., Selph, S., McDonagh,M., Peterson, K.,Tiwari, A., Chou, R., Helfand, M. (2013). Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis. Ann. Intern. Med. 158, 890–902.
63. Krumholz, H.M., Ross, J.S., Gross, C.P., Emanuel, E.J., Hodshon, B., Ritchie, J.D., Low, J.B., Lehman, R. (2013). A historic moment for open science: the Yale University Open Data Access project and medtronic. Ann. Intern. Med. 158, 910–911.
64. Kwon, B.K., Casha, S., Hurlbert, R.J., Yong, V.W. (2011). Inflammatory and structural biomarkers in acute traumatic spinal cord injury. Clin. Chem. Lab. Med. 49, 425–433.
65. Casha, S., Zygun, D., McGowan, M.D., Bains, I., Yong, V.W., Hurlbert, R.J. (2012). Results of a phase II placebo-controlled randomized trial of minocycline in acute spinal cord injury. Brain 135, 1224–1236.
66. Slobogean, G.P., Sprague, S., Bhandari, M. (2012). The tactics of large randomized trials. J. Bone Jt. Surg. Am. 94, Suppl 1, 19–23.

Address correspondence to:
Marcel F. Dvorak, MD
Blusson Spinal Cord Centre
818 West 10th Avenue
Vancouver, British Columbia V5Z 1M9
Canada

E-mail: Marcel.Dvorak@ubc.ca

Chapter 4

The surgical management of scoliosis:

A scoping review of the literature

Reprinted from *Scoliosis*; 10(1); Nathan Evaniew, Tahira Devji, Brian Drew, Devin Peterson, Michelle Ghert, and Mohit Bhandari; The surgical management of scoliosis: a scoping review of the literature; Pages 1-8; Copyright 2015; doi:10.1186/s13013-014-0026-3.

Evaniew et al. Scoliosis (2015) 10:1
DOI 10.1186/s13013-014-0026-3

Scoliosis

RESEARCH Open Access

The surgical management of scoliosis: a scoping review of the literature

Nathan Evaniew[1*], Tahira Devji[2], Brian Drew[1], Devin Peterson[1], Michelle Ghert[1] and Mohit Bhandari[1,2]

Abstract

Background: Scoping reviews are innovative studies that can map a range of evidence to convey the breadth and depth of a large field. An evidence-based approach to the wide spectrum of surgical interventions for scoliosis is paramount to enhance clinical outcomes. The objectives of this scoping review were to identify critical knowledge gaps and direct future research.

Methods: This study was completed according to the methodology of Arksey and O'Malley. Two reviewers performed duplicate systematic screening of eligibility. Studies were classified according to patient age, scoliosis etiology, outcomes reported, study design, and overall research theme.

Results: There were 1763 eligible studies published between 1966 and 2013. The literature focused on adolescents (83% of studies) with idiopathic scoliosis (72%). There was a dominance of observational designs (88%), and a paucity of randomized trials (4%) or systematic reviews (1%). Fifty six percent of studies were conducted in North America, followed by 23% in Europe and 18% in Asia. Few high-level studies investigated surgical indications, surgical approaches, surgical techniques, or implant selection. Patient important outcomes including function, health-related quality of life, pain, and rates or re-operation were infrequently reported.

Conclusions: Current research priorities are to (1) undertake high-quality knowledge synthesis and knowledge translation activities; (2) conduct a series of planning meetings to engage clinicians, patients, and methodologists; and (3) clarify outcome reporting and strategies for methodological improvement. Higher-quality studies are specifically needed to inform surgical indications, surgical approaches, surgical techniques, and implant selection. Engaging global partners may increase generalizability.

Keywords: Scoliosis, Spinal deformity, Scoping review, Systematic review, Clinical epidemiology

Introduction

'Scoliosis' encompasses a heterogeneous group of coronal and rotational spinal deformities that can affect patients of any age. Adolescent idiopathic scoliosis alone is associated with a substantial burden of health care utilization, but costs are even higher for patients with congenital or neuromuscular etiologies [1-5]. Likewise, degenerative scoliosis may affect up to 68% of adults greater than 70 years old and is a frequent cause of pain and disability [6,7]. An evidence-based approach to the wide range of surgical interventions for scoliosis is paramount to enhance clinical outcomes.

Knowledge translation is the dynamic and iterative process of summarizing, disseminating, exchanging, and applying research findings to improve patient outcomes and strengthen health care systems [8]. Comprehensive systematic reviews are the foundation of most knowledge translation activities, but understanding very broad or complex topics can be challenging. Systematic reviews related to the surgical management of scoliosis have been limited by narrow scope, heterogeneity across the included studies, or insufficient primary evidence [2,5,9,10].

Scoping reviews are innovative studies that can map a range of evidence to convey the breadth and depth of a large field. Scoping reviews are also powerful tools to guide ongoing knowledge synthesis and inform future research [11]. In contrast to standard systematic reviews, scoping reviews ask broader questions and do not perform

* Correspondence: nathan.evaniew@medportal.ca
[1]Division of Orthopaedics, Department of Surgery, McMaster University, 293 Wellington St N, Suite 110, Hamilton, ON L8L 8E7, Canada
Full list of author information is available at the end of the article

detailed assessments of individual studies. Scoping reviews also differ from narrative reviews in that they comprehensively and reproducibly identify relevant articles in order to minimize bias. Arksey and O'Malley's six-stage framework, which involves a systematic literature search, duplicate screening of eligibility, and the identification of overall research themes, is the foundation of modern scoping review methodology [11,12].

This study is a scoping review that was performed to synthesize the available literature reporting on the surgical management of scoliosis. The objectives of this study were to (1) identify critical knowledge gaps and (2) direct future research.

Methods

Eligibility criteria

All therapeutic clinical studies examining the surgical management of scoliosis were included. No restrictions were placed for patient age, scoliosis etiology, or date of publication. Studies of only non-surgical interventions and non-therapeutic study designs such as economic, prognostic, and diagnostic studies were excluded. Non-clinical research studies such as cadaveric biomechanical studies and basic science studies were excluded. Conference proceedings describing unpublished studies and studies that were published in languages other than English or could not be retrieved in English full-text were excluded. Narrative reviews and case reports of less than 5 patients were counted but excluded from the analysis.

Identification of studies

MeSH and EMTREE headings and subheadings were used in various combinations to query MEDLINE and EMBASE (up to June 6, 2013) in Ovid for potentially eligible articles (ie. "scoliosis/su [surgery] AND surgical procedures, operative/or orthopedics/su or spinal fusion/or general surgery/"). The headings were supplemented with free text to increase sensitivity (ie. "[scoliosis.ti,ab. OR curv*.ti,ab.] AND [operation or operative or operate or surgery or surgical).ti,ab.]"). The search strategy was also adapted in PubMed (up to June 6, 2013) to search for articles e-published ahead of print and not yet indexed on Ovid.

Screening and data extraction

Two reviewers performed duplicate screening of all titles and abstracts for eligibility using a piloted electronic screening form (Distiller SR, Evidence Partners 2013, Ottawa ON, Canada). All discrepancies were resolved through consensus.

Patient age and scoliosis etiology were classified according to the recommendations of the Scoliosis Research Society Terminology Committee and Working Group [13]. Reported outcomes were classified as radiological, functional, pain, rates of reoperations, rates of complications, physical exam outcomes, laboratory results, operative variables (such as blood loss or operating time), or other. All applicable classifications were recorded for each study. Total sample size, year of publication, and primary country of were also collected. The geographical distribution of studies was not adjusted for population or research density within each continent.

Study designs and levels of evidence

The two reviewers independently assessed study designs using the Centre for Evidence-Based Medicine in Oxford guidelines for therapeutic studies [14,15]. Higher quality randomized controlled trials (RCTs) were classified as Level I, while lesser quality RCTs and prospective non-randomized controlled studies were classified as level II. Retrospective controlled studies were classified as Level III, and uncontrolled studies were classified as Level IV. Reviewers were not blinded to authors, publication information, or any published level of evidence descriptions [16].

Literature themes

The two reviewers compiled a set of potential primary study themes through discussion and consensus after completing title and abstract screening [11]. The two reviewers then piloted the themes for face validity and content validity using a sample of 50 included studies. Minor revisions were made to clarify existing themes, add additional themes, and document discriminatory criteria for each theme. The single most relevant primary theme for each included study was collected, recognizing that some secondary themes would not be captured.

'Levels' described studies that reported on the selection of spinal levels for fusion; 'Approaches and Stages' described studies that reported on the effects of varying surgical approaches, adjunctive peri-operative interventions, or timing of consecutive procedures; 'Implants and Techniques' described studies that reported on the use of specific implant systems or varying surgical techniques related to implants. 'Indications' described studies that reported on the effect of an intervention in a specific or varying set of populations; 'Grafts' described studies that reported on the effect of varying graft materials; 'Blood' described studies that reported on interventions to minimize blood loss; 'Infection' described studies that reported on interventions to prevent or treat infections; 'Anaesthesia' described studies that reported on anaesthetic agents or techniques; 'Neuromonitoring' described studies that reported on neuromonitoring procedures and techniques; 'Analgesia' described studies that reported on methods to treat post-operative pain; 'Rehabilitation' described studies that reported on interventions related to rehabilitation in operatively treated patients;

'Psychological' described studies that reported on interventions to improve psychological outcomes.

Analysis

Inter-observer agreement for the reviewers' assessments of study eligibility was calculated with Cohen's kappa coefficient of agreement [17]. Inter-observer agreement for the reviewers' assessments of levels of evidence was calculated with the Intraclass Correlation Coefficient (IBM SPSS Version 21; Chicago IL, 2012). Descriptive statistics were used to summarize all other data. Discrete variables are reported as counts or proportions, normally distributed continuous variables are summarized as means with standard deviations, and skewed continuous variables are summarized as medians with interquartile ranges.

Results

Search results

The search strategy identified 15913 potentially relevant articles (Figure 1). Of these, 9313 were removed because they were duplicate references to the same articles from multiple databases. A further 1786 were excluded during screening of titles and 1544 were excluded during screening of titles and abstracts because they either did not relate to surgery or they did not relate to scoliosis. Of 3270 articles eligible for full text review, 618 were excluded because they were narrative reviews, 424 were excluded because they were case reports, 343 were excluded because they were not available as full-texts in English, and 122 were excluded because they were not relevant or were duplicates. In total, 1763 studies were included for data extraction and further analysis. Agreement between the two reviewers for eligibility was satisfactory (kappa = 0.78).

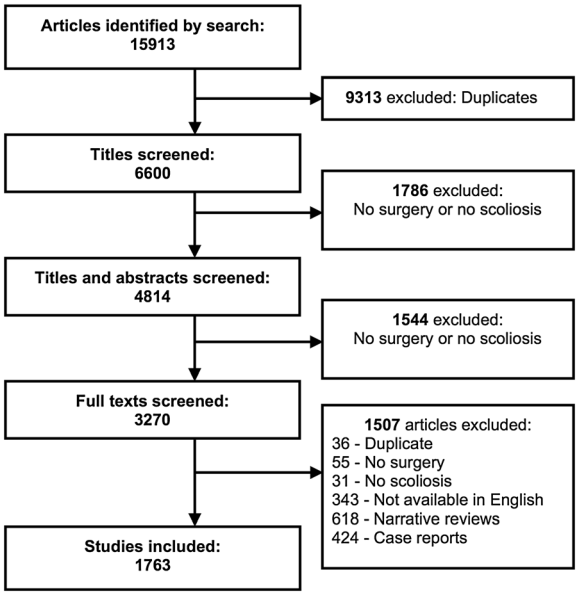


Figure 1 Flow diagram depicting the screening and review of potentially eligible articles.

Characteristics of included studies

Overall, 993 (56%) of the studies were conducted in North America, followed by 413 (23%) in Europe and 320 (18%) in Asia (Figure 2). Twenty-three studies were conducted by Australia and New Zealand together, and only seven each were conducted in each of South America and Africa. The total number of identified studies published globally per year rose from just one in 1966 to more than 130 in each of 2010, 2011, and 2012 (Figure 3a). Studies were most frequently published in *Spine* (711 studies), *European Spine Journal* (167), *Journal of Pediatric Orthopaedics* (142), *Journal of Bone and Joint Surgery - American Volume* (99), and *Clinical Orthopaedics and Related Research* (75).

The most frequently included age category was adolescent (83% of studies) (Figure 4), and the most frequently included etiology of scoliosis was idiopathic (72%) (Figure 5). Patients with neuromuscular scoliosis were included in 28% of studies and patients with congenital scoliosis were included in 17% of studies. Despite being a frequent cause of pain and disability in older adults, patients with degenerative scoliosis were included in just 5% of the identified studies [6,7]. More than one age category of patients was applicable in 33% of studies, and more than one etiological classification of scoliosis was applicable in 23% of studies. Radiological outcomes were reported in 66% of studies, rates of complications were reported in 62% of studies, and rates of reoperations were reported in 27% of studies (Figure 6). Functional outcomes or health-related quality of life were reported in just 20% of studies. The median sample size across all studies was 42 (IQR 24 to 87).

Study designs and levels of evidence

There were 65 prospective randomized controlled trials (4% of the included studies), 115 (7%) prospective non-randomized controlled studies, 571 (32%) retrospective controlled studies, and 983 (56%) uncontrolled studies (case series). Despite a dramatic increase in the total number of studies over time, the proportion of studies that were randomized controlled trials remained low, and has actually relatively decreased following a peak in 2004 (Figure 3b). There were 15 systematic reviews (<1%) and 14 systematic reviews and meta-analyses (<1%). Only three studies were classified as level I (<1%) and only 116 were graded as level II (7%), while 585 were classified as level III (33%) and 1059 were classified as level IV (60%) (Figure 7). Agreement between the two reviewers for levels of evidence was satisfactory (ICC = 0.771).

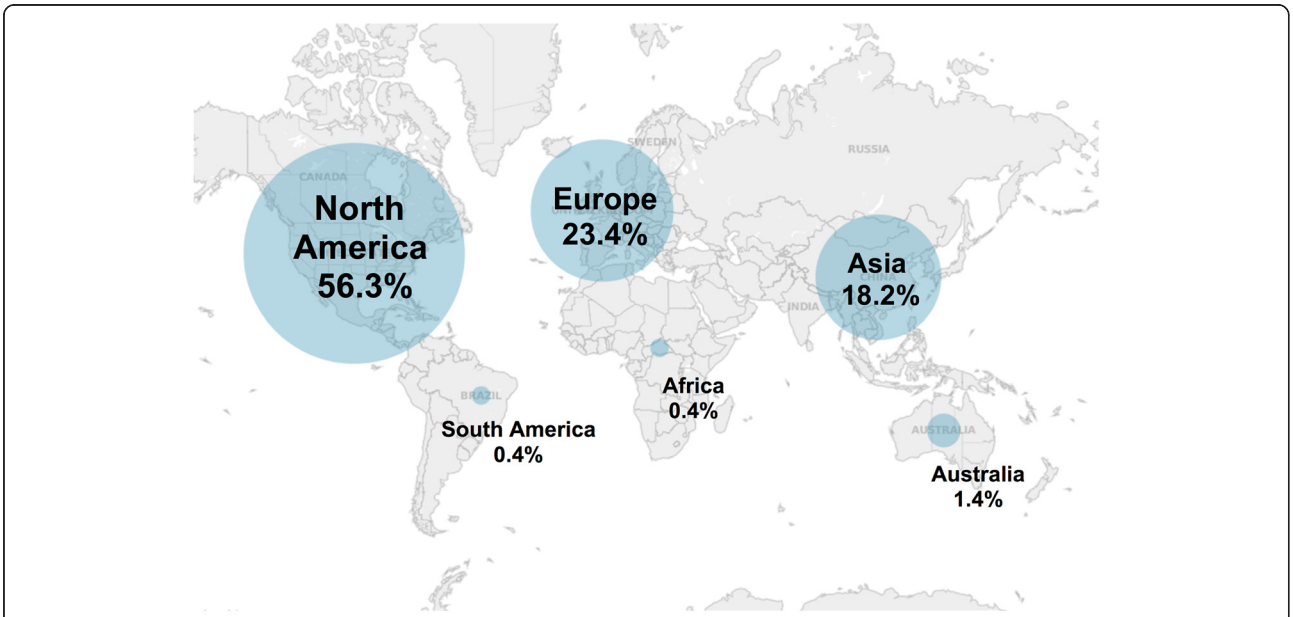


Figure 2 Global distribution of clinical research reporting on the surgical management of scoliosis. Percentages reflect raw proportions and are not adjusted for population or researcher density.

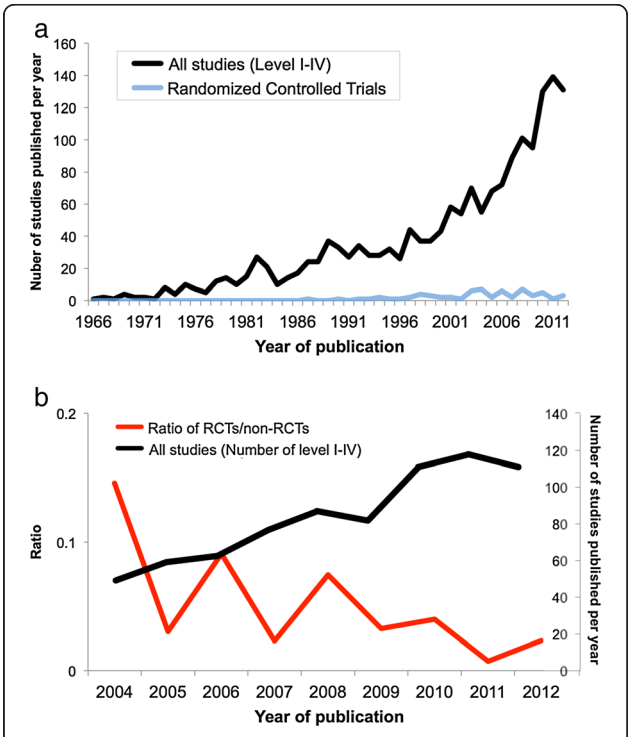


Figure 3 Temporal distribution of clinical research reporting on the surgical management of scoliosis. (a) Total volume of clinical research reporting on the surgical management of scoliosis over time; **(b)** Ratio of randomized controlled trials (RCTs) to non-RCTs since 2004 superimposed against the total volume of clinical research.

Literature themes

Studies most frequently investigated the effects of specific implants and specific surgical techniques (26%), followed by approaches and staging (21%), and indications for surgery (21%) (Figure 8). Ten percent of studies investigated the selection of spinal levels for fusion, 5% investigated neuromonitoring, and 4% investigated strategies to manage blood loss. Three percent investigated anaesthesic management, 3% investigated bone grafts or the use of bone graft substitutes, and 2% investigated post-operative pain management. Only 35 studies (2%) investigated the prevention or management of surgical site infections, 32 (2%) investigated interventions to improve psychological outcomes, and 17 (1%) investigated post-operative rehabilitation.

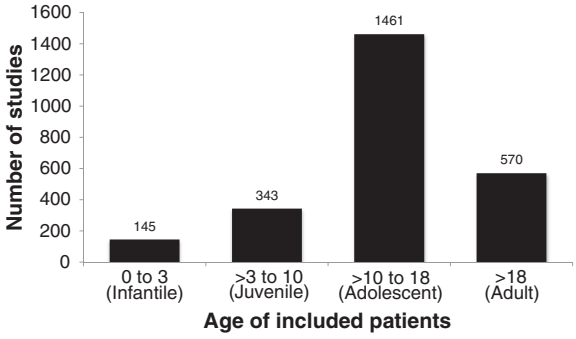


Figure 4 Distribution of clinical research reporting on the surgical management of scoliosis by age of included patients. All applicable age categories were recorded for each study.

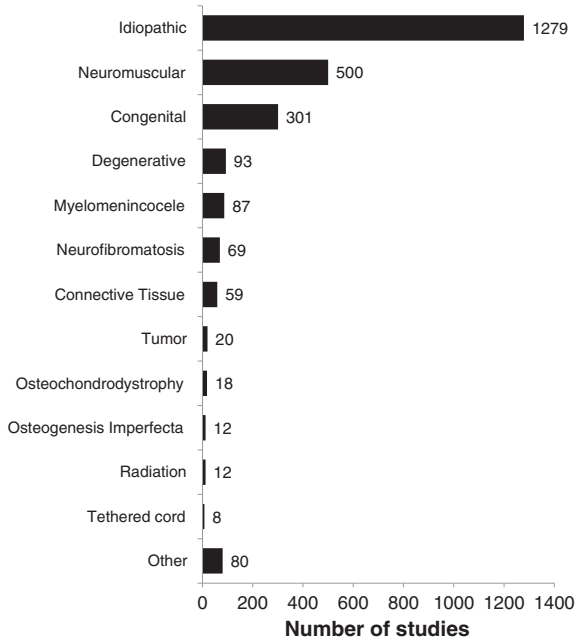


Figure 5 Distribution of clinical research reporting on the surgical management of scoliosis by etiology. All applicable etiological categories were recorded for each study.

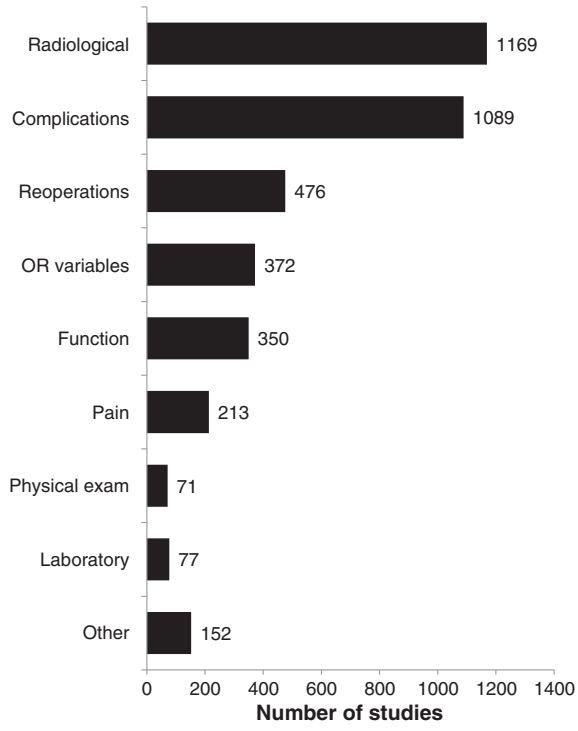


Figure 6 Distribution of clinical research reporting on the surgical management of scoliosis by reported outcomes. All applicable outcomes categories were recorded for each study.

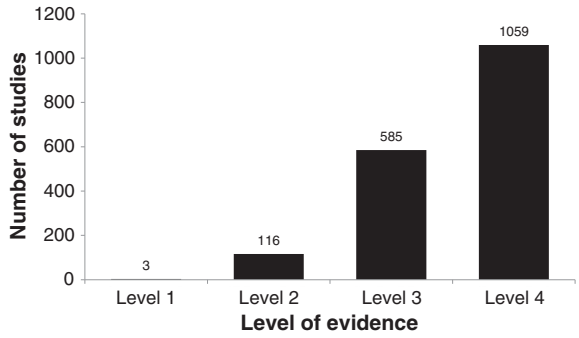


Figure 7 Distribution of clinical research reporting on the surgical management of scoliosis by level of evidence.

Of the 65 identified RCTs, 55 related to perioperative surgical care rather than direct surgical considerations: 19 investigated anesthetic management, 13 investigated strategies to manage blood loss, 13 investigated post-operative pain management, and 5 investigated neuro-monitoring. Four RCTs investigated surgical approaches or staging, 4 investigated bone grafts, 3 investigated implants or techniques, and 1 investigated the selection of spinal levels for fusion. Overall, 61 of the 65 RCTs reported on adolescent patients, and 60 reported on idiopathic curves. The median sample size of the RCTs was 36 (IQR 30 to 50), and the global distribution of RCTs paralleled the overall global distribution shown in Figure 2. Of the 14 meta-analyses, three each investigated indications for surgery, approaches and staging, implants and techniques, and psychological outcomes, and one each investigated blood loss and post-operative pain management.

Discussion

This study was a scoping review performed to summarize the literature available to guide the surgical management of scoliosis, identify critical gaps in current knowledge, and direct future research. The majority of the identified literature focused on adolescent patients with idiopathic scoliosis. There was a clear dominance of uncontrolled studies, and a striking paucity of RCTs. Few high-level studies investigated surgical indications, surgical approaches, surgical techniques, or implant selection. Patient important outcomes including function, health-related quality of life, pain, and rates or re-operation were infrequently reported.

Limitations

Of the 3270 studies identified for full-text screening, 343 full-texts could not be retrieved in English. Retrieving and translating non-English studies for systematic reviews can be technically prohibitive, but excluding them may produce misleading or exaggerated findings, particularly

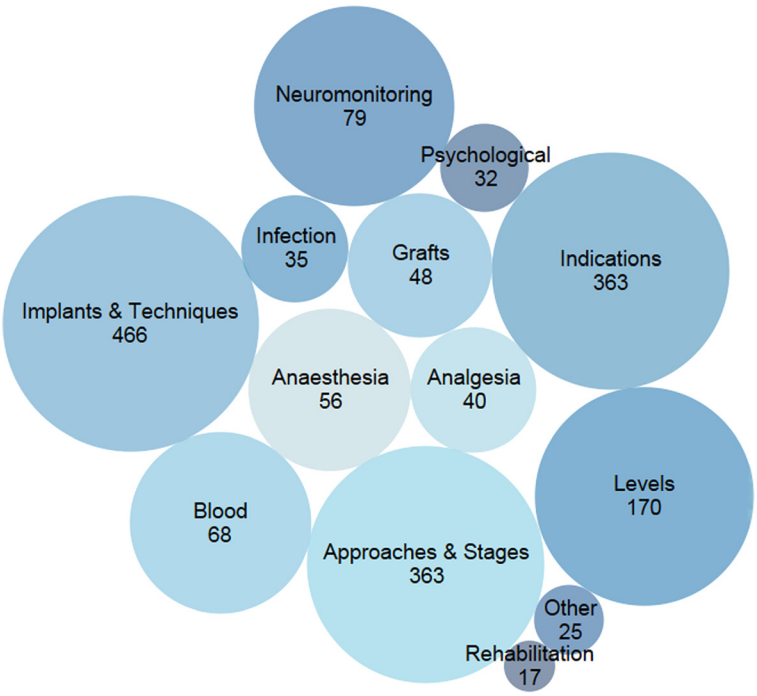


Figure 8 Illustrative plot of the primary research themes across studies reporting on the surgical management of scoliosis. The single most relevant primary theme was selected for each study. The size of each circle is proportional to the number of studies for each primary theme. The circle locations and colors are arbitrary.

when estimating the global distribution of research outside of North America and Europe [18]. Fortunately, large RCTs are most often widely available in high-impact English language journals, and the 343 excluded studies represent only approximately 10 percent of the eligible sample of studies. In their study of 130 systematic reviews, Moher et al. established that language restrictions in systematic reviews of conventional interventions do not seem to produce meaningful bias [18]. The relative lack of studies from India and China may reflect a tendency to publish in journals not indexed in the search databases or it may reflect a developing research infrastructure [19]. This issue highlights an opportunity to engage global partners in future studies [20].

The thematic framework was developed *ad hoc* and the identified domains have not been previously reported. This study's application of the scoping review framework to the scoliosis literature was entirely novel, and the thematic domains were developed after reviewing all titles and abstracts according to the Arksey and O'Malley framework [11,12]. Themes, age, etiology, and reported outcomes were not extracted in duplicate, but consensus

Table 1 Research gaps and future research directions for the surgical management of scoliosis

Research gaps	Future research directions
There are few focused systematic reviews relative to the extensive scoliosis literature, reflecting a lack of emphasis on knowledge synthesis and knowledge translation.	**Knowledge synthesis:** Perform a series of high-quality focused systematic reviews examining important clinical questions.
There is a striking paucity of randomized controlled trials (RCTs), and the existing RCTs are characterized by generally small sample sizes.	**Knowledge translation:** Use existing systematic reviews to inform a series of evidence-based decision aids and preliminary clinical practice guidelines.
Very few high-level studies have investigated surgical indications, surgical approaches, surgical techniques, or implant selection.	**Future RCTs:** Conduct a series of surveys or planning meetings to engage clinicians, patients, methodologists, and other knowledge users in the design and conduct of future large RCTs.
Patient important outcomes such as function, health-related quality of life, pain, and rates or re-operation have been infrequently reported in comparison to radiological outcomes and rates of complications.	**Ongoing scoping work:** Clarify inconsistent outcome reporting and identifying practical strategies for methodological improvement.

meetings were used to document clear definitions and ensure consistency. Further, the themes were piloted by each of the reviewers for face- and content- validity.

Implications for research

Critical knowledge gaps and directions for future research are summarized in Table 1. The first priority is to focus on knowledge synthesis and effective knowledge translation in order to optimize the impact of existing research. This scoping review identified 1763 relevant articles; however, even with such a large number of publications, there were only 15 prior systematic reviews and 14 prior meta-analyses. This scoping review can guide a series of high-quality focused systematic reviews on clinically important topics with identified robust data. Likewise, this scoping review can also inform evidence-based decision aids or preliminary clinical practice guidelines [11].

The second priority is to engage clinicians, patients, methodologists, and other knowledge users in the design and conduct of future large RCTs. A series of planning meetings could clarify research questions, strengthen a collaborative network, and optimize strategies for successful potential funding applications. There were only 65 RCTs identified, and these trials were generally characterized by small sample sizes. Less than one quarter of these trials addressed primarily surgical research questions such as surgical indications, approaches, techniques, or implant selection. Adequately powered large RCTs of surgical interventions are challenging to conduct, but multiple trials over the last decade have demonstrated their feasibility and potential clinical impact [21,22], and the scoliosis literature already contains many examples of multi-center collaborations [23].

The final priority is to clarify inconsistent outcome reporting and identify practical strategies for methodological improvement. Radiological outcomes are critical to understand deformity correction and technical success, but it remains unclear whether particular radiological outcomes are used consistently in the literature. In addition, it is apparent that radiological outcomes may not always correlate with patient reported function, quality of life, or body image [24-26]. Observational designs dominate the scoliosis literature, but they are frequently prone to confounding bias, selection bias, transfer bias, and recall bias [27]. Further research is necessary to investigate whether methodological safeguards can minimize tendencies towards exaggerated or misleading results [28].

Conclusions

There exists a broad and varied body of research to guide the surgical management of scoliosis. Current research priorities are to (1) undertake high-quality knowledge synthesis and knowledge translation activities; (2) conduct a series of planning meetings to engage clinicians, patients, and methodologists; and (3) clarify outcome reporting and strategies for methodological improvement. Higher-quality studies are specifically necessary to evaluate surgical indications, surgical approaches, surgical techniques, and implant selection. Future studies may also consider engaging global partners to increase generalizability.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NE made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critically revising the manuscript for important intellectual content, and administrative and technical support. TD made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, and critically revising the manuscript for important intellectual content. MG and MB made substantial contributions to conception and design, analysis and interpretation of data, critically revising the manuscript for important intellectual content, and supervision. BD and DP made substantial contributions to analysis and interpretation of data, critically revising the manuscript for important intellectual content, and supervision. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Meha Bhatt and Rajeev Jetly for their assistance retrieving full-texts of articles for review.
Investigation performed at McMaster University, Hamilton, ON, Canada.

Author details

[1]Division of Orthopaedics, Department of Surgery, McMaster University, 293 Wellington St N, Suite 110, Hamilton, ON L8L 8E7, Canada. [2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada.

Received: 5 November 2014 Accepted: 15 December 2014
Published online: 09 January 2015

References

1. Kamerlink JR, Quirno M, Auerbach JD, Milby AH, Windsor L, Dean L, et al. Hospital cost analysis of adolescent idiopathic scoliosis correction surgery in 125 consecutive cases. J Bone Joint Surg Am. 2010;92(5):1097–104.
2. Rushton PR, Grevitt MP. What is the effect of surgery on the quality of life of the adolescent with adolescent idiopathic scoliosis? A review and statistical analysis of the literature. Spine (Phila Pa 1976). 2013;38(9):786–94.
3. Roach JW, Mehlman CT, Sanders JO. Does the outcome of adolescent idiopathic scoliosis surgery justify the rising cost of the procedures? J Pediatr Orthop. 2011;31(1 Suppl):S77–80.
4. Diefenbach C, Ialenti MN, Lonner BS, Kamerlink JR, Verma K, Errico TJ. Hospital cost analysis of neuromuscular scoliosis surgery. Bull Hosp Jt Dis. 2013;71(4):272–7.
5. Cheuk DK, Wong V, Wraige E, Baxter P, Cole A. Surgery for scoliosis in duchenne muscular dystrophy. Cochrane Database Syst Rev. 2013;2:CD005375.
6. Schwab F, Dubey A, Gamez L, El Fegoun AB, Hwang K, Pagala M, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. Spine (Phila Pa 1976). 2005;30(9):1082–5.
7. Silva FE, Lenke LG. Adult degenerative scoliosis: evaluation and management. Neurosurg Focus. 2010;28(3):E1.
8. Straus SE, Tetroe J, Graham I. Defining knowledge translation. CMAJ. 2009;181(3–4):165–8.
9. Franic M, Kujundzic Tiljak M, Pozar M, Romic D, Mimica M, Petrak J, et al. Anterior versus posterior approach in 3D correction of adolescent idiopathic thoracic scoliosis: a meta-analysis. Orthop Traumatol Surg Res. 2012;98(7):795–802.

10. Reddi V, Clarke Jr DV, Arlet V. Anterior thoracoscopic instrumentation in adolescent idiopathic scoliosis: a systematic review. Spine (Phila Pa 1976). 2008;33(18):1986–94.
11. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci. 2010;5:69. 5908-5-69.
12. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with arksey and O'malley's framework. BMC Med Res Methodol. 2013;13:48. 2288-13-48.
13. SRS terminology committee and working group on spinal classification revised glossary of terms [homepage on the Internet]. 2000. Available from: http://www.srs.org/professionals/glossary/SRS_revised_glossary_of_terms.htm.
14. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. J Bone Joint Surg Am. 2003;85-A(1):1–3.
15. Slobogean G, Bhandari M. Introducing levels of evidence to the journal of orthopaedic trauma: implementation and future directions. J Orthop Trauma. 2012;26(3):127–8.
16. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania meta-analysis blinding study group. Lancet. 1997;350(9072):185–6.
17. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257–68.
18. Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than english in systematic reviews. Health Technol Assess. 2003;7(41):1–90.
19. Perkovic V, Patil V, Wei L, Lv J, Petersen M, Patel A. Global randomized trials: the promise of india and china. J Bone Joint Surg Am. 2012;94 Suppl 1:92–6.
20. Yeung M, Bhandari M. Uneven global distribution of randomized trials in hip fracture surgery. Acta Orthop. 2012;83(4):328–33.
21. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. JAMA. 2006;296(20):2441–50.
22. Wardlaw D, Cummings SR, Van Meirhaeghe J, Bastian L, Tillman JB, Ranstam J, et al. Efficacy and safety of balloon kyphoplasty compared with non-surgical care for vertebral compression fracture (FREE): A randomised controlled trial. Lancet. 2009 Mar 21 [cited 20090323];373(9668):1016–24. Available from:.
23. Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of bracing in adolescents with idiopathic scoliosis. N Engl J Med. 2013;369(16):1512–21.
24. Wilson PL, Newton PO, Wenger DR, Haher T, Merola A, Lenke L, et al. A multicenter study analyzing the relationship of a standardized radiographic scoring system of adolescent idiopathic scoliosis and the scoliosis research society outcomes instrument. Spine (Phila Pa 1976). 2002;27(18):2036–40.
25. Sobottke R, Siewe J, Hokema J, Schlegel U, Zweig T, Eysel P. Scoliosis surgery: correction not correlated with instrumentation, quality of life not correlated with correction or instrumentation. Acta Orthop Belg. 2010;76(4):536–42.
26. D'Andrea LP, Betz RR, Lenke LG, Clements DH, Lowe TG, Merola A, et al. Do radiographic parameters correlate with clinical outcomes in adolescent idiopathic scoliosis? Spine (Phila Pa 1976). 2000;25(14):1795–802.
27. Bhandari M, Morshed S, Tornetta 3rd P, Schemitsch EH. Design, conduct, and interpretation of nonrandomized orthopaedic studies: a practical approach. (all) evidence matters. J Bone Joint Surg Am. 2009;91 Suppl 3:1.
28. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7):821–30.

Section III: Systematic Reviews and Meta-Analyses

Chapter 5

Strategies to improve the credibility of meta-analyses in spine surgery: A systematic survey

Reprinted from *The Spine Journal*; 15(9); Nathan Evaniew, Leon van der Watt, Mohit Bhandari, Michelle Ghert, Ilyas Aleem, Brian Drew, and Gordon Guyatt; Strategies to improve the credibility of meta-analyses in spine surgery: a systematic survey; Pages 2066 to 2076;

doi:10.1016/j.spinee.2015.05.018

ELSEVIER

CrossMark

The Spine Journal 15 (2015) 2066–2076

THE SPINE JOURNAL

Review Article

Strategies to improve the credibility of meta-analyses in spine surgery: a systematic survey

Nathan Evaniew, MD*, Leon van der Watt, MBChB(SA), Mohit Bhandari, MD, PhD, FRCSC, Michelle Ghert, MD, FRCSC, Ilyas Aleem, MD, FRCSC, Brian Drew, MD, FRCSC, Gordon Guyatt, MD, MSc, FRCPC

McMaster University, Department of Surgery, Division of Orthopaedics, 293 Wellington St. N, Hamilton ON, Canada, L8L 8E7

Received 23 October 2014; revised 9 March 2015; accepted 13 May 2015

Abstract

BACKGROUND CONTEXT: Meta-analyses are powerful tools that can synthesize existing research, inform clinical practice, and support evidence-based care. These studies have become increasingly popular in the spine surgery literature, but the rigor with which they are being conducted has not yet been evaluated.

PURPOSE: Our primary objectives were to evaluate the methodological quality (credibility) of spine surgery meta-analyses and to propose strategies to improve future research. Our secondary objectives were to evaluate completeness of reporting and identify factors associated with higher credibility and completeness of reporting.

STUDY DESIGN: This study is based on a systematic survey of meta-analyses.

OUTCOME MEASURES: We evaluated credibility according to the Users' Guide to the Medical Literature and completeness of reporting according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.

METHODS: We systematically searched MEDLINE, EMBASE, and The Cochrane Library, and two reviewers independently assessed eligibility, credibility, and completeness of reporting. We used multivariable linear regression to evaluate potential associations. Interrater agreement was quantified using kappa and intraclass correlation (ICC) coefficients.

RESULTS: We identified 132 eligible meta-analyses of spine surgery interventions. The mean credibility score was 3 of 7 (standard deviation [SD], 1.4; ICC, 0.86), with agreement for each item ranging from 0.54 (moderate) to 0.83 (almost perfect). Clinical questions were judged as sensible in 125 (95%), searches were exhaustive in 102 (77%), and risk of bias assessments were undertaken in 91 (69%). Seven (5%) meta-analyses addressed possible explanations for heterogeneity using a priori subgroup hypotheses and 24 (18%) presented results that were immediately clinically applicable. Investigators undertook duplicate assessments of eligibility, risk of bias, and data extraction in 46 (35%) and rated overall confidence in the evidence in 24 (18%). Later publication year, increasing Journal Impact Factor, increasing number of databases, inclusion of Randomized Controlled Trials, and inclusion of non-English studies were significantly associated with higher credibility scores ($p<.05$). The mean score for reporting was 18 of 27 (SD, 4.4; ICC, 0.94).

CONCLUSIONS: The credibility of many current spine surgery meta-analyses is limited. Researchers can improve future meta-analyses by performing exhaustive literature searches, addressing

FDA device/drug status: Not applicable.

Author disclosures: ***NE:*** Nothing to disclose. ***LvdW:*** Nothing to disclose. ***MB:*** Consultancy: Smith & Nephew (C), Stryker (C), Amgen (C), Zimmer (C), Moximed (C), Bioventus (C), Merck, Eli Lilly, Sanofi (F); Grants/grants pending: Smith & Nephew (E, paid directly to institution), DePuy, Eli Lily, Bioventus (C), Stryker (F, Paid directly to institution), Zimmer, Amgen (F, Paid directly to institution). ***MG:*** Grant: Canadian Institutes of Health Research (G, Research grant), Canadian Cancer Society (F, Research grant); Personal Fees: Wright Medical (C, Consultant ad hoc). ***IA:*** Nothing to disclose. ***BD:*** Nothing to disclose. ***GG:*** Nothing to disclose.

The disclosure key can be found on the Table of Contents and at www.TheSpineJournalOnline.com.

The authors certify that they, or a member of their immediate families, have no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, and so forth) that might pose a conflict of interest in connection with the submitted article. This study was not funded. Dr Bhandari is funded in part by a Canada Research Chair. This study did not require local institutional ethics board approval.

* Corresponding author. McMaster University, Department of Surgery, Division of Orthopaedics, 293 Wellington St. N, Hamilton ON, Canada, L8L 8E7. Tel.: (905) 527-4322 fax: (905) 523-8781.

E-mail address: nathan.evaniew@medportal.ca (N. Evaniew)

http://dx.doi.org/10.1016/j.spinee.2015.05.018

possible explanations of heterogeneity, presenting results in a clinically useful manner, reproducibly selecting and assessing primary studies, addressing confidence in the pooled effect estimates, and adhering to guidelines for complete reporting.

Keywords: Meta-analysis; Systematic review; Spine surgery; Reporting; Credibility; Research methodology

Introduction

Meta-analyses are systematic reviews in which statistical tests combine the results from similar studies to produce best estimates of the underlying treatment effects [1]. Pooling multiple studies increases sample size, increasing both the accuracy and precision of the results, and provides guidance for clinical care. Meta-analyses also provide unique opportunities to evaluate differences between studies and detect publication bias and to direct future investigations by identifying knowledge gaps.

Publication rates of spine surgery meta-analyses have increased by approximately fivefold over the last 15 years, but the methodological quality, also known methodological credibility, with which they are being conducted has not been evaluated [2,3]. Along with transparent reporting, high methodological credibility is necessary to avoid misleading conclusions. Flawed meta-analyses are at risk of compromising clinical decision making, and limitations have been documented in a variety of other surgical and medical specialties [3–7].

Clinicians, researchers, and other evidence users may evaluate the credibility of a meta-analysis by applying the Users' Guide to the Medical Literature [1,8]. According to the Users' Guide, the credibility of a meta-analysis depends on the extent to which it addresses a sensible clinical question, includes an exhaustive literature search, addresses possible explanations of between-studies differences, presents results in a clinically useful manner, reproducibly selects and assesses primary studies, and addresses confidence in the pooled effect estimates. Credibility is conceptually distinct from completeness of reporting, which describes the extent to which authors comprehensively report the items necessary for users to critically appraise strengths and weaknesses. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist is a 27-item guide that most major surgical and medical journals have adopted by for this purpose [9].

Our primary objectives were to evaluate the credibility of spine surgery meta-analyses and to propose strategies to improve future research. Our secondary objectives were to evaluate completeness of reporting and identify factors associated with higher credibility and completeness of reporting.

Methods

Eligibility criteria

We performed a systematic survey of all meta-analyses of therapeutic interventions related to spine surgery published since 1990. We defined meta-analyses as any summaries of research that attempted to address a focused clinical question in a systematic and reproducible manner and included a quantitative synthesis to yield a best estimate of treatment effect [1,8]. We excluded systematic reviews that summarized the available literature but did not include quantitative syntheses of results across studies. We included meta-analyses that examined preoperative, intraoperative, or postoperative interventions specific to patients undergoing spine surgery and excluded meta-analyses that addressed relevant interventions in general orthopedic surgery or neurosurgery patients. We also excluded meta-analyses of nontherapeutic research questions such as diagnostic and prognostic studies.

When meta-analyses were published in duplicate in *The Cochrane Database of Systematic Reviews* and another journal, we evaluated the Cochrane versions [10]. When meta-analyses were published as updated versions of earlier meta-analyses, we evaluated the most recent versions. We excluded network meta-analyses, protocols of meta-analyses that did not include results, and meta-analyses in languages other than English [11,12].

Identification of studies

We systematically searched MEDLINE (1990 to present), EMBASE (1990 to present), and The Cochrane Library (no date limit) for articles published up to and including June 6, 2014. We used MeSH and EMTREE headings and subheadings in various combinations, supplemented with free text and limited only to humans and English (Appendix 1). We hand searched the reference lists of the included meta-analyses, consulted with experts, used an online database for evidence-based orthopedics [13], and used the "related articles" feature in PubMed to search for additional articles.

Assessments of eligibility and credibility

Two reviewers independently screened all titles and abstracts for eligibility using a piloted electronic screening form. All discrepancies were resolved by consensus through a process that required reviewers to discuss the rationale for their decisions.

We assessed credibility according to the Users' Guide to the Medical Literature for systematic reviews and meta-analyses [1,8]. The two reviewers independently evaluated seven issues:

(1) Did the meta-analysis explicitly address a sensible question?

Therapeutic clinical questions should have a clear focus defined by specific elements. Research questions were considered sensible if across the range of Population, Intervention, Comparison, and Outcomes consistent treatment effects were plausible [14].

(2) Was the search for relevant studies exhaustive?

To be comprehensive, meta-analyses must seek to identify all potentially eligible articles. Searches were considered exhaustive if they included MEDLINE, at least one other electronic database, and at least one other resource (ie, a third electronic database, hand-searching reference lists, textbook bibliographies, specialized registries, and so forth) [8,15,16].

(3) Was the risk of bias of the primary studies assessed?

Risk of bias assessments could have been study- or outcome-specific and were considered adequate if they used any formal instrument, such as the Cochrane Risk of Bias tool for Randomized Controlled Trials (RCTs) or the Newcastle-Ottawa Scale for observational studies [17,18].

(4) Did the meta-analysis address possible explanations of between-study differences (statistical heterogeneity) in the results?

When multiple studies are combined in a meta-analysis, there are always some variabilities in the results across studies. Some of these variabilities could be because of chance, but it can also reflect important differences between the studies that limit confidence in the pooled results. Statistical heterogeneity quantifies the amount of variability that cannot be explained by chance, and explanations of potential statistical heterogeneity require subgroup testing according to hypotheses that, ideally, were formulated a priori [19–21].

(5) Did the review present results that were ready for clinical application?

Pooled relative measures of association and pooled continuous outcomes are often challenging to interpret, and, in some instances, may be misleading [1]. Dichotomous outcomes were considered ready for clinical application if their presentation included any absolute measure of effect, such as absolute risk reduction or number need to treat [22]. Continuous outcomes were considered ready for clinical application if they referenced any patient-important effect size, such as a minimal important difference [22,23].

(6) Were selection and assessments of studies reproducible?

The systematic review process involves judgments that may suffer from random errors. Selection and assessments of studies were considered reproducible if assessments of eligibility and risk of bias, and data extraction was all performed in duplicate [7,24].

(7) Did the review address confidence in the effect estimates?

Confidence ratings are important because they inform evidence users about the quality of the evidence being used to facilitate clinical decision making. Confidence in effect estimates could have been addressed using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach or any other similar system [25]. All discrepancies were resolved by consensus.

Assessments of reporting and other data collection

The two reviewers independently evaluated each of the included meta-analyses according to the PRISMA checklist [9]. The two reviewers also extracted the following additional data using piloted forms: authors and affiliations, year of publication, journal names, funding, clinical characteristics (spinal level, spinal disorder, and intervention under study), and methodological characteristics (types of eligible studies, included studies, sample size, summary statistics, direction of results, tests of heterogeneity, tools used to assess risks of bias, and reproducibility). We also recorded the 2013 Thomson Reuters Journal Impact Factor and the most recent Science Citation Index for each meta-analysis. The Science Citation Index is a metric of citation frequency that reflects the cumulative number of citations to source items indexed within the Web of Science Core Collection [26].

Statistical analyses

We report discrete variables as counts or proportions, normally distributed continuous variables as means with standard deviations (SDs), and skewed continuous variables as medians with interquartile ranges (IQRs). We quantified interobserver agreement for the reviewers' assessments of study eligibility, each item in the Users' Guide, and each item in the PRISMA checklist using Cohen's kappa coefficient, and we interpreted the kappa values according to Landis and Koch as follows: 0 (poor), 0.01–0.20 (slight), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and 0.81–1.00 (almost perfect) [27]. We quantified interobserver agreement for the reviewers' overall assessments using the raw totals of satisfactory items for each meta-analysis according to the intraclass correlation coefficient (ICC).

We used multivariable linear regression to identify characteristics associated with each of higher credibility and completeness of reporting by treating the raw totals of satisfactory items for each meta-analysis as continuous dependent variables. We hypothesized a priori that each

of publication year, Journal Impact Factor, Science Citation Index, the inclusion of randomized controlled trials, the inclusion of unpublished studies, the inclusion of studies in languages other than English, the searching of multiple databases, the presence of an author with a degree or affiliation in epidemiology or biostatistics would be positively associated with credibility and completeness of reporting [3,5,28]. Given earlier research that demonstrated industry-funded studies are more likely to describe positive findings and more likely to report subgroup analyses that have lesser credibility, we also included an a priori hypothesis that industry funding could have a negative association [29–32].

We evaluated direct correlations with the Pearson correlation coefficient. All tests of significance were two-tailed, and p values of less than .05 were considered significant. All analyses were performed using Microsoft Excel (Santa Rosa, CA, USA, 2011) and IBM SPSS, version 21 (Chicago IL, USA, 2012).

Results

Our results are summarized in Tables 1–3 and Figs. 1–3. Further results are also available in Appendices 2–4.

Study selection

Our search strategy identified 2,166 potentially relevant articles. Screening of titles and abstracts and review of full texts led to the final inclusion of 132 spine surgery meta-analyses (Fig. 1). Of the 2,034 excluded articles, 370 were systematic reviews that did not involve quantitative syntheses and 21 were meta-analyses that did not address therapeutic clinical questions. There were six duplicate publications of eligible meta-analyses in both the *Cochrane Database of Systematic Reviews* and another journals. Agreement between the two reviewers for eligibility was substantial (kappa=0.70).

Characteristics of studies

The number of spine surgery meta-analyses published per year increased over time from 0 in 1990 to 35 in 2013 (Fig. 2). Meta-analyses were most frequently performed in China (52), the United States (37), The Netherlands (11), the United Kingdom (9), and Canada (5) (Fig. 3) and were most frequently published in *Spine* (22), *European Spine Journal* (21), *The Cochrane Database of Systematic Reviews* (13), *Journal of Neurosurgery: Spine* (9), and *PLoS One* (7). The median journal Impact Factor across all the included meta-analyses was 2.5 (IQR, 2.4–3.5), and the median Science Citation Index was 5.5 (IQR, 1.0–18.8). The cumulative frequency of meta-analysis publication in subspecialty journals for spine surgery compared with nonsubspecialty journals over time was similar (Appendix 2).

Table 1
Characteristics of spine surgery meta-analyses

Characteristics	Number of meta-analyses (%), n=132
Spinal levels included	
Cervical	55 (42)
Thoracic	47 (36)
Lumber	95 (72)
Spinal disorders included	
Degenerative	97 (74)
Trauma	40 (30)
Deformity	19 (14)
Tumor/metastases	9 (7)
Infection	5 (4)
Other	7 (5)
Evidence included	
Only RCTs	60 (45)
Only observational studies	30 (23)
RCTs and observational studies	43 (32)
Unpublished studies eligible	34 (26)
Non-English studies eligible	61 (46)
Number of databases searched	
0	1 (1)
1	23 (17)
2	18 (14)
3	42 (32)
4	13 (10)
≥5	35 (27)
Specific databases searched	
MEDLINE	129 (98)
EMBASE	88 (67)
Cochrane	90 (68)
Other	59 (45)
Funding	
Government	19 (14)
Institutional	8 (6)
Industry	6 (5)
None	46 (35)
Other	2 (2)
Not reported	51 (39)

RCT, randomized controlled trial.

The median number of studies included in each meta-analysis was 10 (IQR, 6–19), and the median number of participants was 886 (IQR, 477–1,542). Sixty (45%) meta-analyses included only RCTs, 30 (23%) included only

Table 2
Clinical topics with multiple separate meta-analyses

Clinical topic	Number of meta-analyses (%), n=132
Cervical disc arthroplasty	16 (12)
Vertebral augmentation	15 (11)
Pedicle screws	14 (11)
Lumbar fusion	11 (8)
Lumbar discectomy	8 (6)
BMPs	7 (5)
Antifibrinolytics	6 (5)
Cervical myelopathy	5 (4)
Burst fractures	5 (4)
Spine metastases	4 (3)
Other topics (≤4 each)	45 (34)

BMP, bone morphogenic protein.

Table 3
Methodological credibility of spine surgery meta-analyses

Users' Guide for credibility of the systematic review and meta-analysis process	Number (%) of satisfactory meta-analyses	Interobserver agreement (kappa)
Did the review explicitly address a sensible clinical question?	125 (95)	0.83
Was the search for relevant studies exhaustive?	102 (77)	0.66
Was the risk of bias of the primary studies assessed?	91 (69)	0.67
Did the review address possible explanations of between-study differences in results?	7 (5)	0.56
Did the review present results that are ready for clinical application?	24 (18)	0.68
Were selection and assessments of studies reproducible?	46 (35)	0.54
Did the review address confidence in effect estimates?	24 (18)	0.80

observational studies, and 42 (32%) included both RCTs and observational studies (Table 1). Ninety meta-analyses (68%) included searches in three or more electronic databases, and the most commonly searched databases were MEDLINE (98%), The Cochrane Central Register of Controlled Trials (68%), and EMBASE (67%). Authors included unpublished studies in 34 (26%) and non-English studies in 61 (46%). Authors conducted additional search strategies, such as manually searching reference lists and bibliographies or searching the databases of specialized registries, in 101 (77%). There were several clinical topics that were addressed in more than one published meta-analysis, including cervical disc arthroplasty (16) [33–48], vertebral augmentation (15) [49–63], and pedicle screws (14) [64–77] (Table 2).

Credibility

The mean number of satisfactory Users' Guide items in each spine surgery meta-analysis was three of seven (SD,

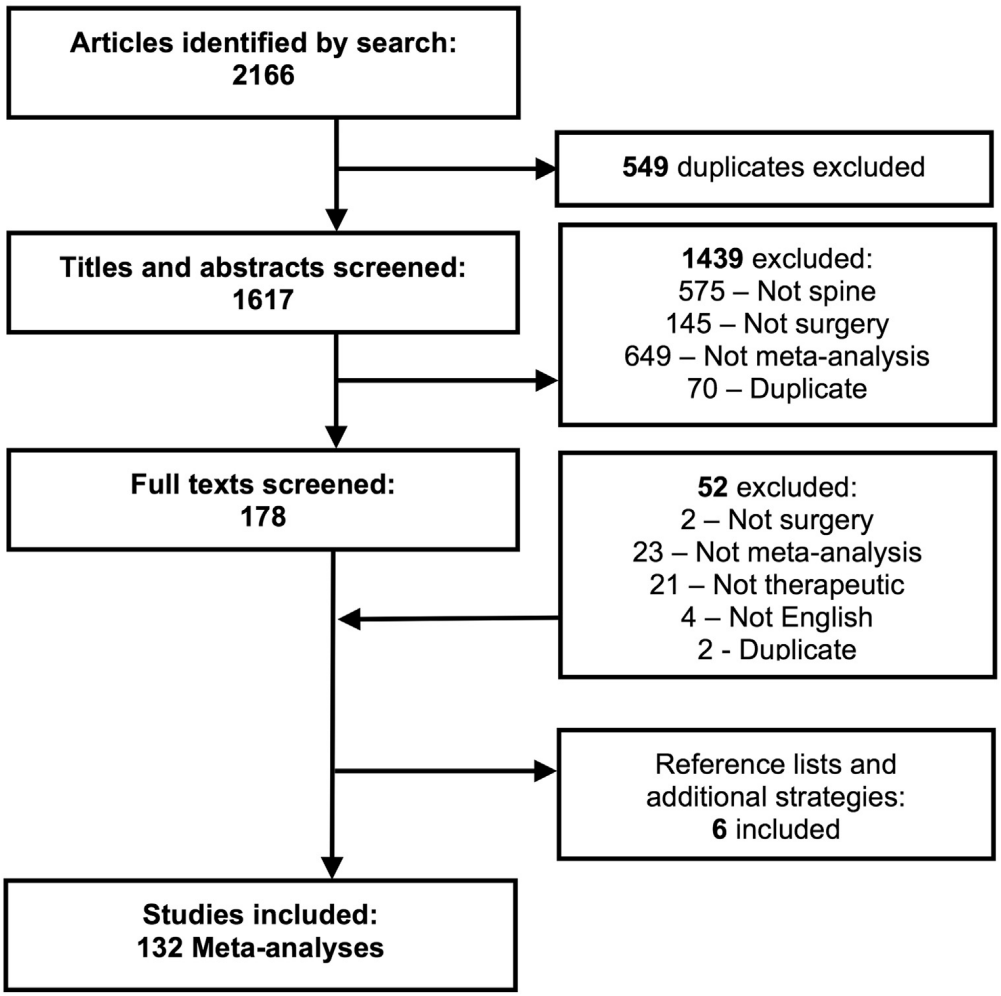


Fig. 1. Flow of articles through systematic survey and reasons for exclusion.

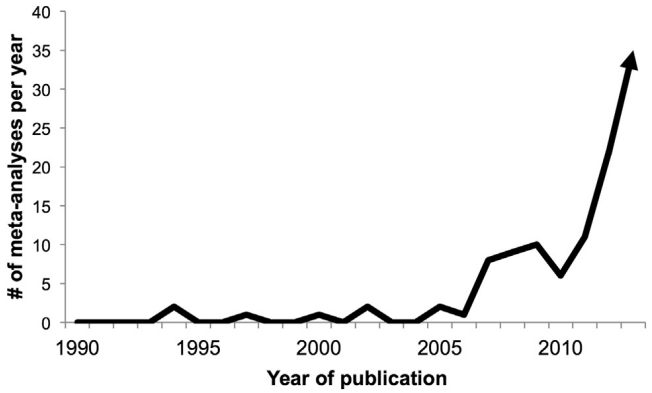


Fig. 2. Number of spine surgery meta-analyses published per year since 1990.

1.4; ICC, 0.86) (Table 3). Almost all authors (125%–95%) addressed sensible clinical questions. Authors conducted comprehensive searches in 102 (77%) and satisfactory risk of bias assessments in 91 (69%), with the most commonly used instrument being the Cochrane Risk of Bias Tool (52, 39%). Seven (5%) meta-analyses addressed possible explanations of heterogeneity using a priori subgroup hypotheses.

Of 24 meta-analyses (18%) that presented results ready for clinical application, minimal important differences were reported in 12, number need to treats in 6, and absolute risk reduction in 13. Duplicate assessments of eligibility were performed in 88 meta-analyses (67%), duplicate assessments of risk of bias in 78 (59%), and duplicate extraction of data in 63 (47%); only 46 (35%) performed all three steps in duplicate. Confidence in the effect estimates was addressed in 24 of the meta-analyses (18%) and 21 (16%) used the GRADE approach. The kappa coefficient for interobserver agreement for individual credibility items ranged from 0.54 (moderate) to 0.83 (almost perfect).

In our multivariate linear regression model (Appendix 3, adjusted R^2=0.58), each of the following were significantly associated with higher credibility ($p<.05$): increasing publication year, increasing Journal Impact Factor, increasing number of databases, inclusion of RCTs, and inclusion of non-English studies. Science Citation Index had a significant negative association with study credibility based on the User's Guide items ($p<.05$) but not when adjusted for Journal Impact Factor ($p=.08$).

Completeness of reporting

The mean number of satisfactory PRISMA items in each spine surgery meta-analysis was 18 of 27 (SD, 4.4; ICC, 0.94) (Appendix 4). Under "Methods" section, 122 (92%) meta-analyses adequately reported eligibility criteria, 117 (87%) reported information sources, and 97 (74%) reported study selection. Eighty meta-analyses (61%) failed to adequately report their search terms or refer to them in an Appendix, 91 (69%) failed to adequately report how risk of bias across studies was assessed, and 127 (96%) failed to report whether risk of bias assessments was study or outcome specific. Under "Results" section, 50 (38%) failed to adequately report risk of bias within individual studies, 24 (18%) failed to adequately report the individual results of each included study, and 100 (76%) failed to adequately report on risk of bias across studies. Under "Discussion" section, study limitations were not adequately reported in 28 (21%), and funding was not adequately reported in 51 (39%). The kappa coefficient for interobserver agreement for individual completeness of reporting items ranged from 0.19 (slight) to 0.91 (almost perfect).

In our multivariable linear regression model (Appendix 3, adjusted R^2=0.50), each of increasing publication year, increasing Journal Impact Factor, increasing number of databases, inclusion of RCTs, and inclusion of non-English studies were significantly associated with a higher

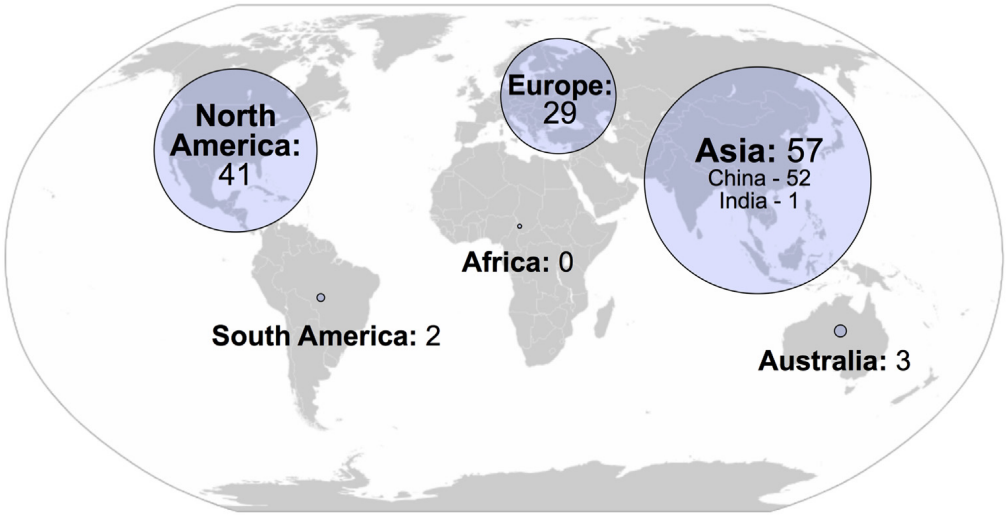


Fig. 3. Global distribution of spine surgery meta-analyses. Numbers indicate absolute counts of meta-analyses published from each continent and are not adjusted for population or researcher density.

completeness of reporting, and industry funding had a significantly negative association ($p<.05$). The correlation between overall credibility and completeness of reporting across all spine surgery meta-analyses was 0.79 ($p<.01$).

Discussion

In 132 eligible meta-analyses of spine surgery interventions, we found that the mean number of satisfactory items for credibility was 3 of 7 (SD, 1.4; ICC, 0.86) and the mean number of satisfactory items for completeness of reporting was 18 of 27 (SD, 4.4; ICC, 0.94). Each of more recent publication, increasing Journal Impact Factor, increasing number of databases, inclusion of RCTs, and inclusion of non-English studies were associated with higher credibility and completeness of reporting.

Strengths and limitations

Credibility is conceptually distinct from completeness of reporting, but complete reporting is necessary to optimally appraise credibility. Van Oldenrijk et al. [78] showed that spine surgery RCTs frequently fail to report essential methodological safeguards such as allocation concealment, blinding, and adherence to the intention-to-treat principle, but Devereaux et al. [79] showed in an observational study of 105 RCTs that safeguards against bias in RCTs were often adequate despite incomplete reporting. Chan et al. [80] surveyed the authors of 43 orthopedic surgery RCTs and identified that 28% to 40% of the trials had blinding of relevant groups despite unclear reporting.

Thus, judgments regarding credibility in the presence of incomplete reporting could have been flawed, but many of the PRISMA items required for the assessments of the User's Guide items were adequately reported. Reporting completeness was greater than 80% for each of objectives, information sources, summary measures, and the results of the risk of bias assessments. However, it was lower for the items required to judge reproducibility, and PRISMA does not evaluate whether authors reported a priori hypotheses to explain heterogeneity or whether they reported confidence in the overall estimates of effect. Although authors may have tested subgroup hypotheses and omitted mention of negative results from the publication, it seems unlikely that they would have omitted positive subgroup findings or that they would have performed but not reported confidence ratings.

The Users' Guide items and the PRISMA checklist were not designed as scoring instruments and neither has been validated for this purpose. We treated tallies of satisfactory items for each instrument as continuous dependent variables, but this approach has not been formally tested. Moderate to almost perfect interrater agreement for all seven Users' Guide items and high interrater overall agreement support the merit of the chosen approach. The strong correlation between the Users' Guide and PRISMA scores suggests that credibility and completeness of reporting may not be distinct and that PRISMA might actually function as a credibility instrument for differentiating between more and less credible reviews. On the other hand, from a normative perspective, the Users' Guide overall scores highlight limitations, whereas the PRISMA scores suggest much better performance.

The exclusion of meta-analyses in languages other than English raises the possibility of language bias. Moher et al. [12] evaluated 130 systematic reviews and concluded that language restrictions did not appear to bias estimates of effect, but the generalizability of these findings to novel surgical interventions and their applicability to credibility are unknown. We selected the interval from 1990 to present because it corresponded with pioneering work in the use and evaluation of meta-analyses and overlapped with earlier research evaluating meta-analyses in orthopedic surgery [3,4,81,82]. That there were no meta-analyses published from 1990 to 1993 supports our choice.

The cumulative frequency of meta-analysis publication in subspecialty spine surgery journals compared with non-subspecialty spine surgery journals over time was similar, which suggests that other factors might be more important to explain the rise in publication rate. Potential factors could include increasing emphasis on evidence-based medicine, increasing awareness of meta-analysis methodology, increasing collaboration with biostatisticians, and increasing availability of dedicated statistical software packages for meta-analyses, but we did not investigate these factors in our study.

Relation to previous work

Investigators have previously identified limitations in the conduct and reporting of meta-analyses in other surgical subspecialties [3–7]. In neurosurgery, Klimo et al. [83] used the Assessment of Multiple Systematic Reviews instrument and PRISMA and identified major deficiencies in search strategies, study selection, data extraction, and assessments of heterogeneity, publication bias, and study quality in 72 meta-analyses. Assessment of Multiple Systematic Reviews has been widely used in the previous literature, but was not used in this study because it does not produce reliably quantifiable assessments, it does not account for modern innovations in meta-analysis methodology, and it does not distinguish credibility from completeness of reporting [1,84,85].

In orthopedic surgery, Dijkman et al. [3] compared 45 meta-analyses from 2005 to 44 meta-analyses from 2008 using the index of Oxman and Guyatt and found that up to 30% had major or extensive flaws. Similar findings were reported by Sharma et al. [5] in their evaluation of 77 total joint arthroplasty meta-analyses and Kowalczuk et al. [6] in their evaluation of 22 femoroacetabular impingement meta-analyses. Each of these studies suggested that adherence to standardized checklists could potentially assist peer

Table 4
Users' Guide for credibility of the systematic review and meta-analysis process

Address a sensible clinical question
• Therapeutic questions should have a clear focus defined by specific PICO.
• It should be plausible that the intervention will have a similar effect across the selected range of patients, interventions or exposures, and outcomes.
Perform an exhaustive search for relevant studies
• MEDLINE, EMBASE, and the Cochrane Database may not uncover all eligible articles for many questions.
• Additional resources include trial registries, bibliographies, abstract presentations, contact with experts, and specialized registries.
Assess risk of bias of the primary studies
• Confidence in the pooled effect estimates depends on the safeguards against bias inherent in the primary studies.
• There are multiple strategies to assess risk of bias, and different study designs require different approaches.
Address possible explanations of between-study differences in results
• Inconsistent results may be because of varying characteristics of the populations, interventions, comparisons, outcomes, and designs of the included studies.
• A small number of credible hypotheses to explain heterogeneity should be specified a priori and tested in subgroup analyses.
Present results that are ready for clinical application
• Relative measures of association and continuous outcomes and are often challenging to interpret and in some instances may be misleading.
• Dichotomous outcomes should be presented as absolute effects compared with baseline risks, and continuous outcomes should reference patient-important effect sizes in familiar units.
Ensure that the selection and assessments of studies are reproducible
• The systematic review process involves judgments that are at risk of random and systematic errors.
• Two or more authors should independently perform each of eligibility assessments, risk of bias assessments, and data abstraction.
Address confidence in pooled effect estimates
• Confidence in the pooled effect estimates depends on study design, risk of bias, imprecision, inconsistency, indirectness, publication bias, and other factors.
• The GRADE Working Group has developed a system that has been adopted by over 70 major health research organizations.
Report transparently according to the PRISMA checklist
• Quality of reporting is conceptually distinct from methodological credibility. It is the extent to which authors transparently report the items necessary for users to critically appraise strengths and weaknesses.

GRADE, Grades of Recommendation, Assessment, Development, and Evaluation; PICO, populations, interventions, comparisons, and outcomes; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

reviewers and journal editors, but none proposed strategies to improve future studies, none evaluated credibility according to the Users' Guide, and none systematically examined the spine literature.

Implications

The credibility of many current spine surgery meta-analyses is limited, and surgeons should be cautious when interpreting their results and applying them to the care of their patients. The mean credibility score of spine surgery meta-analyses was just 3 of 7 Users' Guide items, which may leave them at risk of producing misleading conclusions. Although confidence ratings can specifically help to inform surgeons about the quality of the evidence being used for clinical decision making, only 18% of meta-analyses addressed confidence in the pooled effect estimates according to GRADE or a similar approach.

Adherence to the PRISMA checklist is important to ensure transparent reporting, but it does not include all items necessary for high credibility. Researchers should also adhere to the Users' Guide for credibility of the systematic review and meta-analysis process (Table 4) and should specifically consider performing sufficiently exhaustive literature searches, addressing possible explanations of heterogeneity, presenting results in a clinically useful manner, reproducibly selecting and assessing primary studies, and addressing confidence in the pooled effect estimates [1,8]. Our results suggest the advisability of ongoing education regarding the process of systematic reviews and meta-analysis.

Further investigation is warranted to understand the potential clinical impact of the observed deficiencies by comparing similar meta-analyses and engaging evidence users. We identified several clinical topics that were addressed in more than one published meta-analyses, including some topics that had more meta-analyses than there are randomized trials [86,87]. Meta-analyses by different authors addressing the same topic frequently include different primary trials [88,89], and these differences can produce discordant results [89,90]. In a review of eight meta-analyses, Ford et al. [7] identified missed eligible trials and errors in dichotomous data extraction that led to a ≥10% or more relative difference in treatment effects in five studies and a change in the statistical significance in four studies.

Conclusions

Meta-analyses are powerful tools that can synthesize existing research, inform clinical practice, and directly support evidence-based care. They can be profoundly impactful when high-quality primary evidence and high-quality methodology align, but the credibility of many current spine surgery meta-analyses is limited. Researchers can improve the credibility of future meta-analyses by performing sufficiently exhaustive literature searches, addressing possible explanations of heterogeneity, presenting results in a

clinically useful manner, reproducibly selecting and assessing primary studies, addressing confidence in the pooled effect estimates, and adhering to guidelines for transparent reporting.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.spinee.2015.05.018.

References

[1] Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA 2014;312:171–9.

[2] Jacobs WC, Kruyt MC, Verbout AJ, Oner FC. Spine surgery research: on and beyond current strategies. Spine J 2012;12:706–13.

[3] Dijkman BG, Abouali JA, Kooistra BW, Conter HJ, Poolman RW, Kulkarni AV, et al. Twenty years of meta-analyses in orthopaedic surgery: has quality kept up with quantity? J Bone Joint Surg Am 2010;92:48–57.

[4] Bhandari M, Morrow F, Kulkarni AV, Tornetta P III. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. J Bone Joint Surg Am 2001;83-A:15–24.

[5] Sharma R, Vannabouathong C, Bains S, Marshall A, MacDonald SJ, Parvizi J, et al. Meta-analyses in joint arthroplasty: a review of quantity, quality, and impact. J Bone Joint Surg Am 2011;93: 2304–9.

[6] Kowalczuk M, Adamich J, Simunovic N, Farrokhyar F, Ayeni OR. Methodological quality of systematic reviews addressing femoroacetabular impingement. Knee Surg Sports Traumatol Arthrosc 2014 Jul 19. [Epub ahead of print].

[7] Ford AC, Guyatt GH, Talley NJ, Moayyedi P. Errors in the conduct of systematic reviews of pharmacological interventions for irritable bowel syndrome. Am J Gastroenterol 2010;105:280–8.

[8] Murad MH, Jaeschke R, Devereaux PJ, Prasad K, Carrasco-Labra A, Agoritsas T, et al. The process of a systematic review and meta-analysis: users' guide to the medical literature. In: Guyatt G, Rennie D, Meade MO, Cook DJ, eds. Users' Guides to the medical literature: a manual for evidence-based clinical practice. 3rd ed. New York, NY: McGraw-Hill Professional, 2014.

[9] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009;62:1006–12.

[10] Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of cochrane reviews with articles published in paper-based journals. JAMA 1998;280:278–80.

[11] Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PLoS One 2014;9:e92508.

[12] Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. Health Technol Assess 2003;7:1–90.

[13] OE—OrthoEvidence [homepage on the Internet]; 2014. Available at: www.myorthoevidence.com. Accessed June 6, 2014.

[14] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395–400.

[15] Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev 2007;MR000001.

[16] Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. JAMA 1993;269:2749–53.

[17] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.

[18] Higgins JPT, Greene S, (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. [updated March 2011]. The Cochrane Collaboration, 2011. Available at: www.cochrane-handbook.org. Accessed June 6, 2014.

[19] Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 2010;340:c117.

[20] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. rating the quality of evidence—inconsistency. J Clin Epidemiol 2011;64:1294–302.

[21] Ghert M, Petrisor B. Subgroup analyses: when should we believe them? J Bone Joint Surg Am 2012;94(Suppl 1):61–4.

[22] Alonso-Coello P, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, Akl EA, Sun X, et al. A methodological survey of the analysis, reporting and interpretation of absolute risk ReductiOn in systematic revieWs (ARROW): a study protocol. Syst Rev 2013;2:113. 4053–2-113.

[23] Johnston BC, Thorlund K, Schunemann HJ, Xie F, Murad MH, Montori VM, et al. Improving the interpretation of quality of life evidence in meta-analyses: the application of minimal important difference units. Health Qual Life Outcomes 2010;8:116. 7525–8-116.

[24] Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. J Clin Epidemiol 2006;59:697–703.

[25] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94.

[26] Science citation index [homepage on the Internet]; 2014. Available at: http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jloptions.cgi?PC=K. Accessed July 20, 2014.

[27] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.

[28] Bala MM, Akl EA, Sun X, Bassler D, Mertz D, Mejza F, et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. J Clin Epidemiol 2013;66:286–95.

[29] Okike K, Kocher MS, Mehlman CT, Bhandari M. Conflict of interest in orthopaedic research. An association between findings and funding in scientific presentations. J Bone Joint Surg Am 2007;89:608–13.

[30] Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. The influence of study characteristics on reporting of subgroup analyses in randomised controlled trials: systematic review. BMJ 2011;342:d1569.

[31] Bhandari M, Busse JW, Jackowski D, Montori VM, Schunemann H, Sprague S, et al. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. CMAJ 2004;170:477–80.

[32] Amiri AR, Kanesalingam K, Cro S, Casey ATH. Level of evidence of clinical spinal research and its correlation with journal impact factor. Spine J 2013;13:1148–53.

[33] Alvin MD, Abbott EE, Lubelski D, Kuhns B, Nowacki AS, Steinmetz MP, et al. Cervical arthroplasty: a critical review of the literature. Spine J 2014;14:2231–45.

[34] Anderson PA, Puschak TJ, Sasso RC. Comparison of short-term SF-36 results between total joint arthroplasty and cervical spine decompression and fusion or arthroplasty. Spine 2009;34:176–83.

[35] Anderson PA, Sasso RC, Hipp J, Norvell DC, Raich A, Hashimoto R. Kinematics of the cervical adjacent segments after disc arthroplasty compared with anterior discectomy and fusion: a systematic review and meta-analysis. Spine 2012;37(22 Suppl): S85–95.

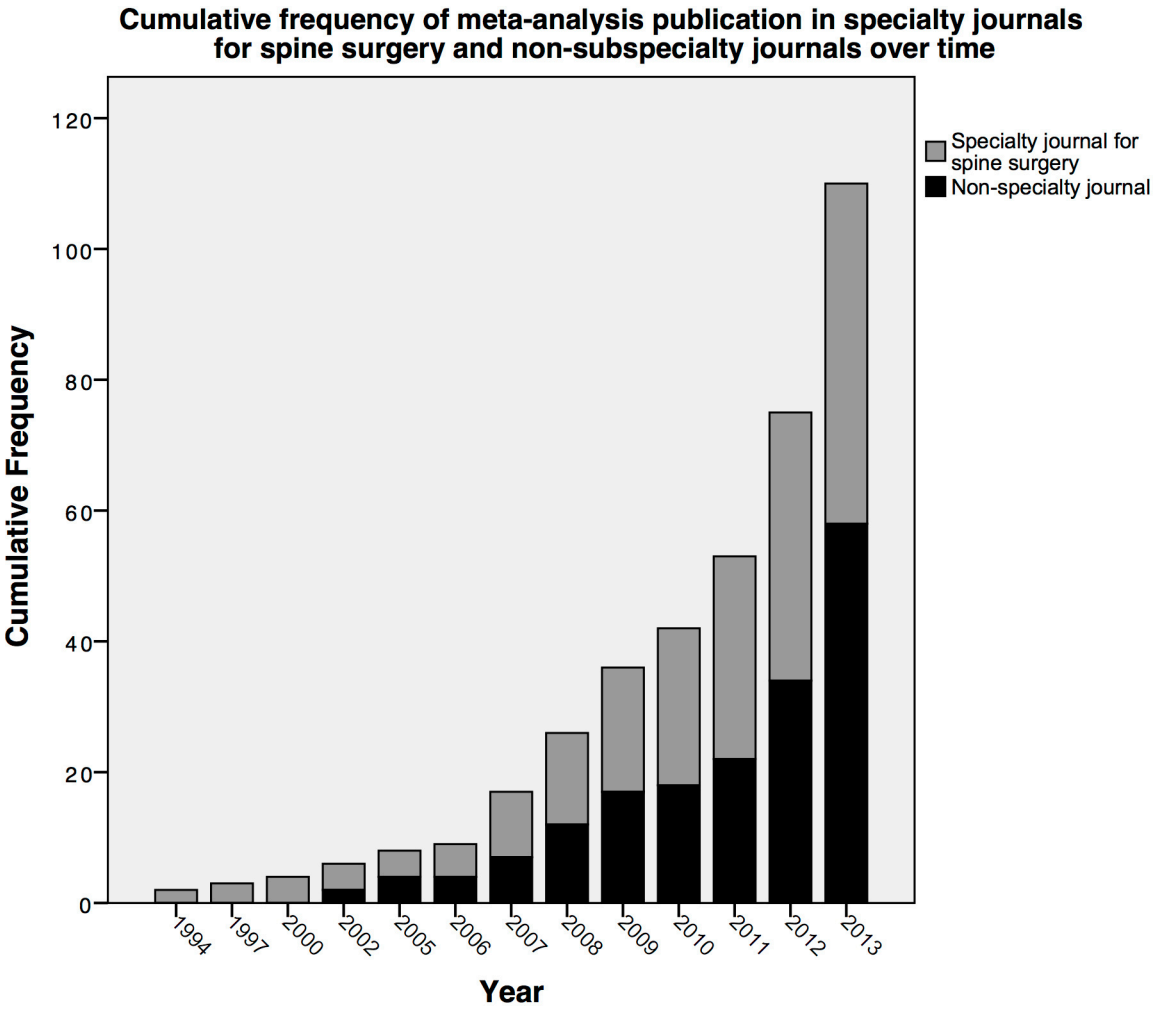
[36] Bartels RH, Donk R, Verbeek AL. No justification for cervical disk prostheses in clinical practice: a meta-analysis of randomized controlled trials. Neurosurgery 2010;66:1153–60; discussion 1160.
[37] Boselie TF, Willems PC, van Mameren H, de Bie R, Benzel EC, van Santbrink H. Arthroplasty versus fusion in single-level cervical degenerative disc disease. Cochrane Database Syst Rev 2012;9: CD009173.
[38] Chen J, Fan SW, Wang XW, Yuan W. Motion analysis of single-level cervical total disc arthroplasty: a meta-analysis. Orthop Surg 2012;4:94–100.
[39] Fallah A, Akl EA, Ebrahim S, Ibrahim GM, Mansouri A, Foote CJ, et al. Anterior cervical discectomy with arthroplasty versus arthrodesis for single-level cervical spondylosis: a systematic review and meta-analysis. PLoS One 2012;7:e43407.
[40] Gao Y, Liu M, Li T, Huang F, Tang T, Xiang Z. A meta-analysis comparing the results of cervical disc arthroplasty with anterior cervical discectomy and fusion (ACDF) for the treatment of symptomatic cervical disc disease. J Bone Joint Surg 2013;95:555–61.
[41] McAfee PC, Reah C, Gilder K, Eisermann L, Cunningham B. A meta-analysis of comparative outcomes following cervical arthroplasty or anterior cervical fusion: results from 4 prospective multicenter randomized clinical trials and up to 1226 patients. Spine 2012;37:943–52.
[42] Ren C, Song Y, Xue Y, Yang X. Mid- to long-term outcomes after cervical disc arthroplasty compared with anterior discectomy and fusion: a systematic review and meta-analysis of randomized controlled trials. Eur Spine J 2014;23:1115–23.
[43] Verma K, Gandhi SD, Maltenfort M, Albert TJ, Hilibrand AS, Vaccaro AR, et al. Rate of adjacent segment disease in cervical disc arthroplasty versus single-level fusion: meta-analysis of prospective studies. Spine 2013;38:2253–7.
[44] Xing D, Ma XL, Ma JX, Wang J, Ma T, Chen Y. A meta-analysis of cervical arthroplasty compared to anterior cervical discectomy and fusion for single-level cervical disc disease. J Clin Neurosci 2013;20:970–8.
[45] Yang B, Li H, Zhang T, He X, Xu S. The incidence of adjacent segment degeneration after cervical disc arthroplasty (CDA): a meta analysis of randomized controlled trials. PLoS One 2012;7:e35032.
[46] Yin S, Yu X, Zhou S, Yin Z, Qiu Y. Is cervical disc arthroplasty superior to fusion for treatment of symptomatic cervical disc disease? A meta-analysis. Clin Orthop Relat Res 2013;471:1904–19.
[47] Yu L, Song Y, Yang X, Lv C. Systematic review and meta-analysis of randomized controlled trials: comparison of total disk replacement with anterior cervical decompression and fusion. Orthopedics 2011;34:e651–8.
[48] Jiang H, Zhu Z, Qiu Y, Qian B, Qiu X, Ji M. Cervical disc arthroplasty versus fusion for single-level symptomatic cervical disc disease: a meta-analysis of randomized controlled trials. Arch Orthop Trauma Surg 2012;132:141–51.
[49] Anderson PA, Froyshteter AB, Tontz WL Jr. Meta-analysis of vertebral augmentation compared with conservative treatment for osteoporotic spinal fractures. J Bone Miner Res 2013;28:372–82.
[50] Bouza C, Lapez-Cuadrado T, Cediel P, Saz-Parkinson Z, Amate JM. Balloon kyphoplasty in malignant spinal fractures: a systematic review and meta-analysis. BMC Palliat Care 2009;8:12.
[51] Bouza C, Lopez T, Magro A, Navalpotro L, Amate JM. Efficacy and safety of balloon kyphoplasty in the treatment of vertebral compression fractures: a systematic review. Eur Spine J 2006;15:1050–67.
[52] Eck JC, Nachtigall D, Humphreys SC, Hodges SD. Comparison of vertebroplasty and balloon kyphoplasty for treatment of vertebral compression fractures: a meta-analysis of the literature. Spine J 2008;8:488–97.
[53] Gill JB, Kuper M, Chin PC, Zhang Y, Schutt R Jr. Comparing pain reduction following kyphoplasty and vertebroplasty for osteoporotic vertebral compression fractures. Pain Physician 2007;10:583–90.
[54] Li LH, Sun TS, Liu Z, Zhang JZ, Zhang Y, Cai YH, et al. Comparison of unipedicular and bipedicular percutaneous kyphoplasty for treating osteoporotic vertebral compression fractures: a meta-analysis. Chin Med J 2013;126:3956–61.
[55] Lin J, Zhang L, Yang H. Unilateral versus bilateral balloon kyphoplasty for osteoporotic vertebral compression fractures. Pain Physician 2013;16:447–53.
[56] Ma XL, Xing D, Ma JX, Xu WG, Wang J, Chen Y. Balloon kyphoplasty versus percutaneous vertebroplasty in treating osteoporotic vertebral compression fracture: grading the evidence through a systematic review and meta-analysis. Eur Spine J 2012;21:1844–59.
[57] Levine SA, Perin LA, Hayes D, Hayes WS. An evidence-based evaluation of percutaneous vertebroplasty. Manag Care 2000;9. 56–60, 63.
[58] Han S, Wan S, Ning L, Tong Y, Zhang J, Fan S. Percutaneous vertebroplasty versus balloon kyphoplasty for treatment of osteoporotic vertebral compression fracture: a meta-analysis of randomised and non-randomised controlled trials. Int Orthop 2011;35:1349–58.
[59] Yang H, Liu T, Zhou J, Meng B, Wang G, Zhu X. Kyphoplasty versus vertebroplasty for painful osteoporotic vertebral compression fractures—which one is better? A systematic review and meta-analysis. Int J Spine Surg 2013;7:e45–57.
[60] Taylor RS, Fritzell P, Taylor RJ. Balloon kyphoplasty in the management of vertebral compression fractures: an updated systematic review and meta-analysis. Eur Spine J 2007;16:1085–100.
[61] Stevenson M, Gomersall T, Lloyd Jones M, Rawdin A, Hernandez M, Dias S, et al. Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for the treatment of osteoporotic vertebral fractures: a systematic review and cost-effectiveness analysis. Health Technol Assess 2014;18:1–290.
[62] Xing D, Ma JX, Ma XL, Wang J, Xu WG, Chen Y, et al. A meta-analysis of balloon kyphoplasty compared to percutaneous vertebroplasty for treating osteoporotic vertebral compression fractures. J Clin Neurosci 2013;20:795–803.
[63] Yang LY, Wang XL, Zhou L, Fu Q. A systematic review and meta-analysis of randomized controlled trials of unilateral versus bilateral kyphoplasty for osteoporotic vertebral compression fractures. Pain Physician 2013;16:277–90.
[64] Cheng LM, Wang JJ, Zeng ZL, Zhu R, Yu Y, Li C, Wu ZR. Pedicle screw fixation for traumatic fractures of the thoracic and lumbar spine. Cochrane Database Syst Rev 2013;5:CD009073.
[65] Dickman CA, Yahiro MA, Lu HTC, Melkerson MN. Surgical treatment alternatives for fixation of unstable fractures of the thoracic and lumbar spine: a meta-analysis. Spine 1994;19(20 SUPPL):2266S–73S.
[66] Ding W, Chen Y, Liu H, Wang J, Zheng Z. Comparison of unilateral versus bilateral pedicle screw fixation in lumbar interbody fusion: a meta-analysis. Eur Spine J 2014;23:395–403.
[67] Han YC, Liu ZQ, Wang SJ, Li LJ, Tan J. Comparison of unilateral versus bilateral pedicle screw fixation in degenerative lumbar diseases: a meta-analysis. Eur Spine J 2014;23:974–84.
[68] Hu XQ, Wu XL, Xu C, Zheng XH, Jin YL, Wu LJ, et al. A systematic review and meta-analysis of unilateral versus bilateral pedicle screw fixation in transforaminal lumbar interbody fusion. PLoS One 2014;9:e87501. Article Number: e8.
[69] Kosmopoulos V, Schizas C. Pedicle screw placement accuracy: a meta-analysis. Spine 2007;32:E111–20.
[70] Ledonio CG, Polly DW Jr, Vitale MG, Wang Q, Richards BS. Pediatric pedicle screws: comparative effectiveness and safety: a systematic literature review from the scoliosis research society and the pediatric orthopaedic society of North America task force. J Bone Joint Surg Am 2011;93:1227–34.
[71] Luo J, Gong M, Gao M, Huang S, Yu T, Zou X. Both unilateral and bilateral pedicle screw fixation are effective for lumbar spinal fusion—A meta-analysis-based systematic review. J Orthop Translation 2014;2:66–74.
[72] Lin X, Xu CP, Yang T, Yin QS, Zhang Y, Xia H. Unilateral versus bilateral fixation for lumbar spinal fusion: a systemic review and meta-analysis. Eur J Orthop Surg Traumatol 2014;24:247–55.
[73] Shin BJ, James AR, Njoku IU, Hartl R. Pedicle screw navigation: a systematic review and meta-analysis of perforation risk for

computer-navigated versus freehand insertion - a review. J Neurosurg Spine 2012;17:113–22.
[74] Tian NF, Xu HZ. Image-guided pedicle screw insertion accuracy: a meta-analysis. Int Orthop 2009;33:895–903.
[75] Tian NF, Huang QS, Zhou P, Zhou Y, Wu RK, Lou Y, et al. Pedicle screw insertion accuracy with different assisted methods: a systematic review and meta-analysis of comparative studies. Eur Spine J 2011;20:846–59.
[76] Verma R, Krishan S, Haendlmayer K, Mohsen A. Functional outcome of computer-assisted spinal pedicle screw placement: a systematic review and meta-analysis of 23 studies including 5,992 pedicle screws. Eur Spine J 2010;19:370–5.
[77] Yuan C, Chen K, Zhang H, Zhang H, He S. Unilateral versus bilateral pedicle screw fixation in lumbar interbody fusion: a meta-analysis of complication and fusion rate. Clin Neurol Neurosurg 2014;117:28–32.
[78] van Oldenrijk J, van Berkel Y, Kerkhoffs GM, Bhandari M, Poolman RW. Do authors report surgical expertise in open spine surgery related randomized controlled trials? A systematic review on quality of reporting. Spine 2013;38:857–64.
[79] Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schunemann HJ, et al. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol 2004;57:1232–6.
[80] Chan S, Bhandari M. The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. J Bone Joint Surg Am 2007;89:1970–8.
[81] Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991;44:1271–8.
[82] Oxman AD, Guyatt GH. Guidelines for reading literature reviews. CMAJ 1988;138:697–703.
[83] Klimo P Jr, Thompson CJ, Ragel BT, Boop FA. Methodology and reporting of meta-analyses in the neurosurgical literature. J Neurosurg 2014;120:796–810.
[84] Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J 2010;4:84–91.
[85] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009;62:1013–20.
[86] Evaniew N, Khan M, Drew B, Peterson D, Bhandari M, Ghert M. Intrawound vancomycin to prevent infections after spine surgery: a systematic review and meta-analysis. Eur Spine J 2015;24:533–42.
[87] Chiang HY, Herwaldt LA, Blevins AE, Cho E, Schweizer ML. Effectiveness of local vancomycin powder to decrease surgical site infections: a meta-analysis. Spine J 2014;14:397–407.
[88] Mascarenhas R, Chalmers PN, Sayegh ET, Bhandari M, Verma NN, Cole BJ, et al. Is double-row rotator cuff repair clinically superior to single-row rotator cuff repair: a systematic review of overlapping meta-analyses. Arthroscopy 2014;30:1156–65.
[89] Campbell J, Bellamy N, Gee T. Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee. Osteoarthritis Cartilage 2007;15:1424–36.
[90] Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. CMAJ 1997;156:1411–6.

Appendix 1 – Electronic search strategy for MEDLINE

1	Spine or spinal).mp. [mp=ti, ab, sh, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui]
2	spine.mp.
3	(surgery or surgical or operation or operate).mp.
4	surgery.mp.
5	exp Spine/su
6	(1 or 2) and (3 or 4)
7	6 or 5
8	exp meta-analysis/
9	meta-analysis.mp.
10	exp systematic review/
11	systematic review.mp.
12	8 or 9 or 10 or 11
13	7 and 12
14	limit 13 to english language
15	limit 14 to yr="1990 -Current"
16	limit 15 to humans

APPENDIX 2


Cumulative frequency of meta-analysis publication in specialty journals for spine surgery and non-subspecialty journals over time
Cumulative Frequency
120
100
80
60
40
20
0
1994
1997
2000
2002
2005
2006
2007
2008
2009
2010
2011
2012
2013
Year
Specialty journal for spine surgery
Non-specialty journal

Appendix 3. Factors associated with methodological credibility and quality of reporting

	Methodological credibility (Users' guide)					Quality of reporting (PRISMA)			
	Adjusted R-square = 0.58					Adjusted R-square = 0.50			
	Standardized Coefficient	P-value	95% CI			Standardized Coefficient	P-value	95% CI	
			lower	*upper*				lower	upper
Year of publication	0.23	**<0.01**	0.03	0.15		0.39	**<0.01**	0.27	0.67
Impact Factor	0.23	**<0.01**	0.06	0.24		0.19	**<0.01**	0.10	0.68
Science Citation Index	-0.18	**0.02**	-0.02	0.00		-0.04	0.67	-0.03	0.02
Number of databases	0.33	**<0.01**	0.13	0.31		0.22	**<0.01**	0.15	0.76
Inclusion of RCTs	0.18	**0.01**	0.12	0.61		0.20	**<0.01**	0.48	2.10
Inclusion of unpublished studies	-0.05	0.50	-0.64	0.31		0.13	0.11	-0.28	2.84
Inclusion of non-English studies	0.31	**0.01**	0.52	1.26		0.18	**0.01**	0.38	2.83
Degree/affiliation in epidemiology or biostatistics	-0.04	0.56	-0.53	0.29		-0.05	0.42	-1.88	0.80
Industry funding	-0.06	0.34	-1.26	0.43		-0.13	**0.04**	-5.64	-0.08

Appendix 4. Completeness of reporting of spine surgery meta-analyses

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist	Number (%) of satisfactory meta-analyses	Inter-observer agreement (kappa)
Title		
Title	88 (67)	0.91
Abstract		
Structured summary	123 (93)	0.56
Introduction		
Rationale	130 (99)	0.66
Objectives	125 (95)	0.83
Methods		
Protocol and registration	18 (14)	0.58
Eligibility criteria	122 (92)	0.40
Information sources	117 (87)	0.60
Search	52 (39)	0.67
Study selection	97 (74)	0.31
Data collection process	82 (62)	0.32
Data items	103 (78)	0.45
Risk of bias in individual studies	5 (4)	0.91
Summary measures	126 (96)	0.27
Synthesis of results	97 (74)	0.45
Risk of bias across studies	41 (31)	0.88
Additional analyses	58 (44)	0.52
Results		
Study selection	108 (82)	0.19
Study characteristics	110 (83)	0.71
Risk of bias within studies	82 (62)	0.68
Results of individual studies	108 (82)	0.69
Synthesis of results	101 (77)	0.69
Risk of bias across studies	32 (24)	0.73
Additional analysis	55 (42)	0.44
Discussion		
Summary of evidence	24 (18)	0.64
Limitations	104 (79)	0.55
Conclusions	130 (99)	0.49
Funding		
Funding	72 (55)	0.74

Chapter 6

Methylprednisolone for the treatment of patients with acute spinal cord injuries: A systematic review and meta-analysis

Reproduced from *Journal of Neurotrauma* Nov. 3 [Epub ahead of print]; Nathan Evaniew, Emilie P Belley-Côté, Nader Fallah, Vanessa K Noonan, Carly S Rivers, Marcel F Dvorak; Methylprednisolone for the Treatment of Patients with Acute Spinal Cord Injuries: A systematic review and meta-analysis; Copyright 2015; Publisher: Mary Ann Liebert, Inc.
doi:10.1089/neu.2015.4192

Methylprednisolone for the treatment of patients with acute spinal cord injuries: A systematic review and meta-analysis

Nathan Evaniew, Emilie P Belley-Côté, Nader Fallah, Vanessa K Noonan, Carly S Rivers, Marcel F Dvorak

Nathan Evaniew (corresponding author)
Department of Surgery, McMaster University
293 Wellington Street North, Suite 110
Hamilton, Ontario L8L 8E7, Canada
Tel: 905 527 4322 x44983
Fax: 905 523 8781
nathan.evaniew@medportal.ca

Emilie P Belley-Côté
Department of Clinical Epidemiology and Biostatistics, McMaster University
237 Barton Street East, Room 1C1
Hamilton, Ontario L8L 2X2, Canada
Tel: 905 527 4322 x44091
Fax: 905 577 8017
emilie.belley-cote@medportal.ca

Nader Fallah
University of British Columbia
Rick Hansen Institute
6400-818 West 10th Avenue, Vancouver, BC V5Z 1M9, Canada
Tel: 604 827 1694
Fax: 604 827 1669
nfallah@rickhanseninstitute.org

Vanessa K Noonan
University of British Columbia
Rick Hansen Institute
6400-818 West 10th Avenue, Vancouver, BC V5Z 1M9, Canada
Tel: 604 827 1694
Fax: 604 827 1669
vnoonan@rickhanseninstitute.org

Carly S Rivers
University of British Columbia
Rick Hansen Institute
6400-818 West 10th Avenue, Vancouver, BC V5Z 1M9, Canada
Tel: 604 827 1694
Fax: 604 827 1669
crivers@rickhanseninstitute.org

Marcel F Dvorak
Department of Orthopaedics, University of British Columbia
Blusson Spinal Cord Centre
818 West 10th Avenue, Vancouver, BC V5Z 1M9, Canada
Tel: 604 875 5859
Fax: 604 875 8223
Marcel.Dvorak@ubc.ca

ABSTRACT

Previous meta-analyses of methylprednisolone for patients with acute traumatic spinal cord injuries (TSCIs) have not addressed confidence in the quality of evidence used for pooled effect estimates and new primary studies have been recently published. We aimed to determine whether methylprednisolone improves motor recovery and is associated with increased risks for adverse events.

We searched MEDLINE, EMBASE, and The Cochrane Library, and two reviewers independently screened articles, extracted data, and evaluated risk of bias. We pooled outcomes from randomized controlled trials (RCTs) and controlled observational studies separately and used the Grades of Recommendation, Assessment, Development, and Evaluation approach to evaluate confidence.

We included four RCTs and 17 observational studies. Methylprednisolone was not associated with an increase in long-term motor score recovery (two RCTs: 335 participants, mean difference [MD] -1.11, 95% CI -4.75 to 2.53, p=0.55, low confidence; two observational studies: 528 participants, MD 1.37, 95% CI -3.08 to 5.83, p=0.55, very low confidence) or improvement by at least one motor grade (three observational studies: 383 participants, risk ratio [RR] 0.84, 95% CI 0.53 to 1.33, p=0.46, very low confidence). Evidence from two RCTs demonstrated superior short-term motor score improvement if methylprednisolone was administered within eight hours of injury (two RCTs: 250 participants; MD 4.46, 95% CI 0.97 to 7.94, p=0.01; low confidence), but risk of bias and imprecision limit confidence in these findings. Observational studies demonstrated a significantly increased risk for gastrointestinal bleeding (nine studies: 2857 participants, RR 2.18; 95% CI 1.13 to 4.19; p=0.02, very low confidence), but RCTs did not.

Pooled evidence does not demonstrate a significant long-term benefit for methylprednisolone in patients with acute TSCIs and suggests it may be associated with increased gastrointestinal bleeding. These findings support current guidelines against routine use, but strong recommendations are not warranted because confidence in the effect estimates is limited.

KEY WORDS

methylprednisolone; spinal cord injury; systematic review; meta-analysis; motor score

INTRODUCTION

Patients with acute traumatic spinal cord injuries (TSCIs) often experience severe loss of function and profoundly impaired quality of life, and the development of interventions to improve motor recovery is critically important.[1,2] More than 500,000 people suffer acute TSCIs worldwide each year, and global prevalence is expected to increase.[3–5]

In the landmark Second National Spinal Cord Injury Study (NASCIS-II), 437 participants with acute TSCIs were randomized to an initial bolus of 30 mg/kg of methylprednisolone followed by an infusion of 5.4 mg/kg per hour for 23 hours versus either naloxone or placebo.[6,7] Although subgroup analyses suggested a small benefit attributable to methylprednisolone for motor recovery, other studies reported conflicting results and utilization has declined sharply in the last decade.[8–12] Potential harms of methylprednisolone include risks for infections and gastrointestinal bleeding, potentially leading to increased mortality.[1,13] Most current guidelines do not recommend routine administration of methylprednisolone for acute TSCIs.[13–15]

Systematic reviews and meta-analyses are powerful tools that can synthesize conflicting literature and inform clinical practice, but they require rigorous methodology and must be frequently updated to avoid

misleading conclusions.[16,17] New primary studies have been recently published and previous meta-analyses evaluating methylprednisolone for patients with TSCIs have not addressed confidence in the quality of the evidence used for the pooled effect estimates.[1,8,18–20] Therefore, we aimed to determine whether methylprednisolone improves motor recovery and is associated with an increase in adverse events in patients with acute TSCIs in comparison to placebo or no treatment.

METHODS

We performed a systematic review and meta-analysis according to the methods of the Cochrane Handbook for Systematic Reviews of Interventions, and we report according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[21,22]

Eligibility criteria

We included all randomized controlled trials (RCTs) and controlled observational studies that compared methylprednisolone against placebo or no treatment in adult patients with acute TSCIs. Studies that reported exclusively on pediatric patients (less than 18 years of age) and studies of corticosteroids other than methylprednisolone were excluded. Studies that combined pediatric patients with adult patients were included. No exclusions were made on the basis of open versus closed injuries,

language, publication status, timing of outcome assessment, setting, or regimen of methylprednisolone.

Identification of studies

We searched MEDLINE (1946 to present), EMBASE (1974 to present), and The Cochrane Library (no date limit) on June 6, 2015 using MeSH and EMTREE headings and subheadings in various combinations, supplemented with free text (**Appendix 1**). We also reviewed reference lists from included studies and previous reviews, consulted with experts, and used the "related articles" feature in PubMed. To identify potential unpublished studies, we searched *clinicaltrials.gov* and reviewed annual conference proceedings from 2012 to present for the North American Spine Society, the Spine Society of Europe, and the Canadian Spine Society.

Two reviewers (NE, EBC) independently screened all titles and abstracts, and then screened the full texts of potentially eligible studies for final inclusion. Studies in languages other than English were translated. All discrepancies were resolved by consensus.

Data extraction

The two reviewers independently evaluated risk of bias for each study using the Cochrane Collaboration's Risk of Bias tool for RCTs and the Methodological Index for Non-Randomized Studies (MINORS) for observational studies, and all discrepancies were resolved by consensus.[21,23]

We classified outcomes by consensus as critical, important but not critical, or of limited importance to patients and decision-makers, and we extracted data for those outcomes considered critical or important.[21] Motor recovery and specific adverse events (mortality, sepsis, pneumonia, gastrointestinal bleeding, decubitus ulcer, urinary tract infection, venous thromboembolism, surgical site infection) were considered critical or important based on clinical significance and previous literature.[24–26] Sensory recovery, length of stay, hyperglycemia, and other outcomes were considered of limited importance.

The two reviewers independently extracted the following data points using piloted electronic data forms: study design, first author, journal, year of publication, patient characteristics, surgical co-intervention and surgical timing, injury severity, sample size and losses in each group, duration of follow-up, methylprednisolone regimen, motor recovery outcomes, and adverse events. Scales for motor recovery outcomes included

International Standards for Neurologic Classification of Spinal Cord Injury (ISNCSCI) American Spinal Injury Association (ASIA) total motor scores (continuous), and improvement by one grade or more on the Frankel or ASIA Impairment Scale (AIS) (dichotomous).[27]

We contacted authors and reviewed data reported in previous meta-analyses for clarifications when needed.[8] We estimated standard deviations (SD) for motor score improvement when necessary by imputing the median SDs for all patients that received the same treatment (steroids versus no steroids) from all studies that reported on the same outcome (motor score improvement).[21] When studies investigated additional interventions, we extracted outcome data only for comparisons of methylprednisolone versus placebo or no treatment.

Data synthesis

We pre-specified that we would not pool data from RCTs with data from observational studies,[21,28] and we pre-specified subgroup hypotheses that we would test to explain potential high heterogeneity: cervical versus thoracolumbar injuries; complete (AIS A) versus incomplete (AIS B/C/D) injuries; presence and timing of surgical co-intervention; and risk of bias.[16,29–31] We also planned subgroup analyses for motor recovery outcomes including only studies in which methylprednisolone was

administered within eight hours of injury regardless of heterogeneity because of established clinical interest.[1,8] We pooled motor recovery data at follow-up durations of six months or earlier (short-term) and greater than six months (long-term), and adverse event data at final follow-up from each study.[1,6,24,30,32] We used the numbers of participants reported as followed-up at each specific time-point in each trial where possible.

We used the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach to evaluate confidence in the pooled effect estimates.[16,33] According to GRADE, data from randomized controlled trials are considered high quality evidence but can be rated down according to risk of bias, imprecision, inconsistency, indirectness, or publication bias. Data from observational studies are considered low quality evidence but can be rated up due to a large treatment effect, evidence of a dose–response relationship, or if all plausible biases would not undermine the conclusions. We rated down for imprecision if the 95% confidence intervals (CIs) failed to exclude benefit or harm and if the pooled sample would have been underpowered to detect the point estimate (Optimal Information Size criterion).[34] We rated down for inconsistency if statistically significant heterogeneity could not be explained by our pre-specified subgroup hypotheses.[35]

Statistical analysis

We quantified inter-observer agreement for the reviewers' assessments using Cohen's kappa and interpreted values according to Landis and Koch as: 0, poor; 0.01 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect.[36] We combined outcome data according to the inverse variance method using a random effects model.[21] We report pooled estimates as mean differences (MD) with 95% CIs for continuous outcomes and risk ratios (RR) with 95% CIs for dichotomous outcomes. We constructed funnel plots to assess for publication bias, we quantified heterogeneity using the chi-squared test and the I^2 statistic. We planned sensitivity analyses to test the importance of estimated data by omitting studies requiring estimation and to test the importance of losses to follow-up across a range of plausible assumptions about the nature of any losses.[37] Tests of significance were two-tailed and p-values <0.05 were considered significant. All analyses were performed using IBM SPSS Version 21 (SPSS Inc., Chicago IL; 2012) and Review Manager 5.3 (The Nordic Cochrane Center, The Cochrane Collaboration, 2014, Copenhagen, Denmark).

RESULTS

Included studies

Our search strategy identified 2062 potential articles, 82 of which we reviewed as full texts (**Figure 1**). Thereafter, we included four RCTs (n = 548 participants)[6,38–40] and 17 controlled observational studies (n = 3967)[16,19,20,41–54] that compared methylprednisolone to placebo or no treatment in adult patients with acute TSCIs. Inter-observer agreement was substantial for screening of titles and abstracts (kappa = 0.75) and almost perfect for review of full texts (kappa = 0.95). We used a previous meta-analysis[8] to clarify motor scores for three studies[6, 38, 39] and adverse events for one study.[38] Contact with authors led to clarification of adverse events and quality assessment for one study.[53]

Of the four RCTs, two compared the NASCIS-II regimen of methylprednisolone against placebo[6,40] and two against no treatment[38,39] (**Table 1**). Two reported on motor score improvement at short-term follow-up and two at long-term follow-up,[7,39] and all four reported on at least one adverse event of interest. Three reported that all participants were treated within eight hours of their injuries and one reported that only 45% were treated within eight hours but had published subgroup data available.

Of the 17 observational studies, 15 compared the NASCIS-II regimen of methylprednisolone against no treatment and two compared “high dose” methylprednisolone against no treatment (**Table 1**). Seven reported on

motor score improvement, of which five did so at short-term follow-up and two at long-term follow-up. Seven reported on improvement by one grade or more on the Frankel or ASIA Impairment Scale, including three at short-term follow-up and two at long-term follow-up. One reported on motor grade improvement at the "time of discharge" and was pooled at short-term follow-up and one did not report follow-up duration for motor recovery and was pooled at long-term follow-up. Fifteen observational studies reported on at least one of the adverse events of interest.

Risk of bias

All four RCTs were at unclear risk of bias for sequence generation, allocation concealment, and addressing incomplete outcome data, and low risk for selective reporting (**Appendix 2**). Two were at low risk and two were at high risk for blinding. None adequately reported loss to follow-up. One study was at high risk because there were a large number of unexplained post-randomization exclusions.

Fifteen of the 17 observational studies were retrospective and five used historical rather than contemporary control groups (**Appendix 2**). None incorporated unbiased assessment of outcomes, 14 did not demonstrate similarity at baseline and only two performed adjusted analyses, 9 did not

specify consecutive patient enrolment, and 11 did not report losses to follow-up.

Our funnel plots to detect publication bias were symmetric, but the small number of studies for each outcome limited interpretability (**Appendix 3**).[21]

Motor recovery

Methylprednisolone was not associated with a significant motor score improvement at short-term follow-up according to evidence from two RCTs (414 participants; MD 1.19, 95% CI -2.33 to 4.71, p=0.51; heterogeneity p=0.23, I^2=30%; low confidence) and five observational studies (308 participants; MD 3.04, 95% CI -2.81 to 8.90, p=0.31; heterogeneity p<0.05, I^2=77%; very low confidence; **Figure 2a**). Methylprednisolone was also not associated with a significant motor score improvement at long-term follow-up according to evidence from two RCTs (335 participants; MD -1.11, 95% CI -4.75 to 2.53, p=0.55; heterogeneity p=0.52, I^2=0%; low confidence) and two observational studies (528 participants; MD 1.37, 95% CI -3.08 to 5.83, p=0.55; heterogeneity p=0.26, I^2=22%; very low confidence; **Figure 2b**). We rated down the quality of the evidence for motor score improvement to low for RCTs due to risk of bias and imprecision, and to very low for observational studies due to study design, risk of bias, and imprecision (**Table 2**). Heterogeneity among the

observational studies with short-term outcomes was resolved by including only those studies in which all patients received surgical co-intervention according to a pre-specified subgroup analysis (167 participants, two studies; MD -0.99, 95% CI -6.02 to 4.04, p=0.70; heterogeneity p=0.85, I^2=0%; very low confidence).

According to evidence from observational studies only, methylprednisolone was not associated with a significant improvement by one grade or more on the Frankel or ASIA Impairment Scale at short-term follow-up (675 participants, four studies; RR 1.27, 95% CI 0.75 to 2.17, p=0.37; heterogeneity p<0.05, I^2=76%; very low confidence; **Figure 3a**) or long-term follow-up (383 participants, three studies; RR 0.84, 95% CI 0.53 to 1.33, p=0.46; heterogeneity p=0.96, I^2=0%; **Figure 3b**). Heterogeneity at short-term follow-up was not explained by our pre-specified subgroup hypotheses, so we rated down the quality of the evidence for inconsistency.

According to evidence from RCTs, methylprednisolone initiated within eight hours of injury was associated with a significant motor score improvement at short-term follow-up (250 participants, two studies; MD 4.46, 95% CI 0.97 to 7.94, p=0.01; heterogeneity p=0.81, I^2=0%; low confidence), but not at long-term follow-up (177 participants, two studies;

MD 1.97, 95% -7.78 to 11.73, p=0.69; heterogeneity p=0.16, I^2=50; low confidence; **Appendix 4**). According to evidence from observational studies, methylprednisolone initiated within eight hours of injury was not associated with a significant motor score improvement at short-term follow-up (275 participants, four studies; MD 4.48; 95% CI -2.49 to 11.45, p=0.21; heterogeneity p<0.05, I^2=81%; very low confidence) or long-term follow-up (224 patients, one study; MD -1.80, 95% CI -8.79 to 5.19, p=0.61; very low confidence), and was not associated with a significant improvement by one grade or more on the Frankel or ASIA Impairment Scale at short-term follow-up (675 participants, four studies; RR 1.27, 95% CI 0.75 to 2.17, p=0.37; heterogeneity p<0.05, I^2=76%; very low confidence) or long-term follow-up (368 participants, two studies; RR 0.86, 95% CI 0.53 to 1.38, p=0.53; heterogeneity p=0.97, I^2=0%; very low confidence). We rated down the quality of the evidence for motor score improvement when methylprednisolone was initiated within eight hours of injury to low for RCTs due to risk of bias and imprecision, and to very low for observational studies due to study design, risk of bias, and imprecision

Adverse events

Methylprednisolone was not associated with significantly increased risks for mortality, sepsis, pneumonia, decubitus ulcer, urinary tract infection, venous thromboembolism, surgical site infection, or total adverse events

according to pooled evidence from RCTs and observational studies (**Table 3** and **Appendix 5**). Evidence from observational studies suggested a significantly increased risk for gastrointestinal bleeding (2857 participants, nine studies; RR 2.18, 95% CI 1.13 to 4.19, p=0.02; heterogeneity p=0.17, I^2=33%; very low confidence), but evidence from RCTs did not (444 participants, three studies; RR 1.99, 95% CI 0.74 to 2.13, p=0.40; heterogeneity p=0.54, I^2=0%). The quality of the evidence for all adverse events was rated down to low for RCTs due to risk of bias and imprecision, and to very low for observational studies due to study design and risk of bias (**Table 2**).

Sensitivity analyses

Our results were robust in sensitivity analyses to test the importance of estimated data (**Appendix 6**). Sensitivity analyses to test the importance of loss to follow-up were not performed because they were adequately reported only in five observational studies (**Appendix 2**).

DISCUSSION

According to evidence from RCTs and controlled observational studies, methylprednisolone did not significantly improve long-term motor score recovery or recovery by at least one motor grade in patients with acute TSCIs in comparison to placebo or no treatment. Limited data suggested

short-term motor score improvements if methylprednisolone was administered within eight hours of injury, but there was no significant benefit after more than six months. Evidence from observational studies suggested an association between methylprednisolone and an increased rate of gastrointestinal bleeding, but evidence from RCTs did not. The risks for other adverse events were not significantly different between groups. The quality of the evidence for all outcomes was low or very low, which means that confidence in the effect estimates is limited and the true effects may be substantially different.

Limitations

We included evidence from RCTs and evidence from controlled observational studies, and this approach risked trading off imprecise but unbiased estimates for precise but biased estimates.[21] Statistical heterogeneity was minimal for most outcomes, but it is plausible that unreported differences in treatment decisions, administration of co-interventions, timing of baseline neurological examinations, or methods of outcomes assessment could have introduced important variability.[25,55,56] For example, patients given methylprednisolone within eight hours of their injuries would be most likely to have had their baseline neurological exams performed within eight hours of injury whereas those who did not receive methylprednisolone may have been treated outside that window and had

their baseline examinations delayed. Differential timing of baseline examinations could therefore bias motor recovery outcomes in favor of methylprednisolone by creating a greater opportunity to recovery.[1]

In order to avoid unacceptable error and misleading conclusions, we presented the effect estimates from RCTs and observational studies separately and we rated down confidence for each study design when risk of bias was unclear or high.[21,33] Given that poorly designed or poorly executed RCTs can sometimes be more problematic than well-designed observational studies, our inclusion of both study designs provides readers a broader view of the literature.[28,57]

Pooled effect estimates should ideally be interpreted in light of patient-important effect sizes to facilitate clinical application, but it is unknown what magnitude of motor score improvement represents a minimal important difference (MID).[16,55,58] A current multi-center trial of riluzole in TSCI has been powered to detect a nine-point difference[59] and some have considered MIDs of up to 20 points.[24] However, others have argued that even very small amounts of motor improvement may be meaningful or that importance may vary depending on the anatomical level and baseline severity of patients' injuries or other contextual factors.[29,55,60] We considered that even small differences could be important to patients, so

we conservatively rated down confidence when the 95% confidence intervals (CIs) failed to exclude any amount of benefit or harm, rather than rating down only if they failed to exclude certain thresholds of MID.[34]

A multivariate analysis of 411 participants from The Surgical Timing in Acute Spinal Cord Injury Study (STASCIS) suggested potential confounding between methylprednisolone and timing of surgical decompression.[61,62] We excluded this study because controlled outcome data were not reported. Although we proposed subgroup analyses based on surgical co-intervention to explain potential heterogeneity, surgical timing was infrequently reported and conventional meta-analyses are poorly equipped to statistically adjust for potential confounders. Access to participant-level data from the included studies could facilitate meta-regression for this purpose.[63]

Relation to previous literature

Botelho *et al.* performed a systematic review and concluded that serious potential harms of methylprednisolone in patients with acute TSCIs outweighed small potential benefits,[18] but Bracken performed a meta-analysis and concluded that methylprednisolone improves motor recovery if started within eight hours.[8] In comparison to Bracken's findings, our point estimate for motor score improvement among all patients was

smaller at short-term follow-up and in the direction of harm rather than benefit at long-term follow-up; our confidence intervals indicate similar imprecision. Neither of these reviews ensured that the selection of studies was reproducible by using two or more reviewers, neither explored between-study differences in results with pre-specified subgroup hypotheses, and neither addressed confidence in their effect estimates.[1,16] Hurlbert *et al.* reported a systematic review performed for the 2013 Update of the Guidelines for the Management of Acute Cervical Spine and Spinal Cord Injuries and provided a level I recommendation against methylprednisolone.[14,15] Their recommendation cited a lack of compelling evidence from RCTs or controlled observational studies to support clinical benefit and consistent evidence to suggest harm, including indirect evidence of increased mortality from an RCT of 10,008 participants with head injuries.[64]

Our meta-analysis advances current understanding because it is the first to incorporate the GRADE approach for evaluating confidence in the pooled effect estimates. Confidence ratings are important because they inform evidence users about the quality of the evidence available for clinical decision-making by transparently integrating study design, risk of bias, imprecision, inconsistency, indirectness, and publication bias.[16] The GRADE approach has been adopted by more than 70 major health

research organizations, including the Cochrane Collaboration, the World Heath Organization, and the American College of Physicians.[33]

Our meta-analysis also advances current understanding because it includes recent studies not pooled previously, including two matched cohort studies that were at low risk of bias. Chikuda et al.'s report of 812 pairs from a national administrative database in Japan is the largest controlled study of adverse events, and Evaniew et al.'s report of 44 matched pairs from a national spinal cord injury registry in Canada was the first to adjust for potential confounding due to patients' neurological level of injury and baseline severity of impairment.[1,19] Other older but previously overlooked studies were also retrieved due to our broad search strategy and rigorous methodology.[41,52,54]

Implications

Our results support current guideline recommendations against routine administration of methylprednisolone for patients with acute TSCIs, but strong recommendations are not warranted.[14,65] Guidelines panels must integrate confidence in effect estimates with the balance of desirable and undesirable consequences for alternative management strategies, estimated values and preferences of typical patients, and potential use of healthcare resources.[66,67] Methylprednisolone did not provide significant

long-term benefit, but all of the 95% confidence intervals for motor recovery were compatible with benefit or harm, the effect estimates for gastrointestinal bleeding were conflicting, confidence for all outcomes was low or very low, and the values and preferences of typical patients are likely to be variable.[68–70]

Further research could increase confidence in the effect estimates and clarify the influence of potential confounders or effect modifiers, but utilization of methylprednisolone has already declined sharply at many centers in the last decade and equipoise among individual clinicians may be lacking. Although some clinicians report belief in efficacy or medicolegal concerns, Hurlbert et al. found that more than 75% of Canadian spine surgeons do not prescribe methylprednisolone and Schroeder et al. found a 37% relative decrease in utilization among members of the Cervical Spine Research Society.[9,10] The clinical burden, cost, and multi-center infrastructure required to conduct a well-designed trial may also be prohibitive.

Conclusions

Pooled evidence from multiple RCTs and observational studies does not demonstrate a significant long-term benefit for patients with acute TSCIs and suggests that methylprednisolone may be associated with increased

gastrointestinal bleeding. These findings support current guidelines against routine use, but strong recommendations are not warranted because confidence in the effect estimates is limited. Further research could increase confidence in the effect estimates and clarify the influence of potential confounders or effect modifiers.

Author Disclosure Statement: No competing financial interests exist.

Acknowledgement: Dr. Dvorak holds the Cordula & Günter Paetzold Chair in Spinal Cord Injury Clinical Research at the International Collaboration On Repair Discoveries at the University of British Columbia.

REFERENCES

1. Evaniew, N., Noonan, V., Fallah, N., Kwon, B.K., Rivers, C.S., Ahn, H., Bailey, C., Christie, S., Fourney, D.R., Hurlbert, R.J., Linassi, G., Fehlings, M., and Dvorak, M.F. (2015). Methylprednisolone for the treatment of patients with acute spinal cord injuries: A propensity score-matched cohort study from a Canadian multicenter spinal cord injury registry. J. Neurotrauma .
2. Witiw, C.D., and Fehlings, M.G. (2015). Acute Spinal Cord Injury. J. Spinal Disord. Tech. 28, 202–210.
3. Sekhon, L.H., and Fehlings, M.G. (2001). Epidemiology, demographics, and pathophysiology of acute spinal cord injury. Spine 26, S2–12.
4. Singh, A., Tetreault, L., Kalsi-Ryan, S., Nouri, A., and Fehlings, M.G. (2014). Global prevalence and incidence of traumatic spinal cord injury. Clin Epidemiol 6, 309–31.
5. WHO. (2013). Spinal Cord Injury, Fact Sheet.Available from: http://www.who.int/mediacentre/factsheets/fs384/en/.
6. Bracken, M.B., Shepard, M.J., Collins, W.F., Holford, T.R., Young, W., Baskin, D.S., Eisenberg, H.M., Flamm, E., Leo-Summers, L., Maroon, J., and al, et. (1990). A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury. Results of the Second National Acute Spinal Cord Injury Study. N Engl J Med 322, 1405–11.
7. Bracken, M.B., Shepard, M.J., Collins, W.F., Holford, T.R., Baskin, D.S., Eisenberg, H.M., Flamm, E., Leo-Summers, L., Maroon, J.C., Marshall, L.F., and al, et. (1992). Methylprednisolone or naloxone treatment after acute spinal cord injury: 1-year follow-up data. Results of the second National Acute Spinal Cord Injury Study. J Neurosurg 76, 23–31.
8. Bracken, M.B. (2012). Steroids for acute spinal cord injury. Cochrane Database Syst Rev 1, Cd001046.
9. Hurlbert, R.J., and Hamilton, M.G. (2008). Methylprednisolone for acute spinal cord injury: 5-year practice reversal. Can J Neurol Sci 35, 41–5.
10. Schroeder, G.D., Kwon, B.K., Eck, J.C., Savage, J.W., Hsu, W.K., and Patel, A.A. (2014). Survey of cervical spine research society members on the use of high-dose steroids for acute spinal cord injuries. Spine Phila Pa 1976 39, 971–7.
11. Miekisiak, G., Kloc, W., Janusz, W., Kaczmarczyk, J., Latka, D., and Zarzycki, D. (2014). Current use of methylprednisolone for acute spinal cord injury in Poland: survey study. Eur J Orthop Surg Traumatol 24 Suppl 1, S269–73.

12. Druschel, C., Schaser, K.D., and Schwab, J.M. (2013). Current practice of methylprednisolone administration for acute spinal cord injury in Germany: a national survey. Spine Phila Pa 1976 38, E669–77.
13. Hurlbert, R.J. (2014). Methylprednisolone for the treatment of acute spinal cord injury: point. Neurosurgery 61 Suppl 1, 32–5.
14. Walters, B.C., Hadley, M.N., Hurlbert, R.J., Aarabi, B., Dhall, S.S., Gelb, D.E., Harrigan, M.R., Rozelle, C.J., Ryken, T.C., and Theodore, N. (2013). Guidelines for the management of acute cervical spine and spinal cord injuries: 2013 update. Neurosurgery 60 Suppl 1, 82–91.
15. Hurlbert, R.J., Hadley, M.N., Walters, B.C., Aarabi, B., Dhall, S.S., Gelb, D.E., Rozzelle, C.J., Ryken, T.C., and Theodore, N. (2013). Pharmacological therapy for acute spinal cord injury. Neurosurgery 72 Suppl 2, 93–105.
16. Evaniew, N., van der Watt, L., Bhandari, M., Ghert, M., Aleem, I., Drew, B., and Guyatt, G. (2015). Strategies to improve the credibility of meta-analyses in spine surgery: A systematic survey. Spine J. .
17. Murad, M.H., Montori, V.M., Ioannidis, J.P., Jaeschke, R., Devereaux, P.J., Prasad, K., Neumann, I., Carrasco-Labra, A., Agoritsas, T., Hatala, R., Meade, M.O., Wyer, P., Cook, D.J., and Guyatt, G. (2014). How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA 312, 171–9.
18. Botelho, R.V., Daniel, J.W., Boulosa, J.L., Colli, B.O., Farias Rde, L., Moraes, O.J., Pimenta, W.E., Ribeiro, C.H., Ribeiro, F.R., Taricco, M.A., Carvalho, M.V., and Bernardo, W.M. (2009). [Effectiveness of methylprednisolone in the acute phase of spinal cord injuries--a systematic review of randomized controlled trials]. Rev Assoc Med Bras 55, 729–37.
19. Chikuda, H., Yasunaga, H., Takeshita, K., Horiguchi, H., Kawaguchi, H., Ohe, K., Fushimi, K., and Tanaka, S. (2014). Mortality and morbidity after high-dose methylprednisolone treatment in patients with acute cervical spinal cord injury: a propensity-matched analysis using a nationwide administrative database. Emerg Med J 31, 201–6.
20. Khan, M.F., Burks, S.S., Al-Khayat, H., and Levi, A.D. (2014). The effect of steroids on the incidence of gastrointestinal hemorrhage after spinal cord injury: a case-controlled study. Spinal Cord 52, 58–60.
21. Higgins, J.P., and Green, S. ([date unknown]). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. [cited 2015 Jun 7] Available from: www.cochrane-handbook.org.
22. Moher, D., Liberati, A., Tetzlaff, J., and Altman, D.G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-

Analyses: The PRISMA Statement. J. Clin. Epidemiol. 62, 1006–1012.
23. Slim, K., Nini, E., Forestier, D., Kwiatkowski, F., Panis, Y., and Chipponi, J. (2003). Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J. Surg. 73, 712–716.
24. Fawcett, J.W., Curt, A., Steeves, J.D., Coleman, W.P., Tuszynski, M.H., Lammertse, D., Bartlett, P.F., Blight, A.R., Dietz, V., Ditunno, J., Dobkin, B.H., Havton, L.A., Ellaway, P.H., Fehlings, M.G., Privat, A., Grossman, R., Guest, J.D., Kleitman, N., Nakamura, M., Gaviria, M., and Short, D. (2007). Guidelines for the conduct of clinical trials for spinal cord injury as developed by the ICCP panel: spontaneous recovery after spinal cord injury and statistical power needed for therapeutic clinical trials. Spinal Cord 45, 190–205.
25. Street, J.T., Thorogood, N.P., Cheung, A., Noonan, V.K., Chen, J., Fisher, C.G., and Dvorak, M.F. (2013). Use of the Spine Adverse Events Severity System (SAVES) in patients with traumatic spinal cord injury. A comparison with institutional ICD-10 coding for the identification of acute care adverse events. Spinal Cord 51, 472–6.
26. Guyatt, G.H., Oxman, A.D., Kunz, R., Atkins, D., Brozek, J., Vist, G., Alderson, P., Glasziou, P., Falck-Ytter, Y., and Schünemann, H.J. (2011). GRADE guidelines: 2. Framing the question and deciding on important outcomes. J. Clin. Epidemiol. 64, 395–400.
27. Kirshblum, S., and Waring, W. (2014). Updates for the International Standards for Neurological Classification of Spinal Cord Injury. Phys Med Rehabil Clin N Am 25, 505–17, vii.
28. Evaniew, N., Khan, M., Drew, B., Peterson, D., Bhandari, M., and Ghert, M. (2015). Intrawound vancomycin to prevent infections after spine surgery: a systematic review and meta-analysis. Eur. Spine J. 24, 533–542.
29. Dvorak, M.F., Noonan, V.K., Fallah, N., Fisher, C.G., Rivers, C.S., Ahn, H., Tsai, E.C., Linassi, A.G., Christie, S.D., Attabib, N., Hurlbert, R.J., Fourney, D.R., Johnson, M.G., Fehlings, M.G., Drew, B., Bailey, C.S., Paquet, J., Parent, S., Townson, A., Ho, C., Craven, B.C., Gagnon, D., Tsui, D., Fox, R., Mac-Thiong, J.M., and Kwon, B.K. (2014). Minimizing errors in acute traumatic spinal cord injury trials by acknowledging the heterogeneity of spinal cord anatomy and injury severity: an observational Canadian cohort analysis. J Neurotrauma 31, 1540–7.
30. Dvorak, M.F., Noonan, V.K., Fallah, N., Fisher, C.G., Finkelstein, J., Kwon, B.K., Rivers, C.S., Ahn, H., Paquet, J., Tsai, E.C., Townson, A., Attabib, N., Bailey, C.S., Christie, S.D., Drew, B., Fourney, D.R., Fox, R., Hurlbert, R.J., Johnson, M.G., Linassi, A.G., Parent, S., and Fehlings, M.G. (2014). The Influence of Time from Injury to Surgery

on Motor Recovery and Length of Hospital Stay in Acute Traumatic Spinal Cord Injury: An Observational Canadian Cohort Study. J Neurotrauma .
31. Schulz, K.F., Chalmers, I., Hayes, R.J., and Altman, D.G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273, 408–412.
32. Kirshblum, S., Millis, S., McKinley, W., and Tulsky, D. (2004). Late neurologic recovery after traumatic spinal cord injury. Arch Phys Med Rehabil 85, 1811–7.
33. Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E., Falck-Ytter, Y., Meerpohl, J., Norris, S., and Guyatt, G.H. (2011). GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64, 401–6.
34. Guyatt, G.H., Oxman, A.D., Kunz, R., Brozek, J., Alonso-Coello, P., Rind, D., Devereaux, P.J., Montori, V.M., Freyschuss, B., Vist, G., Jaeschke, R., Williams, J.W., Murad, M.H., Sinclair, D., Falck-Ytter, Y., Meerpohl, J., Whittington, C., Thorlund, K., Andrews, J., and Schünemann, H.J. (2011). GRADE guidelines 6. Rating the quality of evidence--imprecision. J. Clin. Epidemiol. 64, 1283–1293.
35. Guyatt, G.H., Oxman, A.D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Glasziou, P., Jaeschke, R., Akl, E.A., Norris, S., Vist, G., Dahm, P., Shukla, V.K., Higgins, J., Falck-Ytter, Y., Schünemann, H.J., and GRADE Working Group. (2011). GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J. Clin. Epidemiol. 64, 1294–1302.
36. Sim, J., and Wright, C.C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys. Ther. 85, 257–268.
37. Akl, E.A., Briel, M., You, J.J., Sun, X., Johnston, B.C., Busse, J.W., Mulla, S., Lamontagne, F., Bassler, D., Vera, C., Alshurafa, M., Katsios, C.M., Zhou, Q., Cukierman-Yaffe, T., Gangji, A., Mills, E.J., Walter, S.D., Cook, D.J., Schunemann, H.J., Altman, D.G., and Guyatt, G.H. (2012). Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ 344, e2809.
38. Otani, K., Abe, H., and Kadoya, S. (1994). Beneficial effect of methylprednisolone sodium succinate in the treatment of acute spinal cord injury (translation of Japanese). Sekitsui Sekizui 7, 633–647.
39. Pointillart, V., Petitjean, M.E., Wiart, L., Vital, J.M., Lassié, P., Thicoipé, M., and Dabadie, P. (2000). Pharmacological therapy of spinal cord injury during the acute phase. Spinal Cord 38, 71–76.
40. Matsumoto, T., Tamaki, T., Kawakami, M., Yoshida, M., Ando, M., and Yamada, H. (2001). Early complications of high-dose

methylprednisolone sodium succinate treatment in the follow-up of acute cervical spinal cord injury. Spine 26, 426–430.
41. Gäbler, C., and Maier, R. (1995). [Clinical experiences and results of high-dosage methylprednisolone therapy in spinal cord trauma 1991 to 1993]. Unfallchirurgie 21, 20–29.
42. Gerhart, K.A., Johnson, R.L., Menconi, J., Hoffman, R.E., and Lammertse, D.P. (1995). Utilization and effectiveness of methylprednisolone in a population-based sample of spinal cord injured persons. Paraplegia 33, 316–321.
43. Levy, M.L., Gans, W., Wijesinghe, H.S., SooHoo, W.E., Adkins, R.H., and Stillerman, C.B. (1996). Use of methylprednisolone as an adjunct in the management of patients with penetrating spinal cord injury: outcome analysis. Neurosurgery 39, 1141–1148; discussion 1148–1149.
44. Heary, R.F., Vaccaro, A.R., Mesa, J.J., Northrup, B.E., Albert, T.J., Balderston, R.A., and Cotler, J.M. (1997). Steroids and gunshot wounds to the spine. Neurosurgery 41, 576–583; discussion 583–584.
45. Gerndt, S.J., Rodriguez, J.L., Pawlik, J.W., Taheri, P.A., Wahl, W.L., Micheals, A.J., and Papadopoulos, S.M. (1997). Consequences of high-dose steroid therapy for acute spinal cord injury. J. Trauma 42, 279–284.
46. Pollard, M.E., and Apple, D.F. (2003). Factors associated with improved neurologic outcomes in patients with incomplete tetraplegia. Spine 28, 33–39.
47. Tsutsumi, S., Ueta, T., Shiba, K., Yamamoto, S., and Takagishi, K. (2006). Effects of the Second National Acute Spinal Cord Injury Study of high-dose methylprednisolone therapy on acute cervical spinal cord injury-results in spinal injuries center. Spine 31, 2992–2996; discussion 2997.
48. Suberviola, B., González-Castro, A., Llorca, J., Ortiz-Melón, F., and Miñambres, E. (2008). Early complications of high-dose methylprednisolone in acute spinal cord injury patients. Injury 39, 748–752.
49. Ito, Y., Sugimoto, Y., Tomioka, M., Kai, N., and Tanaka, M. (2009). Does high dose methylprednisolone sodium succinate really improve neurological status in patient with acute cervical cord injury?: a prospective study about neurological recovery and early complications. Spine 34, 2121–2124.
50. Sribnick, E.A., Hoh, D.J., and Dhall, S.S. (2014). Traumatic high-grade cervical dislocation: treatment strategies and outcomes. World Neurosurg. 82, 1374–1379.
51. Prendergast, M.R., Saxe, J.M., Ledgerwood, A.M., Lucas, C.E., and Lucas, W.F. (1994). Massive steroids do not reduce the zone of injury

after penetrating spinal cord injury. J. Trauma 37, 576–579; discussion 579–580.
52. Yokota, H., Kawai, M., Kato, K., Mashiko, K., Yamamoto, Y., Henmi, H., and Otsuka, T. (1995). Significance of methylprednisolone therapy in acute spinal cord injury with special reference to short-term follow-up. J. Jpn. Assoc. Acute Med. 6, 349–354.
53. George, E.R., Scholten, D.J., Buechler, C.M., Jordan-Tibbs, J., Mattice, C., and Albrecht, R.M. (1995). Failure of methylprednisolone to improve the outcome of spinal cord injuries. Am. Surg. 61, 659–663; discussion 663–664.
54. Aomar Millan, M., Cortinas Saenz, M., Delgado Tapia, J., Geronimo Pardo, M., Calatayud Perez, V., and Peyro Garcia, R. (2011). [Assessment of neurologic function and complications in a retrospective cohort of patients with acute spinal cord injury due to trauma treated with large-dose methylprednisolone]. [Spanish]. Rev. Esp. Anestesiol. Reanim. .
55. Furlan, J.C., Fehlings, M.G., Tator, C.H., and Davis, A.M. (2008). Motor and sensory assessment of patients in clinical trials for pharmacological therapy of acute spinal cord injury: psychometric properties of the ASIA Standards. J Neurotrauma 25, 1273–301.
56. Ioannidis, J.P., Haidich, A.B., Pappa, M., Pantazis, N., Kokori, S.I., Tektonidou, M.G., Contopoulos-Ioannidis, D.G., and Lau, J. (2001). Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 286, 821–830.
57. Hoppe, D.J., Schemitsch, E.H., Morshed, S., Tornetta, P., and Bhandari, M. (2009). Hierarchy of evidence: where observational studies fit in and why we need them. J Bone Jt. Surg Am 91 Suppl 3, 2–9.
58. Zhang, Y., Zhang, S., Thabane, L., Furukawa, T.A., Johnston, B.C., and Guyatt, G.H. (2015). Although not consistently superior, the absolute approach to framing the minimally important difference has advantages over the relative approach. J. Clin. Epidemiol. .
59. Fehlings, M.G., Nakashima, H., Nagoshi, N., Chow, D.S.L., Grossman, R.G., and Kopjar, B. (2015). Rationale, design and critical end points for the Riluzole in Acute Spinal Cord Injury Study (RISCIS): a randomized, double-blinded, placebo-controlled parallel multi-center trial. Spinal Cord .
60. Bannuru, R.R., Vaysbrot, E.E., and McIntyre, L.F. (2014). Did the American Academy of Orthopaedic Surgeons osteoarthritis guidelines miss the mark? Arthroscopy 30, 86–89.
61. Fehlings, M.G., Vaccaro, A., Wilson, J.R., Singh, A., W Cadotte, D., Harrop, J.S., Aarabi, B., Shaffrey, C., Dvorak, M., Fisher, C., Arnold, P., Massicotte, E.M., Lewis, S., and Rampersaud, R. (2012). Early versus delayed decompression for traumatic cervical spinal cord

injury: results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). PloS One 7, e32037.
62. Wilson, J.R., Arnold, P.M., Singh, A., Kalsi-Ryan, S., and Fehlings, M.G. (2012). Clinical prediction model for acute inpatient complications after traumatic cervical spinal cord injury: a subanalysis from the Surgical Timing in Acute Spinal Cord Injury Study. J. Neurosurg. Spine 17, 46–51.
63. Simmonds, M.C., Brown, J.V., Heirs, M.K., Higgins, J.P., Mannion, R.J., Rodgers, M.A., and Stewart, L.A. (2013). Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. Ann Intern Med 158, 877–89.
64. Edwards, P., Arango, M., Balica, L., Cottingham, R., El-Sayed, H., Farrell, B., Fernandes, J., Gogichaisvili, T., Golden, N., Hartzenberg, B., Husain, M., Ulloa, M.I., Jerbi, Z., Khamis, H., Komolafe, E., Laloe, V., Lomas, G., Ludwig, S., Mazairac, G., Munoz Sanchez Mde, L., Nasi, L., Olldashi, F., Plunkett, P., Roberts, I., Sandercock, P., Shakur, H., Soler, C., Stocker, R., Svoboda, P., Trenkler, S., Venkataramana, N.K., Wasserberg, J., Yates, D., and Yutthakasemsunt, S. (2005). Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. Lancet 365, 1957–9.
65. Hurlbert, R.J., and Moulton, R. (2002). Why do you prescribe methylprednisolone for acute spinal cord injury? A Canadian perspective and a position statement. Can J Neurol Sci 29, 236–9.
66. Andrews, J.C., Schünemann, H.J., Oxman, A.D., Pottie, K., Meerpohl, J.J., Coello, P.A., Rind, D., Montori, V.M., Brito, J.P., Norris, S., Elbarbary, M., Post, P., Nasser, M., Shukla, V., Jaeschke, R., Brozek, J., Djulbegovic, B., and Guyatt, G. (2013). GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J. Clin. Epidemiol. 66, 726–735.
67. Schünemann, H.J., Wiercioch, W., Etxeandia, I., Falavigna, M., Santesso, N., Mustafa, R., Ventresca, M., Brignardello-Petersen, R., Laisaar, K.-T., Kowalski, S., Baldeh, T., Zhang, Y., Raid, U., Neumann, I., Norris, S.L., Thornton, J., Harbour, R., Treweek, S., Guyatt, G., Alonso-Coello, P., Reinap, M., Brozek, J., Oxman, A., and Akl, E.A. (2014). Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ 186, E123–142.
68. Garrino, L., Curto, N., Decorte, R., Felisi, N., Matta, E., Gregorino, S., Actis, M.V., Marchisio, C., and Carone, R. (2011). Towards personalized care for persons with spinal cord injury: a study on patients' perceptions. J. Spinal Cord Med. 34, 67–75.

69. Lindberg, J., Kreuter, M., Taft, C., and Person, L.-O. (2013). Patient participation in care and rehabilitation from the perspective of patients with spinal cord injury. Spinal Cord 51, 834–837.
70. Whitehurst, D.G., Suryaprakash, N., Engel, L., Mittmann, N., Noonan, V.K., Dvorak, M.F., and Bryan, S. (2014). Perceptions of individuals living with spinal cord injury toward preference-based quality of life instruments: a qualitative exploration. Health Qual Life Outcomes 12, 50.

TABLES

Table 1. Included studies

MPS = methylprednisolone; NR = not reported; NASCIS-II = Second National Spinal Cord Injury Study;
AIS = ASIA Impairment Scale

Authors	Year	Size	Follow-up (months)	Mean age (years)	Males (%)	Cervical (%)	Open (%)	Surgery (%)	Complete (%)	MPS regimen	Within 8 hours (%)	Outcomes
Randomized controlled trials												
Bracken et al.[6–8]	1990/ 1992	333	6, 12	NR	86	NR	1	99	62	NASCIS-II	45	Motor scores, adverse events
Otani et al.[38]	1994	117	6	NR	NR	NR	NR	NR	NR	NASCIS-II	100	Motor scores, adverse events
Pointillart et al.[39]	2000	52	12	30	90	NR	0	NR	45	NASCIS-II	100	Motor scores, adverse events
Matsumoto et al.[40]	2001	46	24	61	91	100	0	0	33	NASCIS-II	100	Adverse events
Observational studies												
Prendergast et al.[51]	1994	54	2	36	80	NR	57	NR	46	NASCIS-II	NR	Motor scores, adverse events
Gäbler et al.[41]	1995	144	>12	39	74	39	NR	100	42	NASCIS-II	100	Frankel >1, adverse events
George et al.[53]	1995	145	<1	34	77	55	9	68	64	NASCIS-II	NR	Adverse events
Gerhart et al.[42]	1995	278	NR	NR	NR	55	6	NR	NR	NASCIS-II	NR	Frankel >1
Yokota et al.[52]	1995	38	<1	37	87	63	NR	61	NR	NASCIS-II	100	Motor scores, adverse events
Levy et al.[43]	1996	236	6	26	94	22	100	7	55	NASCIS-II	100	Frankel >1, adverse events
Heary et al.[44]	1997	224	56	26	91	30	100	15	75	NASCIS-II	100	Motor scores, Frankel >1, adverse events
Gerndt et al.[45]	1997	140	NR	32	77	56	0	77	NR	NASCIS-II	100	Adverse events
Pollard et al.[46]	2003	304	24	NR	NR	100	NR	78	0	NASCIS-II	NR	Motor scores
Tsutsumi et al.[47]	2006	70	6	51	89	100	NR	NR	61	NASCIS-II	100	Motor scores, adverse events
Suberviola et al.[48]	2008	82	<1	42	84	54	NR	21	54	NASCIS-II	100	Frankel >1, adverse events
Ito et al.[49]	2009	79	3	58	80	100	NR	72	27	NASCIS-II	100	Motor scores, AIS >1, adverse events
Aomar Millan et al.[54]	2011	96	6	36	79	NR	NR	NR	47	NASCIS-II	100	Adverse events
Chikuda et al.[19]	2014	1624	<1	61	79	100	NR	23	NR	"High-dose"	NR	Adverse events
Khan et al.[20]	2014	350	NR	44	76	68	NR	100	43	NASCIS-II	100	Adverse events
Sribnick et al.[50]	2014	15	22	37	60	100	NR	100	60	"High-dose"	NR	AIS >1, adverse events
Evaniew et al.[1]	2015	88	4	45	88	73	NR	86	45	NASCIS-II	100	Motor scores, adverse events

Table 2. Summary of findings: methylprednisolone versus placebo or no treatment for patients with acute traumatic spinal cord injuries

RCT = randomized controlled trial. OBS = observational study. MD = mean difference. RR = relative risk.

Outcome	Follow-up	Data source	Number of participants	Quality of evidence[1] (GRADE)	Anticipated effects
Motor score	Short	RCT	414 (2 studies)[6,38]	**LOW** Risk of bias, imprecision	No significant difference between groups (MD 1.19; 95% CI -2.33 to 4.71; p=0.51).
		OBS	308 (5 studies)[1,47,49,51,52]	**VERY LOW** Study design, risk of bias, imprecision	No significant difference between groups (MD 3.04; 95% CI -2.81 to 8.90; p=0.31).
	Long	RCT	335 (2 studies)[6,39]	**LOW** Risk of bias, imprecision	No significant difference between groups (MD -1.11; 95% CI -4.75 to 2.53; p=0.55).
		OBS	528 (2 studies)[44,46]	**VERY LOW** Study design, risk of bias, imprecision	No significant difference between groups (MD 1.37; 95% CI -3.08 to 5.83; p=0.55).
Improvement by ≥ 1 Frankel/AIS grade	Short	OBS	675 (4 studies)[42,43,48,49]	**VERY LOW** Study design, risk of bias, imprecision, inconsistency	No significant difference between groups (RR 1.27; 95% CI 0.75 to 2.17; p=0.37).
	Long	OBS	383 (3 studies)[41,44,50]	**VERY LOW** Study design, risk of bias, imprecision	No significant difference between groups (RR 0.84; 95% CI 0.53 to 1.33; p=0.46).
Total adverse events[2]	Up to 24 months	RCT	595 (4 studies)[6,38,39,40]	**LOW** Risk of bias, imprecision,	No significant difference between groups (RR 1.65; 95% CI 0.62 to 4.41; p=0.32).
	Up to 56 months	OBS	3347 (14 studies)[1,19,20,41,43-45,47-51,53,54,]	**VERY LOW** Study design, risk of bias	No significant difference between groups (RR 1.23; 95% CI 1.00 to 1.52; p=0.05).

[1]**High quality**: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect
[2]Composite of total adverse events: mortality, sepsis, pneumonia, gastrointestinal bleeding, decubitus ulcer, urinary tract infection, venous thromboembolism, surgical site infection.

Table 3. Adverse events

RR = relative risk. Bolded results = statistically significant.

Outcome	Pooled effect estimate from randomized controlled trials	Pooled effect estimate from observational studies
Mortality	RR 0.55; 95% CI 0.24 to 1.28; p=0.17 484 participants (2 studies)	RR 0.73; 95% CI 0.54 to 1.07; p=0.10 2624 participants (10 studies)
Sepsis	RR 1.11; 95% CI 0.52 to 2.40; p=0.79 444 participants (3 studies)	RR 1.44; 95% CI 0.72 to 2.89; p=0.30 2078 participants (5 studies)
Pneumonia	RR 1.26; 95% CI 0.74 to 2.13; p=0.40 444 participants (3 studies)	RR 1.19; 95% CI 0.74 to 1.91; p=0.47 2689 participants (10 studies)
Gastrointestinal bleeding	RR 1.99; 95% CI 0.74 to 5.37; p=0.17 444 participants (3 studies)	**RR 2.18; 95% CI 1.13 to 4.19; p=0.02** 2857 participants (9 studies)
Decubitus ulcer	RR 0.94; 95% CI 0.60 to 1.46; p=0.78 379 participants (2 studies)	RR 2.07; 95% CI 0.96 to 4.45; p=0.06 218 participants (2 studies)
Urinary tract infection	RR 1.01; 95% CI 0.81 to 1.27; p=0.91 444 participants (3 studies)	RR 1.01; 95% CI 0.77 to 1.33; p=0.92 2449 participants (8 studies)
Venous thromboembolism	RR 0.89; 95% CI 0.41 to 1.94; p=0.77 333 participants (1 study)	RR 1.10; 95% CI 0.60 to 2.00; p=0.76 2232 participants (5 studies)
Surgical site infection	RR 2.11; 95% CI 0.81 to 5.49; p=0.13 333 participants (1 study)	RR 0.88; 95% CI 0.44 to 1.78; p=0.73 839 participants (7 studies)
Total adverse events[1]	RR 1.65; 95% CI 0.62 to 4.41; p=0.32 595 participants (4 studies)	RR 1.23; 95% CI 1.00 to 1.52; p=0.05 3347 participants (14 studies)

[1]Composite of total adverse events: mortality, sepsis, pneumonia, gastrointestinal bleeding, decubitus ulcer, urinary tract infection, venous thromboembolism, surgical site infection.

FIGURES

Figure 1. Flow of articles through the systematic review

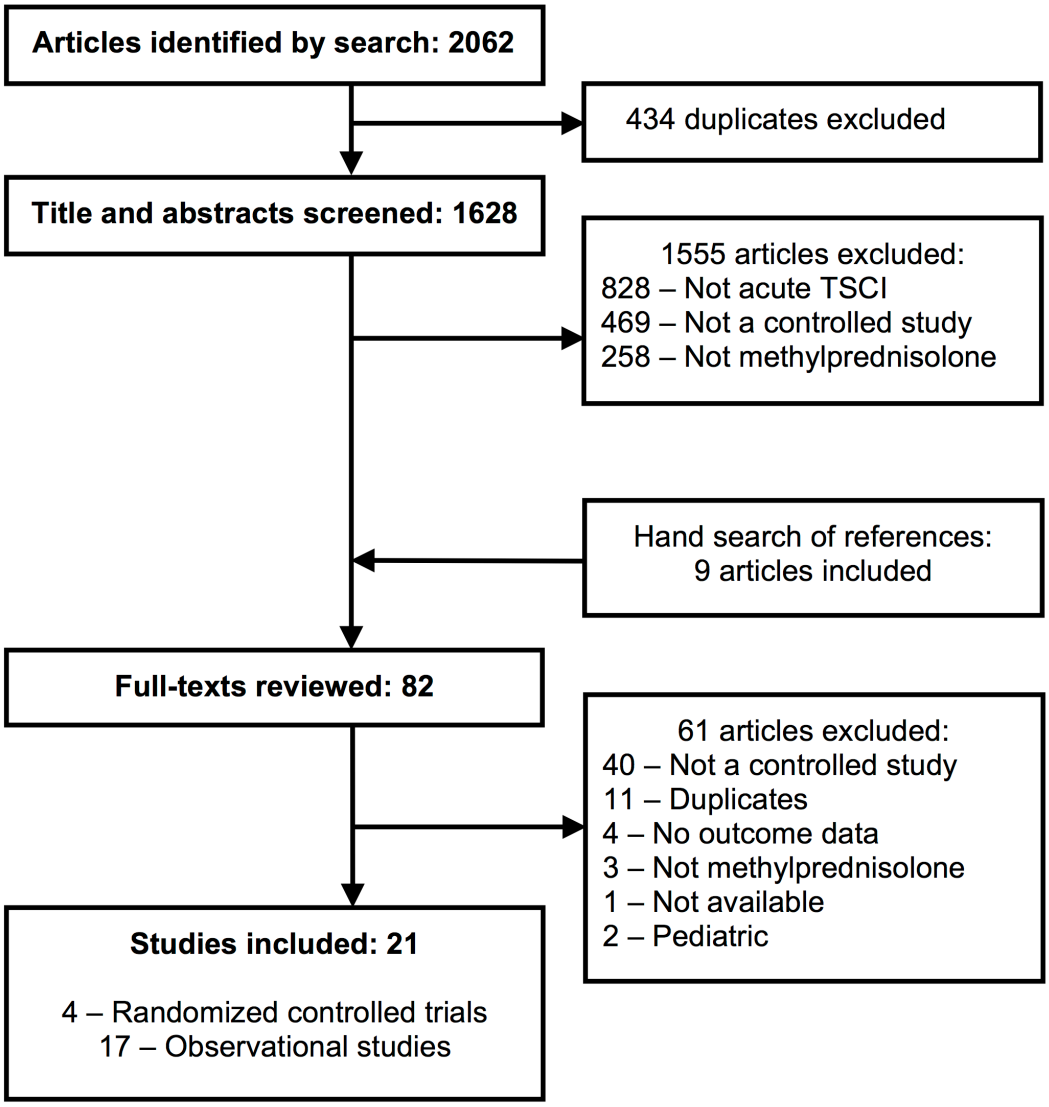
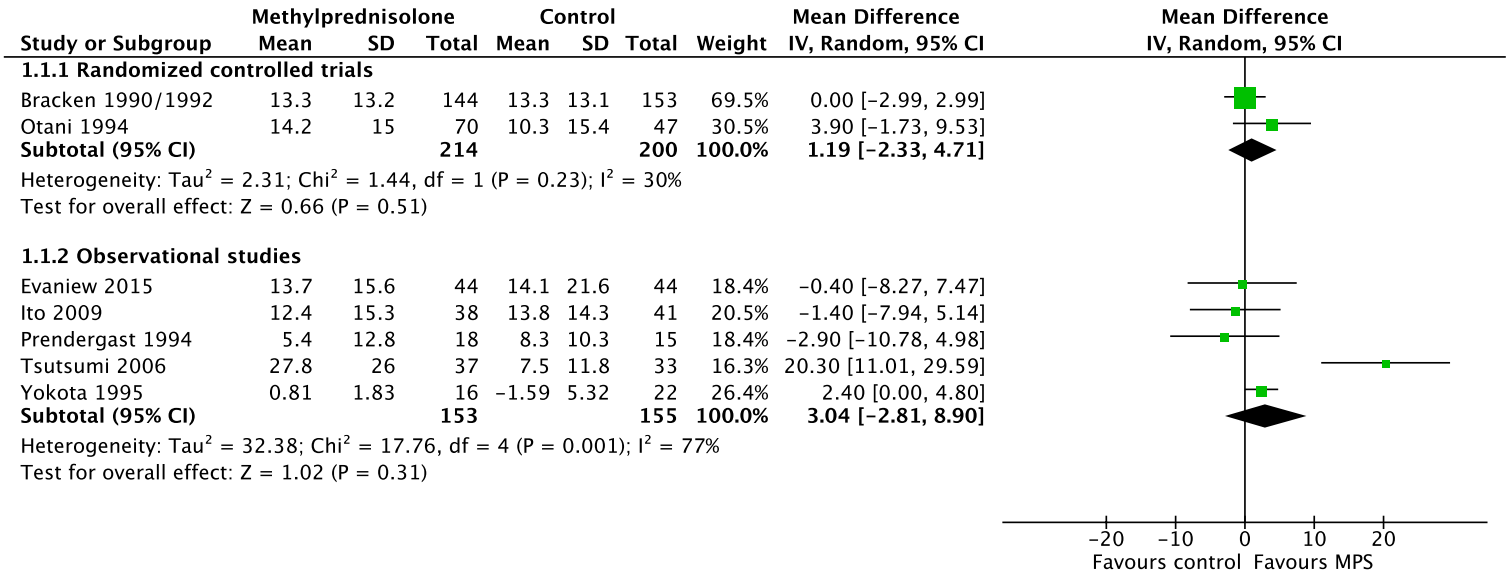


Figure 2. Pooled effect estimates for motor score improvement with methylprednisolone versus placebo or no treatment

(a) Short-term follow-up



(b) Long-term follow-up

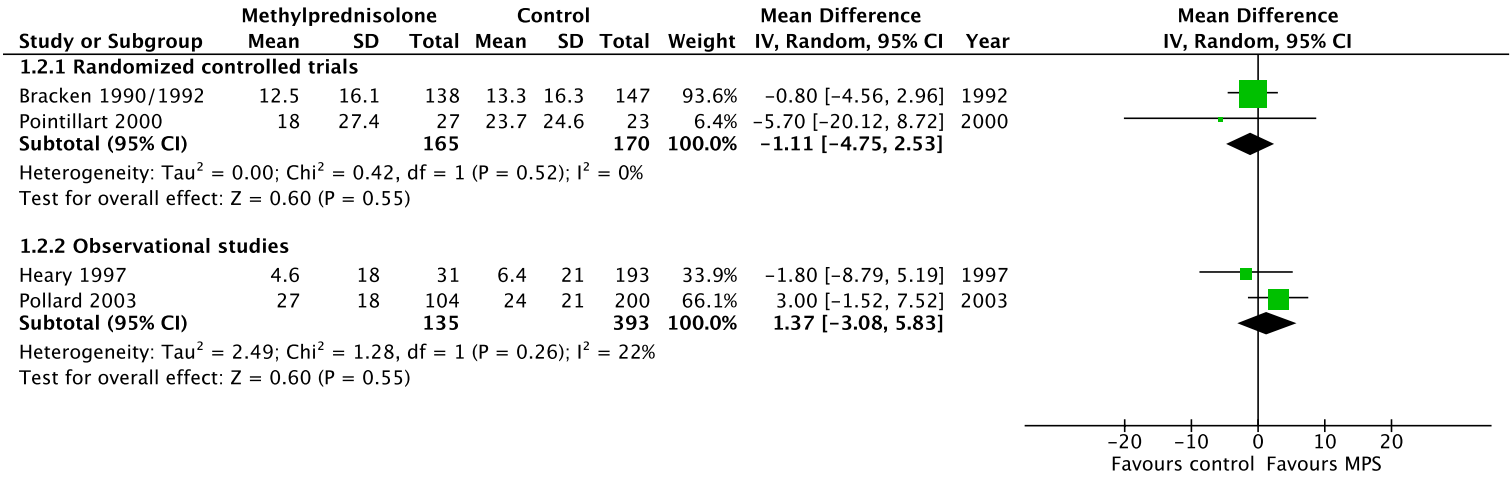
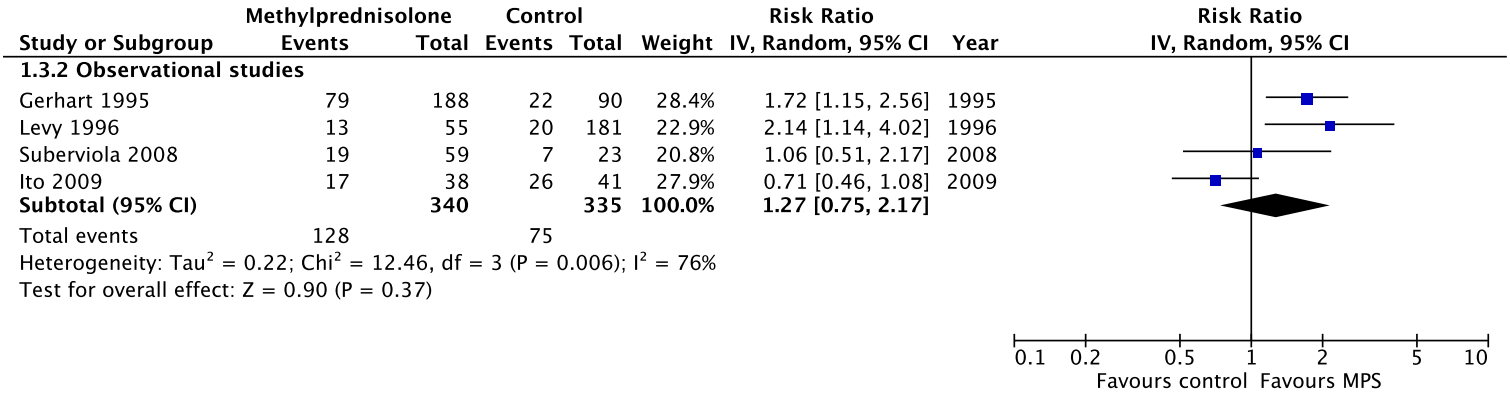
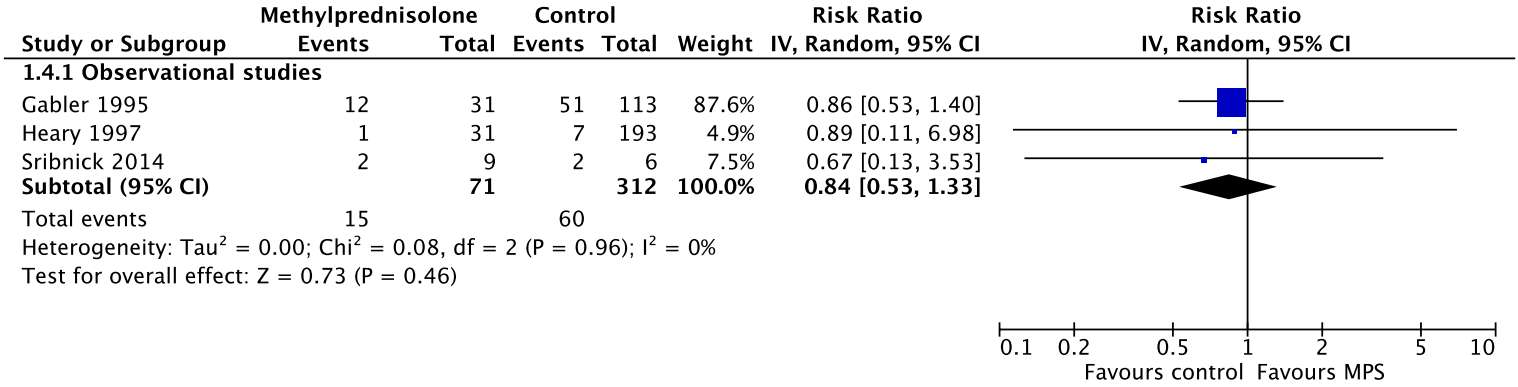


Figure 3. Pooled effect estimates for improvement by one grade or more on the Frankel or ASIA Impairment Scale with methylprednisolone versus placebo or no treatment

(a) Short-term follow-up



(b) Long-term follow-up



APPENDICES

Appendix 1. Electronic search strategy

1. Methylprednisolone/ or methylprednisolone.mp.
2. Adrenal Cortex Hormones/ or corticosteroids.mp.
3. Spinal Cord Injuries/ or spinal cord injury.mp.
4. Spinal Injuries/ or Spinal Fractures/ or spine trauma.mp.
5. (1 or 2) and (3 or 4)
6. limit 5 to humans

Appendix 2. Risk of bias of included studies

(a) Randomized controlled trials - Cochrane Collaboration's Risk of Bias tool

Authors	Year	Sequence	Allocation	Blinding	Data	Reporting	Other
Bracken et al.	1990/1992	unclear	unclear	low	unclear	low	low
Otani et al.	1994	unclear	unclear	high	unclear	low	high
Pointillart et al.	2000	unclear	unclear	high	unclear	low	low
Matsumoto et al.	2001	unclear	unclear	low	unclear	low	low

(b) Observational studies - Methodological Index for Non-Randomized Studies

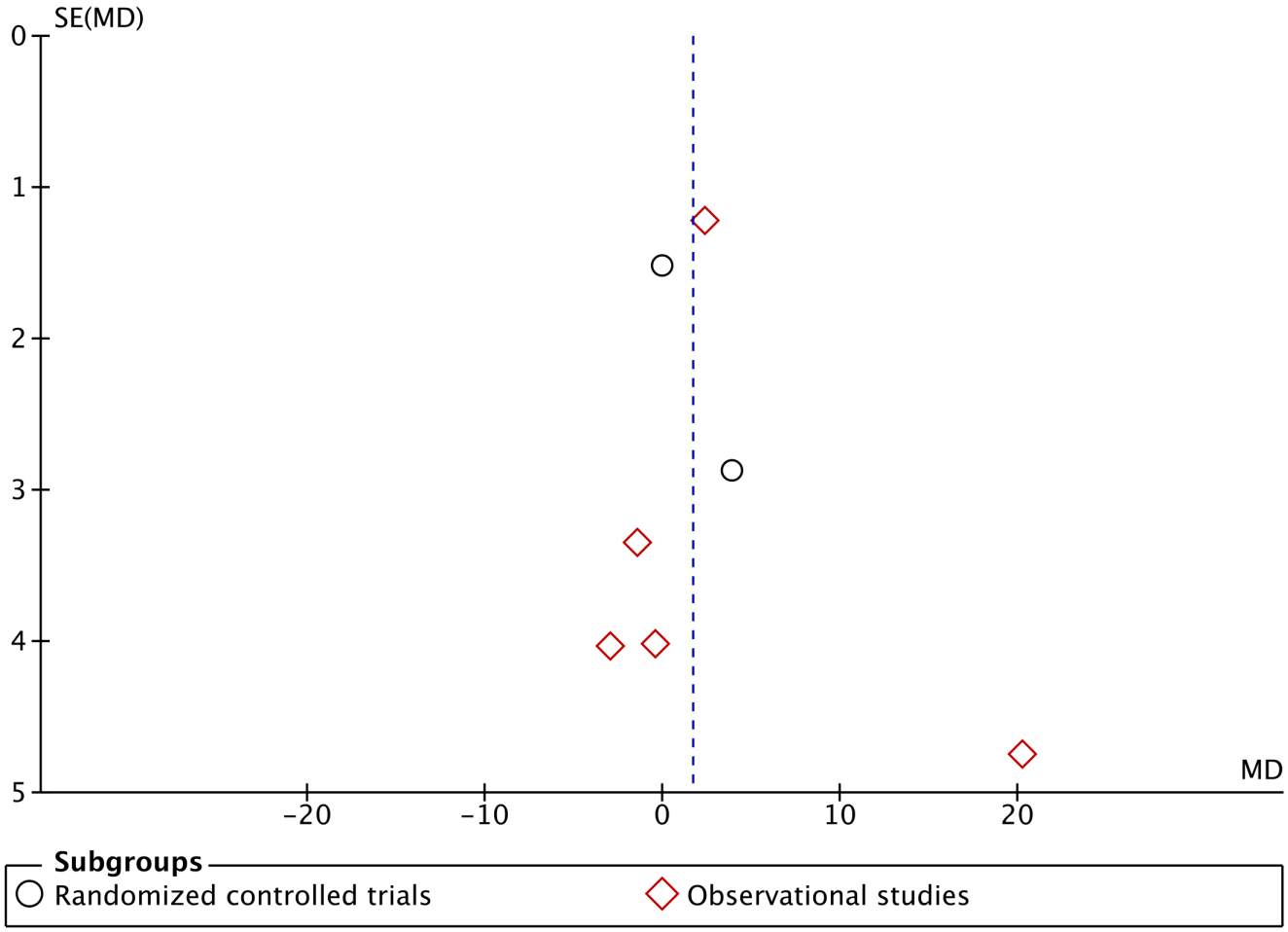
Authors	Year	Clear objective	Consecutive participants	Prospective enrolment	Clear outcomes	Unbiased assessment	Adequate follow-up	Losses <5%	Sample size estimation	Adequate control	Contemporary groups	Baseline equivalence	Statistical analysis	Total score[1]
Prendergast et al.	1994	2	0	1	2	0	2	0	0	2	1	0	1	11
Gabler et al.	1995	2	0	1	2	0	2	0	0	2	1	0	1	11
George et al.	1995	2	2	1	2	0	2	2	0	2	2	1	1	17
Gerhart et al.	1995	2	1	1	2	0	0	0	0	2	2	0	1	11
Yokota et al.	1995	2	2	1	2	0	2	0	0	2	2	0	1	14
Levy et al.	1996	2	0	1	2	0	2	0	0	2	2	0	1	12
Gerndt et al.	1997	2	2	1	2	0	2	0	0	2	1	2	1	15
Heary et al.	1997	2	2	1	2	0	2	2	0	2	2	0	1	14
Pollard et al.	2003	2	0	1	2	0	2	1	0	2	2	0	1	12
Tsutsumi et al.	2006	2	2	1	1	0	2	0	0	2	2	1	1	14
Suberviola et al.	2008	2	2	1	2	0	2	0	0	2	2	1	1	15
Ito et al.	2009	2	2	2	2	0	2	0	0	2	1	1	1	15
Aomar Millan et al.	2011	2	1	1	2	0	1	0	0	2	1	0	1	11
Chikuda et al.	2014	2	0	1	2	0	2	2	0	1	2	2	2	17
Khan et al.	2014	2	0	1	2	1	0	0	0	2	1	1	1	11
Sribnick et al.	2014	2	0	1	2	0	2	1	1	2	2	0	1	14
Evaniew et al.	2015	2	2	2	2	1	2	2	0	2	2	2	2	21

[1]All items were scored as 0 (not reported), 1 (reported but inadequate) or 2 (reported and adequate) towards an ideal score of 24

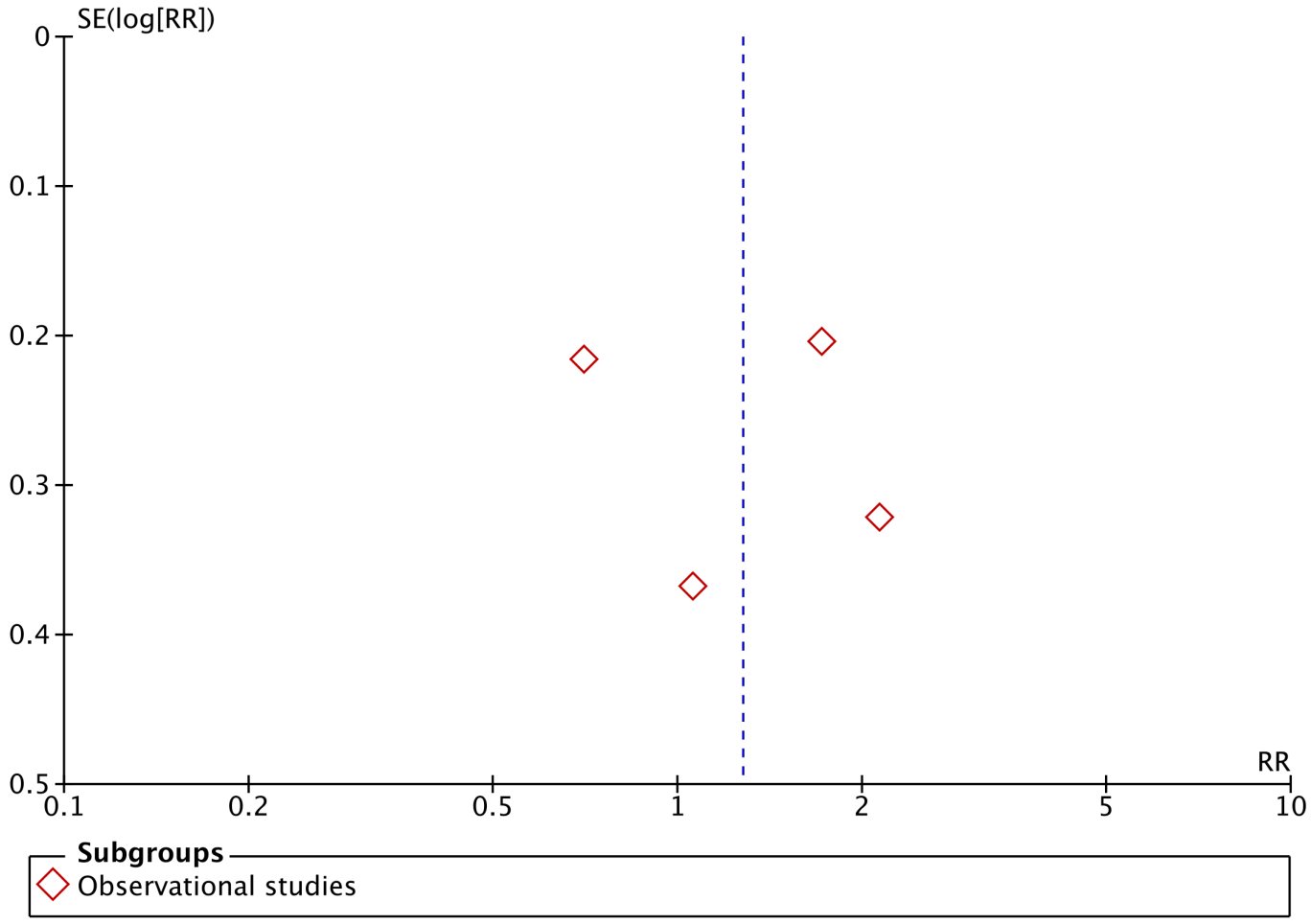
Appendix 3. Funnel plots to evaluate publication bias

(a) Motor scores at short-term follow-up; long-term data were similar.

SE = standard error; MD = mean difference



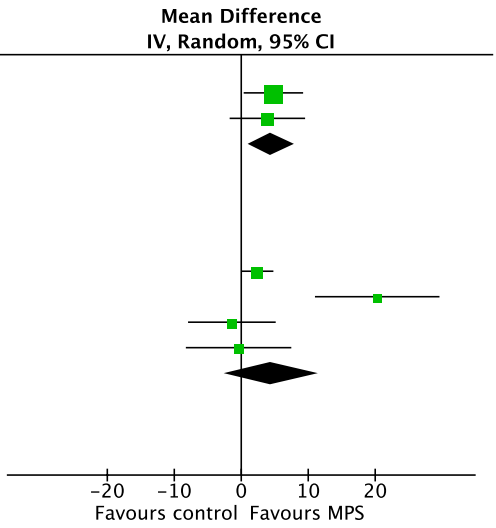
(b) Improvement by one grade or more on the Frankel or ASIA Impairment Scale at short-term follow-up; long-term data were similar. SE = standard error, RR = risk ratio



Appendix 4. Pooled effect estimates for methylprednisolone administered within eight hours of injury versus placebo or no treatment

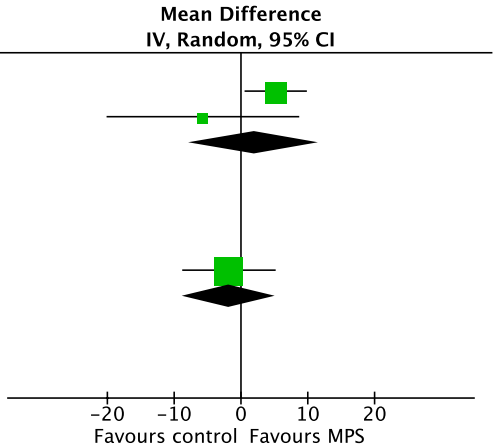
(a) Motor score improvement at short-term follow-up

Study or Subgroup	Methylprednisolone Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.7.1 Randomized controlled trials									
Bracken 1990/1992	16	13.1	65	11.2	13	68	61.7%	4.80 [0.36, 9.24]	1992
Otani 1994	14.2	15	70	10.3	15.4	47	38.3%	3.90 [-1.73, 9.53]	1994
Subtotal (95% CI)			**135**			**115**	**100.0%**	**4.46 [0.97, 7.94]**	
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.06$, df = 1 (P = 0.81); $I^2 = 0\%$									
Test for overall effect: Z = 2.51 (P = 0.01)									
1.7.2 Observational studies									
Yokota 1995	0.81	1.83	16	-1.59	5.32	22	31.2%	2.40 [0.00, 4.80]	1995
Tsutsumi 2006	27.8	26	37	7.5	11.8	33	20.6%	20.30 [11.01, 29.59]	2006
Ito 2009	12.4	15.3	38	13.8	14.3	41	25.2%	-1.40 [-7.94, 5.14]	2009
Evaniew 2015	13.7	15.6	44	14.1	21.6	44	22.9%	-0.40 [-8.27, 7.47]	2015
Subtotal (95% CI)			**135**			**140**	**100.0%**	**4.48 [-2.49, 11.45]**	
Heterogeneity: $Tau^2 = 38.99$; $Chi^2 = 15.94$, df = 3 (P = 0.001); $I^2 = 81\%$									
Test for overall effect: Z = 1.26 (P = 0.21)									



(b) Motor score improvement at long-term follow-up

Study or Subgroup	Methylprednisolone Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.8.1 Randomized controlled trials									
Bracken 1990/1992	17.2	13.4	62	12	13.4	65	70.4%	5.20 [0.54, 9.86]	1992
Pointillart 2000	18	27.4	27	23.7	24.6	23	29.6%	-5.70 [-20.12, 8.72]	2000
Subtotal (95% CI)			**89**			**88**	**100.0%**	**1.97 [-7.78, 11.73]**	
Heterogeneity: $Tau^2 = 29.52$; $Chi^2 = 1.99$, df = 1 (P = 0.16); $I^2 = 50\%$									
Test for overall effect: Z = 0.40 (P = 0.69)									
1.8.2 Observational studies									
Heary 1997	4.6	18	31	6.4	21	193	100.0%	-1.80 [-8.79, 5.19]	1997
Subtotal (95% CI)			**31**			**193**	**100.0%**	**-1.80 [-8.79, 5.19]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 0.50 (P = 0.61)									

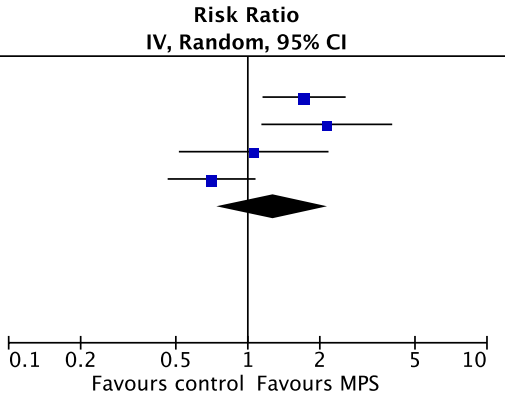


(c) Improvement by one grade or more on the Frankel or ASIA Impairment Scale at short-term follow-up

Study or Subgroup	Methylprednisolone Events	Total	Control Events	Total	Weight	Risk Ratio IV, Random, 95% CI	Year
1.9.1 Observational studies							
Gerhart 1995	79	188	22	90	28.4%	1.72 [1.15, 2.56]	1995
Levy 1996	13	55	20	181	22.9%	2.14 [1.14, 4.02]	1996
Suberviola 2008	19	59	7	23	20.8%	1.06 [0.51, 2.17]	2008
Ito 2009	17	38	26	41	27.9%	0.71 [0.46, 1.08]	2009
Subtotal (95% CI)		**340**		**335**	**100.0%**	**1.27 [0.75, 2.17]**	
Total events	128		75				

Heterogeneity: $Tau^2 = 0.22$; $Chi^2 = 12.46$, df = 3 (P = 0.006); $I^2 = 76\%$

Test for overall effect: Z = 0.90 (P = 0.37)

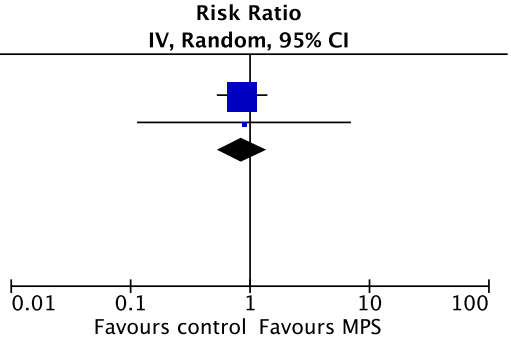


(d) Improvement by one grade or more on the Frankel or ASIA Impairment Scale at long-term follow-up

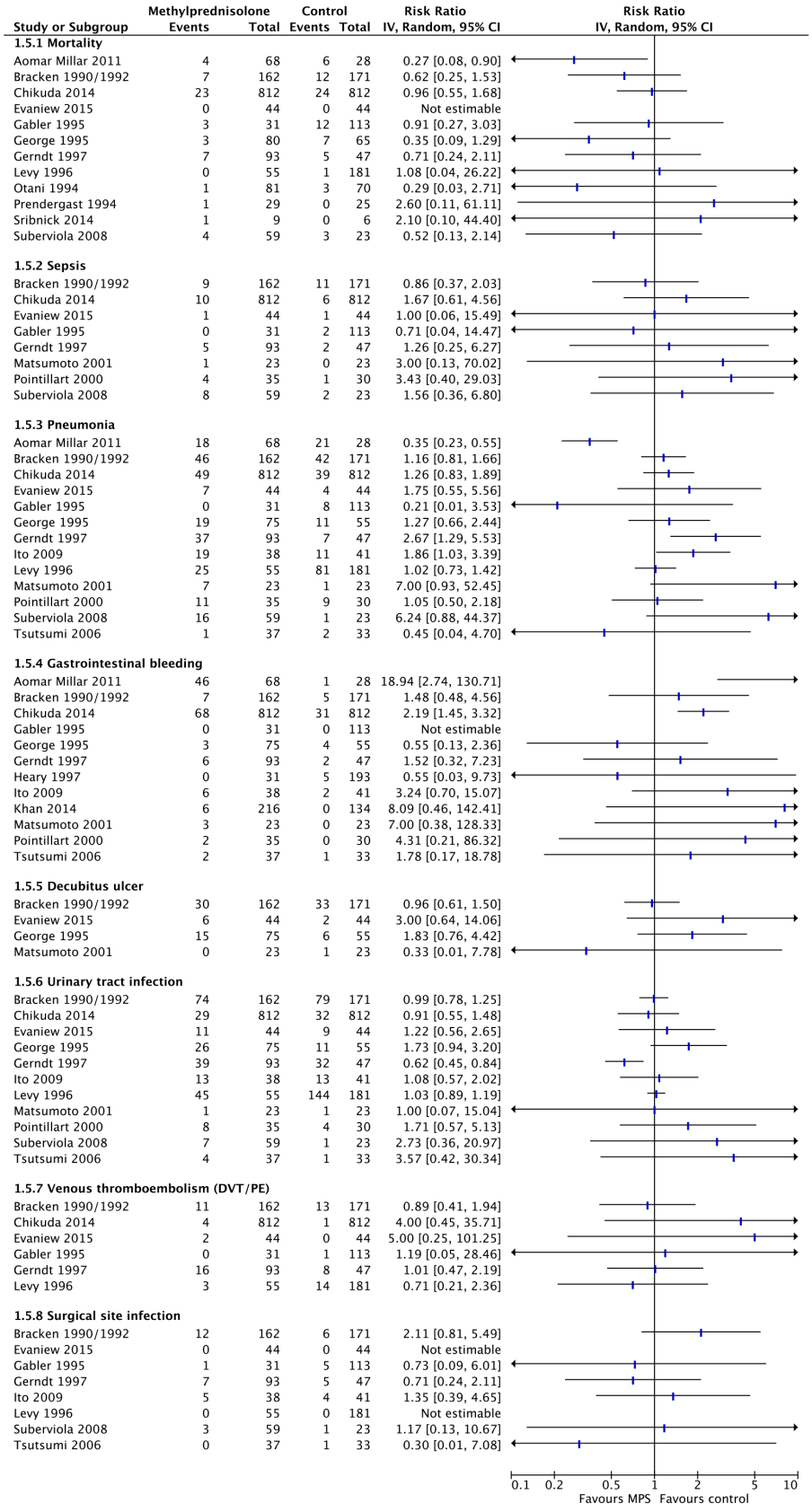
Study or Subgroup	Methylprednisolone Events	Total	Control Events	Total	Weight	Risk Ratio IV, Random, 95% CI	Year
1.10.1 Observational studies							
Gabler 1995	12	31	51	113	94.7%	0.86 [0.53, 1.40]	1995
Heary 1997	1	31	7	193	5.3%	0.89 [0.11, 6.98]	1997
Subtotal (95% CI)		**62**		**306**	**100.0%**	**0.86 [0.53, 1.38]**	
Total events	13		58				

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.00$, df = 1 (P = 0.97); $I^2 = 0\%$

Test for overall effect: Z = 0.63 (P = 0.53)



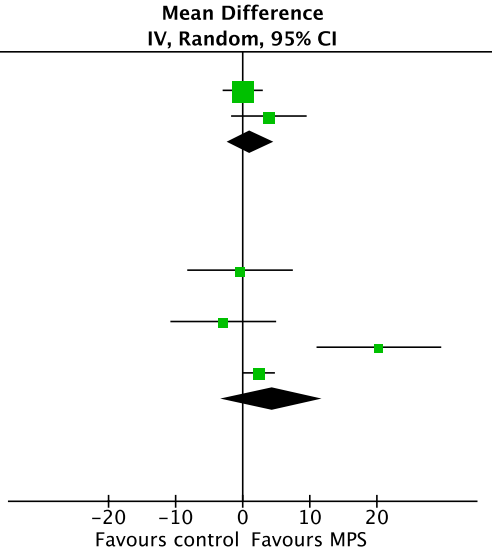
Appendix 5. Individual study data for adverse events at final follow-up



Appendix 6. Sensitivity analyses to test the importance of estimated data

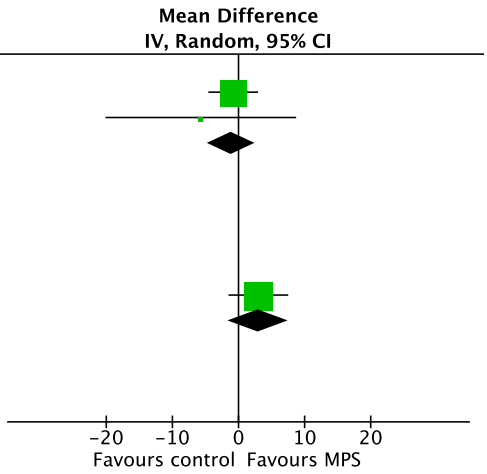
(a) Motor score improvement at short term follow-up with results of study that required estimation of missing SDs (Ito et al., 2009) omitted

	Methylprednisolone			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
1.1.1 Randomized controlled trials								
Bracken 1990/1992	13.3	13.2	144	13.3	13.1	153	69.5%	0.00 [-2.99, 2.99]
Otani 1994	14.2	15	70	10.3	15.4	47	30.5%	3.90 [-1.73, 9.53]
Subtotal (95% CI)			**214**			**200**	**100.0%**	**1.19 [-2.33, 4.71]**
Heterogeneity: Tau^2 = 2.31; Chi^2 = 1.44, df = 1 (P = 0.23); I^2 = 30%								
Test for overall effect: Z = 0.66 (P = 0.51)								
1.1.2 Observational studies								
Evaniew 2015	13.7	15.6	44	14.1	21.6	44	23.7%	-0.40 [-8.27, 7.47]
Ito 2009	12.4	15.3	38	13.8	14.3	41	0.0%	-1.40 [-7.94, 5.14]
Prendergast 1994	5.4	12.8	18	8.3	10.3	15	23.7%	-2.90 [-10.78, 4.98]
Tsutsumi 2006	27.8	26	37	7.5	11.8	33	21.5%	20.30 [11.01, 29.59]
Yokota 1995	0.81	1.83	16	-1.59	5.32	22	31.1%	2.40 [0.00, 4.80]
Subtotal (95% CI)			**115**			**114**	**100.0%**	**4.33 [-3.19, 11.85]**
Heterogeneity: Tau^2 = 45.90; Chi^2 = 16.37, df = 3 (P = 0.0010); I^2 = 82%								
Test for overall effect: Z = 1.13 (P = 0.26)								



(b) Motor score improvement at long term follow-up with results of study that required estimation of missing SDs (Heary et al., 1997) omitted

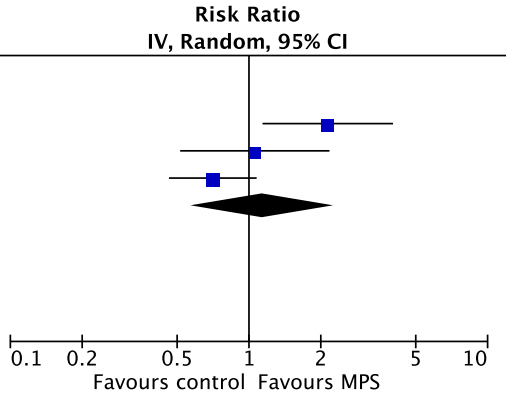
	Methylprednisolone			Control				Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year
1.2.1 Randomized controlled trials									
Bracken 1990/1992	12.5	16.1	138	13.3	16.3	147	93.6%	-0.80 [-4.56, 2.96]	1992
Pointillart 2000	18	27.4	27	23.7	24.6	23	6.4%	-5.70 [-20.12, 8.72]	2000
Subtotal (95% CI)			**165**			**170**	**100.0%**	**-1.11 [-4.75, 2.53]**	
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.42, df = 1 (P = 0.52); I^2 = 0%									
Test for overall effect: Z = 0.60 (P = 0.55)									
1.2.2 Observational studies									
Heary 1997	4.6	18	31	6.4	21	193	0.0%	-1.80 [-8.79, 5.19]	1997
Pollard 2003	27	18	104	24	21	200	100.0%	3.00 [-1.52, 7.52]	2003
Subtotal (95% CI)			**104**			**200**	**100.0%**	**3.00 [-1.52, 7.52]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 1.30 (P = 0.19)									



(c) Improvement by one grade or more on the Frankel or ASIA Impairment Scale at short term follow-up with results of study that reported follow-up as time of discharge (Gerhart et al., 1995) omitted

Study or Subgroup	Methylprednisolone Events	Total	Control Events	Total	Weight	Risk Ratio IV, Random, 95% CI	Year
1.3.2 Observational studies							
Gerhart 1995	79	188	22	90	0.0%	1.72 [1.15, 2.56]	1995
Levy 1996	13	55	20	181	32.3%	2.14 [1.14, 4.02]	1996
Suberviola 2008	19	59	7	23	29.8%	1.06 [0.51, 2.17]	2008
Ito 2009	17	38	26	41	38.0%	0.71 [0.46, 1.08]	2009
Subtotal (95% CI)		**152**		**245**	**100.0%**	**1.14 [0.57, 2.26]**	
Total events	49		53				

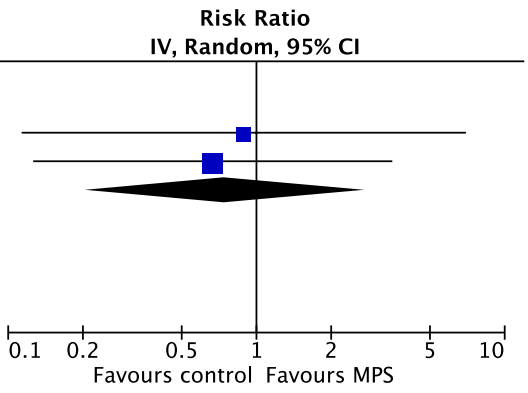
Heterogeneity: $Tau^2 = 0.27$; $Chi^2 = 8.24$, df = 2 (P = 0.02); $I^2 = 76\%$
Test for overall effect: Z = 0.37 (P = 0.71)



(d) Improvement by one grade or more on the Frankel or ASIA Impairment Scale at long term follow-up with results of study that did not report follow-up duration (Gäbler et al., 1995) omitted

Study or Subgroup	Methylprednisolone Events	Total	Control Events	Total	Weight	Risk Ratio IV, Random, 95% CI
1.4.1 Observational studies						
Gabler 1995	12	31	51	113	0.0%	0.86 [0.53, 1.40]
Heary 1997	1	31	7	193	39.5%	0.89 [0.11, 6.98]
Sribnick 2014	2	9	2	6	60.5%	0.67 [0.13, 3.53]
Subtotal (95% CI)		**40**		**199**	**100.0%**	**0.75 [0.20, 2.73]**
Total events	3		9			

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 0.05$, df = 1 (P = 0.83); $I^2 = 0\%$
Test for overall effect: Z = 0.44 (P = 0.66)



Chapter 7

Intrawound vancomycin to prevent infections after spine surgery: A systematic review and meta-analysis

Reprinted from *European Spine Journal*; 24(3); Nathan Evaniew, Moin Khan, Brian Drew, Devin Peterson, Mohit Bhandari, and Michelle Ghert; Intrawound vancomycin to prevent infections after spine surgery: a systematic review and meta-analysis; Pages 533 to 542; Copyright 2014; with permission from Springer Science and Business Media.
doi:10.1007/s00586-014-3357-0.

Eur Spine J (2015) 24:533–542
DOI 10.1007/s00586-014-3357-0

ORIGINAL ARTICLE

Intrawound vancomycin to prevent infections after spine surgery: a systematic review and meta-analysis

Nathan Evaniew · Moin Khan · Brian Drew · Devin Peterson · Mohit Bhandari · Michelle Ghert

Received: 22 March 2014 / Revised: 27 April 2014 / Accepted: 27 April 2014 / Published online: 18 May 2014

Abstract
Purpose Post-operative spine surgical site infections are associated with substantial morbidity, mortality, and economic burden. Intrawound vancomycin may prevent infections after spine surgery, but recent studies have reported conflicting results. The objectives of this systematic review and meta-analysis were to determine: (1) In patients undergoing spine surgery, does the application of intrawound vancomycin lead to reduced rates of post-operative surgical site infections? (2) Are there differences in the estimates of effect between observational studies and randomized trials? (3) What adverse events are reported in the literature?
Methods All published comparative studies of intrawound vancomycin in spine surgery were included. Two reviewers independently screened eligible articles and assessed study quality. Observational studies and randomized trials were pooled separately using a random-effects model.
Results Eight observational studies and one randomized controlled trial met the inclusion criteria. Across observational studies, the odds of infection with intrawound vancomycin was 0.19 times the odds of infection without intrawound vancomycin (95 % CI 0.08–0.47, $p = 0.0003$, $I^2 = 52$ %). The single randomized controlled trial produced a conflicting result (OR 0.96, 95 % CI 0.34–2.66, $p = 0.93$). There were no adverse events attributable to intrawound vancomycin. The quality of the evidence was low or very low.
Conclusions There is a lack of high-quality evidence to inform the use of intrawound vancomycin in spine surgery. Surgeons should be cautious before widely adopting this intervention and should be vigilant in monitoring for adverse effects. Further investigation with additional randomized controlled trials is justified.

Keywords Intrawound vancomycin · Infection · Spine · Evidence-based medicine · Randomized trials

N. Evaniew (✉) · M. Khan · B. Drew · D. Peterson · M. Bhandari · M. Ghert
Division of Orthopaedics, Department of Surgery, McMaster University, Hamilton, ON, Canada
e-mail: nathan.evaniew@medportal.ca

M. Khan
e-mail: mohammed.khan@medportal.ca

B. Drew
e-mail: drewb@hhsc.ca

D. Peterson
e-mail: dpeters@mcmaster.ca

M. Bhandari
e-mail: bhandam@mcmaster.ca

M. Ghert
e-mail: michelle.ghert@jcc.hhsc.ca

M. Bhandari
Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada

Introduction

Post-operative surgical site infections (SSIs) are associated with substantial morbidity, mortality, and economic burden [1–3]. In spine surgery, SSIs are among the most common acute complications and occur in up to 10 % of cases [4–6]. Treatment frequently requires prolonged antibiotic therapy, complex wound management, and multiple operations [4, 7, 8]. Patients' function and health-related quality of life are often dramatically impaired [4, 6, 9].

"Intrawound vancomycin", the application of topical vancomycin powder prior to wound closure, may reduce the incidence of post-operative surgical site infections after spine surgery [10–12]. Approximately 24 % of pediatric spine surgeons currently use intrawound vancomycin, and both adult- and pediatric consensus-based guidelines recommend that intrawound vancomycin be routinely considered in cases with instrumentation, prolonged duration, or significant co-morbidities [13, 14]. It is suggested to be both cost effective and free of systemic complications [15, 16], but potential adverse effects include pseudarthrosis, renal toxicity, and life-threatening anaphylaxis [17–19].

Recent studies on intrawound vancomycin have reported conflicting estimates of efficacy. Several observation studies suggested large benefits, but at least one recent randomized trial has failed to confirm these findings [10, 20, 21]. Some of the discrepancy may be due to the bias inherent in observational designs, but it may also be due to lack of precision. A previous systematic review examined the use of intrawound vancomycin; however, this review included non-spine procedures, did not examine study limitations due to observational study designs, did not estimate zones of clinical equivalence, and did not incorporate a methodological approach to describe the overall quality of the literature [22, 23].

The objectives of this systematic review and meta-analysis were to answer the following questions: (1) In patients undergoing spine surgery for any indication, does the application of intrawound vancomycin in comparison to standard management lead to reduced rates of post-operative surgical site infection? (2) Are there important differences in the estimates of treatment effect between observational studies and randomized trials? (3) What are the adverse events associated with intrawound vancomycin reported in the literature? The current study examines the evidence from observational studies and randomized trials separately, includes estimated zones of clinically equivalent effect sizes, and incorporates the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to summarize evidence to describe the overall quality of the available literature.

Materials and methods

This study was conducted according to the Cochrane Handbook and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [24, 25]. Vancomycin is not FDA approved for intrawound SSI prophylaxis in spine surgery. Intrawound use in this context is unlabeled and investigational.

Eligibility criteria

Studies comparing intrawound vancomycin in spine surgery against standard practice with the outcome of infection were eligible for inclusion. There were no restrictions related to dose of vancomycin, use of instrumentation, definition of infection, age, surgical indication, spinal level(s), co-interventions, year of publication, or language. Non-comparative studies and abstracts from conference proceedings describing unpublished studies were excluded.

Identification of studies

Multiple strategies were used to identify potentially eligible studies. Following training with a professional health sciences librarian, MEDLINE, EMBASE, and the Cochrane databases were systematically searched for articles published up to and including December 4th, 2013. MeSH and EMTREE headings and subheadings were used in various combinations in OVID [i.e., "exp Vancomycin/ AND exp Spine/su (Surgery)"] and supplemented with free text to increase sensitivity [i.e., vancomycin.mp; ("spine or spinal).mp"; "(surgery or surgical or operation or operate).mp"]. The search strategy was adapted in PubMed to search for articles e-published ahead of print. Consultation with experts, hand-searching the reference lists of included full-texts, and the "related articles" feature in PubMed were all used to identify additional studies.

Screening and assessment of eligibility

Two reviewers with methodological and content expertise independently screened all titles and abstracts for eligibility. All discrepancies were resolved by consensus through a process that required reviewers to discuss the rationale for their decisions. Duplicate articles were excluded. There were no articles requiring translation. Reviewers were blinded to author names, journal names, and year of publication.

Assessment of methodological quality

The two reviewers independently graded the methodological quality of each study. Observational studies were evaluated with the Methodological Index for Non-Randomized Studies (MINORS) [26]. MINORS has been validated for the identification of excellent observational studies, and has good inter-reviewer agreement, high test–retest reliability, and strong internal consistency.

The randomized trial was evaluated with the Cochrane Collaboration's Risk of Bias tool [25]. The Cochrane Risk of Bias tool separates judgments about risk of bias from inadequate reporting of methodology. It evaluates each of

selection bias, performance bias, detection bias, attrition bias, reporting bias, and other nonspecific bias.

Extraction of data

Study data were extracted by one reviewer and verified by the second reviewer. Final data (study design, sample size, source of funding, mean age, gender, surgical indication, spinal level, use of instrumentation, dose of intrawound vancomycin, mean follow-up, and infection outcomes) were entered into an electronic spreadsheet for analysis.

Statistical analysis

Inter-observer agreement for the reviewers' assessments of study eligibility was calculated with Cohen's kappa coefficient. Kappa values of >0.65 were considered adequate [27, 28]. Inter-observer agreement for the reviewers' assessments of methodological quality was calculated with the Intraclass Correlation Coefficient (ICC). Kappa and ICC were computed using IBM SPSS Version 21 (Chicago IL, 2012).

Given significant methodological differences between observational studies and randomized controlled trials [25, 29–32], it was decided a priori to pool the infection outcome data from observational studies separately from the infection outcome data from any randomized controlled trials. Odds ratios were computed from the outcome data for each study, weighted by sample size, pooled using the Mantel–Haenszel method and a random-effects model, and displayed with forest plots. Heterogeneity was quantified with the I^2 statistic [25]. A projected zone of clinical equivalence was estimated assuming a 25 % risk reduction in infections [33] and the control event rate generated from the pooled observational data (3.7 %). Pseudarthrosis rates in the study by Sweet et al. were compared using Fisher's exact test. Publication bias was examined by constructing a funnel plot of all studies [25]. The forest plots and the funnel plot were created with Review Manager 5.2 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012).

GRADE quality assessment and summary of findings

The Grades of Recommendation, Assessment, Development and Evaluation (GRADE) Working Group has developed a system for grading the quality of evidence that has been adopted by over 20 major health research organizations and The Cochrane Collaboration [25, 34, 35]. Data from randomized controlled trials were considered high-quality evidence, but could have been rated down according to risk of bias, imprecision, inconsistency, indirectness, or publication bias. Data from moderate-quality trials could have been rated up according to large effect, dose–response, or if all plausible biases would not undermine the conclusions. Data from observational studies were considered low quality, but could have been rated up according to a large treatment effect or evidence of a dose–response relationship. The evidence was graded by two independent assessors and discrepancies were resolved by consensus.

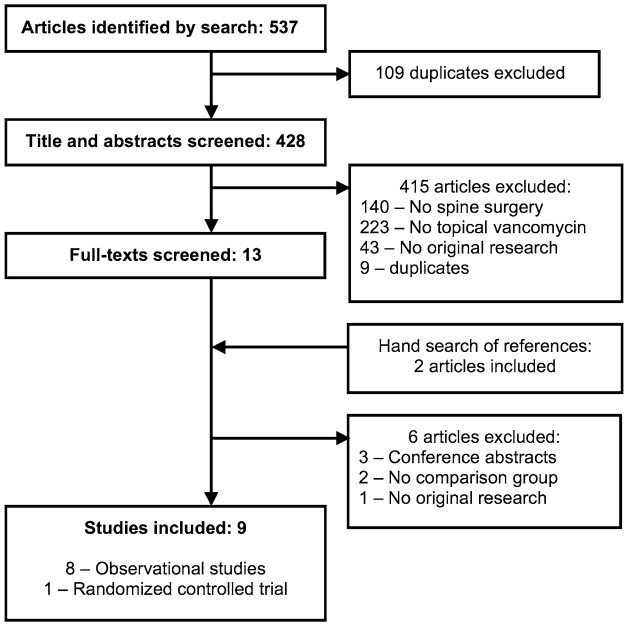


Fig. 1 Flow of articles through the systematic review

Results

Studies included

The electronic literature search identified 537 potentially relevant articles: 450 from EMBASE, 79 from MEDLINE, one from the Cochrane Database, and seven from PubMed. Review of full-texts led to the exclusion of six articles and the inclusion of two studies identified from reference lists. Eight retrospective cohort studies and one randomized controlled trial were included for final analysis, for a total of 5,275 patients (Fig. 1). Agreement between the reviewers for eligibility was satisfactory (kappa = 0.83, 95 % CI 0.66–0.99).

Study characteristics

All of the included studies were published within the last 3 years (Table 1). Sample size in the observational studies ranged from 110 to 1,732, and reported mean follow-up

Table 1 Included studies

Study	Year	Country	Funding	Design	Size	Mean follow-up (years)	Age (mean)	Males (%)	Cervical (%)	Thoracolumbar (%)	Instrumented (%)	Dose (g)	Adverse events
Caroom et al.	2013	USA	Industry	Retrospective cohort	112	NR	57.6	NR	100	–	100	1.0	0
Heller et al.	2013	USA	Institutional	Retrospective cohort	683	NR	52.2	47	NR	NR	100	0.5–1.0	0
Martin et al.	2013	USA	None	Retrospective cohort	306	NR	63.1	32	–	100	100	2.0	0
O'Neill et al.	2011	USA	None	Retrospective cohort	110	0.52	44.0	64	40	60	100	1.0	0
Pahys et al.	2013	USA	None	Retrospective cohort	1001	NR	55.3	58	100	–	86	0.5	0
Strom et al. (1)	2013	USA	None	Retrospective cohort	253	3.08	64.0	56	–	100	65	1.0	0
Strom et al. (2)	2013	USA	None	Retrospective cohort	171	3.44	60.0	59	100	–	100	1.0	0
Sweet et al.	2011	USA	None	Retrospective cohort	1732	2.66	54.6	50	–	100	100	2.0	0
Tubaki et al.	2013	India	Institutional	RCT	907	1.03	45.5	56	12	87	67	1.0	0

RCT randomized controlled trial, *NR* not reported

ranged from 0.52 years to 3.44 years. The single randomized controlled trial had a sample size of 907 and a mean follow-up of 1.03 years. All studies reported exclusively on posterior surgery. Surgical indications included degenerative diseases, trauma, deformity correction, tumor, and other. Instrumentation was used in 4,743 patients (89 %). Only one study reported industry funding (from spinal implant manufacturers) [13].

Intrawound vancomycin

Dosing of intrawound vancomycin varied from 0.5 to 2.0 g per patient, and descriptions about application were variable. Caroom et al. [36] reported placing Gelfoam (Pharmacia and Upjohn Co, Kalamazoo, MI) over exposed dura before applying 1 g of vancomycin powder subfascially along the bone graft and instrumentation, while Heller et al. [37] reported that 0.5–2 g of powder was applied "directly to the wound just prior to closure". Pahys et al. [38] stated that 0.5 mg was "added to the wound". Martin et al. [21], O'Neill et al. [11], Strom et al. (1) [39], and Tubaki et al. [20] described placing powder directly on "muscle, fascia, and subcutaneous tissues" while taking care not to expose bone graft or dura. Strom et al. (2) [40] did not specify whether dura or bone graft was exposed. Sweet et al. [41] reported mixing 1.0 g of vancomycin powder with the bone grafting material before applying an additional 2.0 g of vancomycin power directly into the wound during closure.

Study quality and heterogeneity

MINORS scores ranged from 13 to 17 out of 24 (mean 14.0, 95 % CI 12.8–15.2) across the observational studies (Table 2). All studies used at least some retrospective data, three of the studies did not report whether patients were enrolled consecutively, and seven used historical rather than concurrent control groups. Only two studies included a power analysis, one study reported an Odds Ratio or a relative risk reduction, and none incorporated unbiased outcomes assessment. Agreement between the two reviewers for MINORS assessment was high (ICC = 0.92, 95 % CI 0.88–0.95).

Both reviewers graded the randomized controlled trial to be at 'unclear' risk of bias. The trial reported acceptable sequence generation, but unclear allocation concealment, unclear blinding of participants or outcome assessors, unclear completeness of data, unclear selective outcome reporting, and unclear other sources of bias.

There was moderate-to-large heterogeneity across the observational studies ($I^2 = 52$ %) [32]. The funnel plot was asymmetric and suggestive of publication bias (Fig. 2), but the sample of only nine studies limits interpretability [25].

Table 2 MINORS (Methodological Index for Non-Randomized Studies) scores

Study	Clearly stated objective	Consecutive patients	Prospective study	Unambiguous endpoints	Unbiased assessment	Length of follow-up	Losses <5 %	Adequate sample size calculation	Adequate control group	Contemporary groups	Baseline equivalence	Adequate statistical analyses	Total
Caroom et al.	2	2	1	1	0	1	0	1	2	1	2	1	14
Heller et al.	2	2	1	2	0	1	0	0	2	1	1	1	13
Martin et al.	2	2	1	2	0	1	0	0	2	1	2	2	15
O'Neill et al.	2	1	1	1	0	1	0	1	2	2	2	1	14
Pahys et al.	2	2	1	2	0	1	0	0	1	1	2	1	13
Strom et al. (1)	2	1	1	1	0	2	0	0	2	1	2	1	13
Strom et al. (2)	2	1	1	1	0	2	0	0	2	1	2	1	13
Sweet et al.	2	2	1	2	0	2	2	0	2	1	2	1	17

Items were scored 0 (not reported), 1 (reported but inadequate) or 2 (reported and adequate) towards a global ideal score of 24

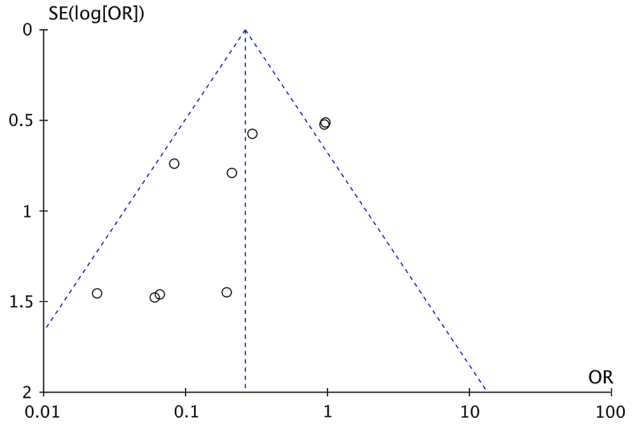


Fig. 2 Funnel plot to assess for publication bias

Surgical site infections

Across the observational studies, 16 surgical site infections were identified among the 1,933 patients who received intrawound vancomycin (0.83 %), as compared to 91 among the 2,435 patients who did not (3.7 %). On the basis of the pooled estimate, the odds of experiencing an infection with intrawound vancomycin was 0.19 times the odds of experiencing an infection without intrawound vancomycin (95 % CI 0.08–0.47, $p = 0.0003$, $I^2 = 52$ %; Fig. 3a).

In the single randomized controlled trial, 7 infections were identified among the 433 patients who received intrawound vancomycin (1.6 %), compared to 8 among the 474 patients who did not (1.7 %). The odds of experiencing an infection with intrawound vancomycin was 0.96 times the odds of experiencing an infection without intrawound vancomycin (95 % CI 0.34–2.66, $p = 0.93$, Fig. 3b).

Adverse events

All studies reported that there were no adverse events attributable to intrawound vancomycin. There were no reported cases of renal failure or anaphylaxis. Sweet et al. [41] were the only group of authors to describe pseudarthrosis rates, reporting three pseudarthroses in their intrawound vancomycin group ($n = 911$) as compared to four in their control group ($n = 821$). There was no statistically significant difference in the pseudarthrosis rate between the two groups in this study ($p = 0.714$). Methods used to investigate possible pseudarthroses were not described.

GRADE quality assessment and summary of findings

Both reviewers rated down the quality of the evidence from the randomized trial from high to low quality based on (1) unclear risk of bias and (2) imprecision due to the fragility

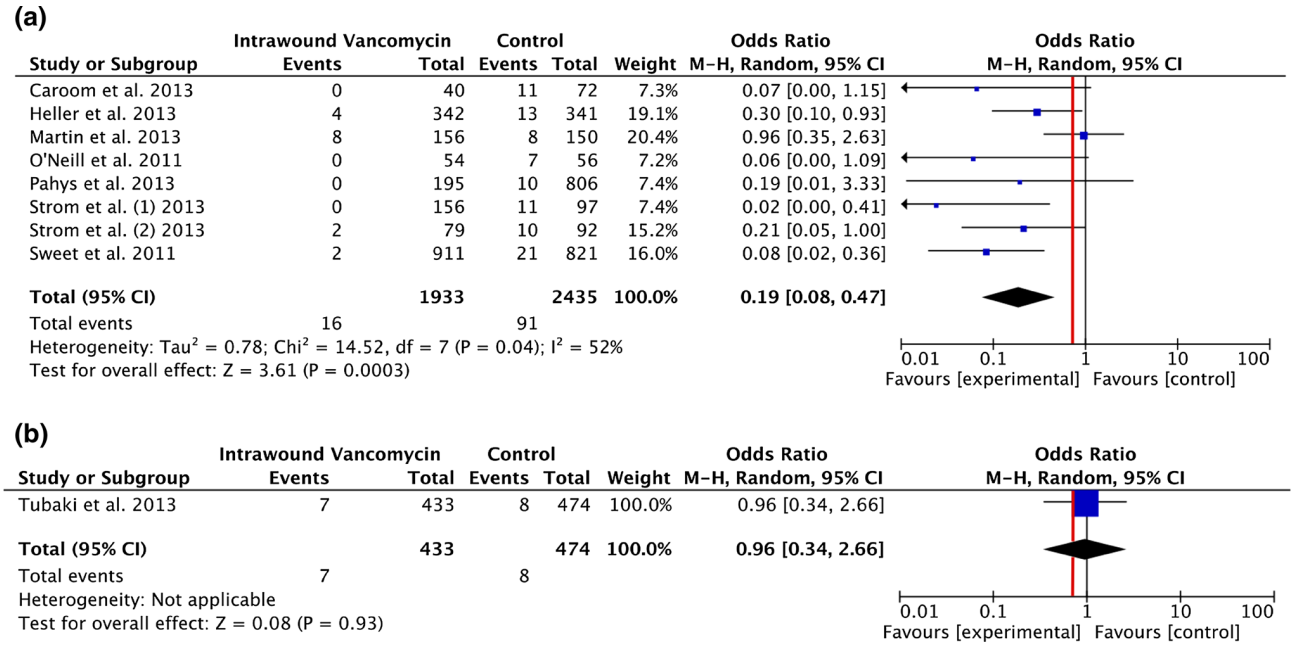


Fig. 3 Forest plots showing pooled estimates of the effect of intrawound vancomycin across **a** observational studies and **b** randomized controlled trials. Red lines project a zone of clinical equivalence based on a 25 % relative risk reduction and a control event rate of 3.7 % (OR = 0.74)

Table 3 GRADE summary of findings (pooled outcomes)

Outcomes	No. of participants (studies) Follow-up	Quality of the evidence (GRADE)	Anticipated effects Time frame is up to 3 years Odds ratio with intrawound vancomycin (95 % CI)
Infection prevention			
Randomized studies	907 (1 study) 12 months	⊕⊕⊖⊖ LOW due to risk of bias, imprecision	The odds ratio of infection in the intervention groups was 0.96 (0.34–2.66), which was not clinically significant[a]
Observational studies	4368 (8 studies) up to 3 years	⊕⊕⊖⊖ LOW due to study design	The odds ratio of infection in the intervention groups was 0.19 (0.08 to 0.47), which was clinically significant[a]
Adverse events			
All studies	5275 (9 studies) up to 3 years	⊕⊖⊖⊖ VERY LOW due to study design, risk of bias, inconsistency	All studies reported no adverse events attributable to intrawound vancomycin. There were no cases of renal failure or anaphylaxis. Sweet et al. reported 3 pseudarthroses in the intrawound vancomycin group ($n = 911$) compared to 4 in the control group ($n = 821$)

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

[a] Clinical equivalence was estimated assuming a 25 % risk reduction in infections [33] and the control event rate generated from the pooled observational data (3.7 %)

associated with very few outcome events (Table 3) [42, 43]. The evidence from the observational studies was rated as low quality by both reviewers on the basis of study design [35]. The overall evidence for the estimated rates of adverse events was initially considered low due to the majority of studies being observational, and was further

rated down to very low on the basis of risk of bias and imprecision. According to the GRADE approach, low-quality evidence indicates that our confidence in the effect estimate is limited and that the true effect may be substantially different from the estimate of the effect [35]. Very-low-quality evidence indicates that we have very little confidence in the effect estimate and the true effect is likely to be substantially different from the estimate of effect. Final assessments are presented in a GRADE summary of findings table with one row for each of the estimate of infection prevention according to randomized trials, the estimate of infection prevention according to observational studies, and the overall estimate of rates of adverse events.

Discussion

The primary objective of this systematic review and meta-analysis was to determine if the available evidence supports a benefit for intrawound vancomycin during spine surgery for the prevention of post-operative SSIs. Secondary objectives were to determine whether there are important differences in the estimates of treatment effect between observational studies and randomized trials and to summarize the adverse events associated with intrawound vancomycin reported in the literature. The pooled estimate from the eight observational studies indicated a statistically significant reduction in odds of infection with the use of intrawound vancomycin (OR 0.19, 95 % CI 0.08–0.47, $p = 0.0003$); however, the single randomized control trial failed to demonstrate any benefit (OR 0.96, 95 % CI 0.34–2.66, $p = 0.93$). Substantial study heterogeneity and recurrent methodological weaknesses limit the validity of these results and draw attention to the need for further high-quality research. All studies reported that there were no adverse events attributable to intrawound vancomycin.

Limitations

This study excluded at least two conference proceedings, both of which reported a beneficial effect for intrawound vancomycin [44, 45]. Their inclusion may have added to the body of evidence favoring intrawound vancomycin, but would have been unlikely to alter the major conclusions of this study. A detailed search of non-peer reviewed literature was not performed and it remains equally possible that abstracts for unpublished negative studies or forthcoming randomized trials were not identified. Two future trials are currently registered on *ClinicalTrials.gov*, but neither has completed recruitment [46, 47]. Sensitivity analyses to control for local co-interventions and surgical or patient risk factors for infection were not possible due to variable reporting and the small number of included studies. Likewise, the optimal dosage and application technique of intrawound vancomycin remain unknown. It is possible that dosing should account for patient weight or the size of their wound, and it has yet to be evaluated whether intrawound vancomycin has a role in anterior or lateral spine approaches [39, 40].

The zones of clinical equivalence were constructed using a 25 % risk reduction and a control event rate of 3.7 % pooled from the observational studies (control $n = 2435$). The 25 % relative risk reduction was based on the preferences of orthopedic trauma surgeons [33]. It may not be a valid estimate of the preferences of spine surgeons, but we are unaware of any similar data. Even if a smaller treatment effect was considered, the randomized data estimate is so imprecise that the conclusions of this meta-analysis would not have changed. Including the data from the single RCT (control infection rate 1.7 %, $n = 474$) in the estimate of control event rate would only minimally shift the zone of clinical equivalence and also not alter our findings. A large survey of spine surgeons should be considered to elicit the preferred absolute and relative risk reductions that would be considered clinically important in future trials.

Implications for practice

Surgical site infections are associated with substantial morbidity, mortality, and economic burden, and effective strategies to prevent their occurrence are paramount [1–4]. All included studies in this review reported no adverse effects attributable to intrawound vancomycin; however, safety data are insufficient. Rare and potentially devastating outcomes such as pseudarthrosis, irreversible renal toxicity, and life-threatening anaphylaxis may be possible effects of this intervention. Basic science studies have shown that vancomycin can inhibit osteoblasts in vitro [19, 48]. A recent case report described sudden and severe circulatory compromise following the application of intrawound vancomycin in a 52-year-old female undergoing vertebrectomy and reconstruction for metastatic bone disease [17]. Acute renal failure has been observed following the use of vancomycin-laden bone cement in total joint arthroplasty [18].

Implications for research

Evidence-based medicine is the process of judiciously integrating best available evidence with patient values and surgeon expertise to facilitate clinical decision making [49]. Although well-designed observational studies occasionally produce estimates of effect that are similar in direction and magnitude to randomized controlled trials [50, 51], observational studies with methodological

weaknesses frequently do not [52, 53]. The medical literature contains many examples in which clinicians were misled by observational studies, sometimes causing serious harm [54]. For example, antiarrhythmic drugs are estimated to have killed more Americans than those died in the Vietnam war [54–56]. Similar data exist for estrogen therapy thought to prevent coronary mortality [57–59], arthroscopic lavage thought to relive pain in osteoarthritic knees [60], and an intracranial-to-extracranial vascular bypass operation thought to prevent strokes [61, 62].

All of the observational studies of intrawound vancomycin in this review had significant methodological limitations. Non-consecutive recruitment may have caused selection bias, historical control groups may have been confounded by unstandardized co-interventions, unequal or unreported losses to follow-up may have cause transfer bias, and attempts to determine infection outcomes retrospectively with varying definitions may have been limited by recall bias. Participants, healthcare providers, outcome assessors, data analysts, and manuscript writers were not blinded in any of the studies, and the diagnoses of infection may have been biased by knowledge of which treatment was provided [63].

Patients in the control group of the single randomized trial received a second-generation cephalosporin (cefuroxime) as intravenous prophylaxis, while most of the patients in the control groups in the observational studies received a first-generation cephalosporin (cefazolin). Although antibiotic prophylaxis with cefuroxime is an accepted standard of care at many institutions, second-generation cephalosporins may be less effective against gram-positive bacteria in comparison to first-generation cephalosporins [64]. Patients in the randomized trial may have been at higher baseline risk for surgical site infection in comparison to patients in the observational studies. Likewise, the single randomized trial was conducted in India whereas all of the observational studies were conducted in the United States. The authors of the randomized trial did not report the manufacturer of the vancomycin that was used, but patients in India may be more likely to receive generic antibiotics [65]. Although generic vancomycin may have similar pharmaceutical bioavailability as the original drug, it may not be therapeutically equivalent [66]. Varied antibiotic regimens or other co-interventions may limit generalizability or contribute to study heterogeneity and should be standardized in future trials.

Conclusions

Based on the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) criteria [34], this systematic review and meta-analysis found a lack of high-quality evidence to inform the use of intrawound vancomycin in spine surgery. Whether this intra-operative technique reduces the risk of post-operative surgical site infections remains controversial. Given the potential risks for serious adverse events, surgeons should be cautious before widely adopting this intervention. They should also be vigilant in monitoring for harm. Further investigation with additional randomized controlled trials would improve the quality of evidence available to inform clinical practice.

Acknowledgments This study was not funded. Dr. Bhandari is funded, in part, by a Canada Research Chair.

Conflict of interest The authors certify that they, or a member of their immediate families, have no funding or commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

Ethical approval This study did not require local institutional ethics board approval.

References

1. Perencevich EN, Sands KE, Cosgrove SE, Guadagnoli E, Meara E, Platt R (2003) Health and economic impact of surgical site infections diagnosed after hospital discharge. Emerg Infect Dis 9(2):196–203
2. Kirkland KB, Briggs JP, Trivette SL, Wilkinson WE, Sexton DJ (1999) The impact of surgical-site infections in the 1990s: attributable mortality, excess length of hospitalization, and extra costs. Infect Control Hosp Epidemiol 20(11):725–730
3. Fry DE (2002) The economic costs of surgical site infection. Surg Infect (Larchmt) 3(Suppl 1):S37–S43
4. Sasso RC, Garrido BJ (2008) Postoperative spinal wound infections. J Am Acad Orthop Surg 16(6):330–337
5. ter Gunne AFP, Cohen DB (2009) Incidence, prevalence, and analysis of risk factors for surgical site infection following adult spinal surgery. Spine (Phila Pa 1976) 34(13):1422–1428
6. Gerometta A, Olaverri JCR, Bitan F (2012) Infections in spinal instrumentation. Int Orthop 36(2):457–464
7. Lazennec J-Y, Fourniols E, Lenoir T, Aubry A, Pissonnier ML, Issartel B, Rousseau MA, French Spine Surgery Society (2011) Infections in the operated spine: Update on risk management and therapeutic strategies. Orthop Traumatol Surg Res 97(6 SUPPL):S107–S116
8. Werner BC, Shen FH, Shimer AL (2011) Infections after lumbar spine surgery: avoidance and treatment. Semin Spine Surg 23(2):142–150
9. Savage JW, Anderson PA (2013) An update on modifiable factors to reduce the risk of surgical site infections. Spine J 13(9):1017–1029
10. Molinari RW, Khera OA, Molinari WJ 3rd (2012) Prophylactic intraoperative powdered vancomycin and postoperative deep spinal wound infection: 1,512 consecutive surgical cases over a 6-year period. Eur Spine J 21(Suppl 4):S476–S482
11. O'Neill K, Smith J, Abtahi A, Archer KR, Spengler DM, McGirt M, Devin CJ (2011) Reduced surgical site infections in patients undergoing posterior spinal stabilization of traumatic injuries using vancomycin powder. Spine J 11(7):641–646
12. Zebala LP, Chuntarapas T, Kelly MP, Talcott M, Greco S, Riew KD (2014) Intrawound vancomycin powder eradicates surgical

wound contamination: an in vivo rabbit study. J Bone Jt Surg Am 96(1):46–51
13. Vitale MG, Riedel MD, Glotzbecker MP, Matsumoto H, Roye DP, Akbarnia BA, Anderson RC, Brockmeyer DL, Emans JB, Erickson M, Flynn JM, Lenke LG, Lewis SJ, Luhmann SJ, McLeod LM, Newton PO, Nyquist AC, Richards BS 3rd, Shah SA, Skaggs DL, Smith JT, Sponseller PD, Sucato DJ, Zeller RD, Saiman L (2013) Building consensus: development of a best practice guideline (BPG) for surgical site infection (SSI) prevention in high-risk pediatric spine surgery. J Pediatr Orthop 33(5):471–478
14. Glotzbecker MP, Vitale MG, Shea KG, Flynn JM (2013) Surgeon practices regarding infection prevention for pediatric spinal surgery. J Pediatr Orthop 33(7):694–699
15. Gans I, Dormans JP, Spiegel DA, Flynn JM, Sankar WN, Campbell RM, Baldwin KD (2013) Adjunctive vancomycin powder in pediatric spine surgery is safe. Spine (Phila Pa 1976) 38:1703–1707
16. Godil SS, Parker SL, O'Neill KR, Devin CJ, McGirt MJ (2013) Comparative effectiveness and cost-benefit analysis of local application of vancomycin powder in posterior spinal fusion for spine trauma: clinical article. J Neurosurg Spine 19(3):331–335
17. Mariappan R, Manninen P, Massicotte EM, Bhatia A (2013) Circulatory collapse after topical application of vancomycin powder during spine surgery case report. J Neurosurg Spine 19(3):381–383
18. Patrick BN, Rivey MP, Allington DR (2006) Acute renal failure associated with vancomycin- and tobramycin-laden cement in total hip arthroplasty. Ann Pharmacother 40(11):2037–2042
19. Rathbone CR, Cross JD, Brown KV, Murray CK, Wenke JC (2011) Effect of various concentrations of antibiotics on osteogenic cell viability and activity. J Orthop Res 29(7):1070–1074
20. Tubaki VR, Rajasekaran S, Shetty AP (2013) Effects of using intravenous antibiotic only versus local intrawound vancomycin antibiotic powder application in addition to intravenous antibiotics on postoperative infection in spine surgery in 907 patients. Spine (Phila Pa 1976) 38(25):2149–2155
21. Martin JR, Adogwa O, Brown CR, Bagley CA, Richardson WJ, Lad SP, Kuchibhatla M, Gottfried ON (2013) Experience with intrawound vancomycin powder for spinal deformity surgery. Spine (Phila Pa 1976) 39:177–184
22. Chiang HY, Herwaldt LA, Blevins AE, Cho E, Schweizer ML (2014) Effectiveness of local vancomycin powder to decrease surgical site infections: a meta-analysis. Spine J 14(3):397–407
23. Chiang HY, Herwaldt L, Schweizer M (2014) Reply to "letter to the editor" by baker and chen regarding "effectiveness of local vancomycin powder to decrease surgical site infections: A meta-analysis". Spine J [Epub ahead of print]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 62(10):1006–1012
25. Higgins JPT, Greene S (2011) Cochrane handbook for systematic reviews of interventions. In: Higgins JPT and Greene S (eds). The cochrane collaboration, version 5.1.0 ed
26. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J (2003) Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 73(9):712–716
27. Sackett D, Haynes R, Guyatt G, Tugwell P (1991) Clinical epidemiology: a basic science for clinicians, 2nd edn. Little Brown, Boston
28. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174
29. Bhandari M, Morshed S, Tornetta P 3rd (2009) Schemitsch EH. Design, conduct, and interpretation of nonrandomized orthopaedic studies: a practical approach. (all) evidence matters. J Bone Jt Surg Am 91(Suppl 3):1
30. Bhandari M, Tornetta P 3rd, Schemitsch EH (2012) Randomized trials in orthopaedic surgery: one step closer. J Bone Jt Surg Am 94(Suppl 1):1
31. Hoppe DJ, Schemitsch EH, Morshed S, Tornetta P 3rd, Bhandari M (2009) Hierarchy of evidence: where observational studies fit in and why we need them. J Bone Jt Surg Am 91(Suppl 3):2–9
32. Guyatt GH, Sackett DL, Cook DJ (1993) Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-based medicine working group. JAMA 270(21):2598–2601
33. Petrisor B, Jeray K, Schemitsch E, Hanson B, Sprague S, Sanders D, Bhandari M, FLOW Investigators (2008) Fluid lavage in patients with open fracture wounds (FLOW): An international survey of 984 surgeons. BMC Musculoskelet Disord 9:7
34. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ (2011) GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64(4):383–394
35. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH (2011) GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64(4):401–406
36. Caroom C, Tullar JM, Benton EG Jr, Jones JR, Chaput CD (2013) Intrawound vancomycin powder reduces surgical site infections in posterior cervical fusion. Spine 38(14):1183–1187
37. Heller A, McIff TE, Lai SM, Burton DC (2013) Intrawound vancomycin powder decreases staphylococcal surgical site infections following posterior instrumented spinal arthrodesis. J Spinal Disord Tech [Epub ahead of print]
38. Pahys JM, Pahys JR, Cho SK, Kang MM, Zebala LP, Hawasli AH, Sweet FA, Lee DH, Riew KD (2013) Methods to decrease postoperative infections following posterior cervical spine surgery. J Bone Jt Surg Am 95(6):549–554
39. Strom RG, Pacione D, Kalhorn SP, Frempong-Boadu AK (2013) Lumbar laminectomy and fusion with routine local application of vancomycin powder: decreased infection rate in instrumented and non-instrumented cases. Clin Neurol Neurosurg 115(9):1766–1769
40. Strom RG, Pacione D, Kalhorn SP, Frempong-Boadu AK (2013) Decreased risk of wound infection after posterior cervical fusion with routine local application of vancomycin powder. Spine 38(12):991–994
41. Sweet FA, Roh M, Sliva C (2011) Intrawound application of vancomycin for prophylaxis in instrumented thoracolumbar fusions: efficacy, drug levels, and patient outcomes. Spine 36(24):2084–2088
42. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, Thabane L, Walter SD, Pogue J, Devereaux PJ (2014) The statistical significance of randomized controlled trial results is frequently fragile: A case for a fragility index. J Clin Epidemiol 67(6):622–628
43. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ (2011) GRADE guidelines 6. Rating the quality of evidence–imprecision. J Clin Epidemiol 64(12):1283–1293
44. Caroom C, Chaput CD, Tullar J, Jones J (2012) Reduced surgical site infections in patients with cervical spondylotic myelopathy undergoing posterior cervical instrumented fusion using vancomycin powder and chlorhexadine skin preparation. Spine J. Conference: 27th Annual Meeting of the North American Spine Society, NASS 2012 Dallas, TX United States. Conference Start:

20121024 Conference End: 20121027. Conference Publication: (var.pagings) 12(9 SUPPL. 1):80S
45. Dewan MC, Godil SS, Zuckerman SL, Mendenhall SK, Shau DN, Parker SL, Devin CJ, McGirt MJ (2013) Comparative effectiveness and cost-benefit analysis of topical vancomycin powder in posterior spinal fusion for spine trauma and degenerative spine disease. Spine J. Conference: 28th Annual Meeting of the North American Spine Society, NASS 2013 New Orleans, LA United States. Conference Start: 20131009 Conference End: 20131012. Conference Publication: (var.pagings). 13(9 SUPPL. 1):56S
46. Vancomycin in spine surgery [Internet]. 2013. Available from: http://clinicaltrials.gov/ct2/show/NCT01977989
47. Comparative effectiveness and cost-benefit analysis of vancomycin powder in high risk spine surgery patients [Internet]. 2013. Available from: http://clinicaltrials.gov/ct2/show/NCT01566422
48. Edin ML, Miclau T, Lester GE, Lindsey RW, Dahners LE (1996) Effect of cefazolin and vancomycin on osteoblasts in vitro. Clin Orthop Relat Res (333):245–251
49. Bhandari M, Giannoudis PV (2006) Evidence-based medicine: what it is and what it is not. Injury 37(4):302–306
50. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials. N Engl J Med 342(25):1878–1886
51. Concato J, Shah N, Horwitz RI (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 342(25):1887–1892
52. Kunz R, Oxman AD (1998) The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 317(7167):1185–1190
53. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J (2001) Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 286(7):821–830
54. Devereaux PJ, Yusuf S (2003) The evolution of the randomized controlled trial and its role in evidence-based decision making. J Intern Med 254(2):105–113
55. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH et al (1991) Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The cardiac arrhythmia suppression trial. N Engl J Med 324(12):781–788
56. Moore TJ (1995) Excess mortality estimates: dearly medicine: why tens of thousands of heart patients died in America's worst drug disaster. Simon & Shuster, New York
57. Stampfer MJ, Colditz GA (1991) Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. Prev Med 20(1):47–63
58. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E (1998) Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. heart and estrogen/progestin replacement study (HERS) research group. JAMA 280(7):605–613
59. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. JAMA 288(3):321–333
60. Reichenbach S, Rutjes AW, Nuesch E, Trelle S, Juni P (2010) Joint lavage for osteoarthritis of the knee. Cochrane Database Syst Rev 12(5):CD007320
61. Popp AJ, Chater N (1977) Extracranial to intracranial vascular anastomosis for occlusive cerebrovascular disease: experience in 110 patients. Surgery 82(5):648–654
62. The EC/IC Bypass Study Group (1985) Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. N Engl J Med 313(19):1191–1200
63. Dowrick AS, Bhandari M (2012) Ethical issues in the design of randomized trials: to sham or not to sham. J Bone Jt Surg Am 18(94 Suppl 1):7–10
64. Petignat C, Francioli P, Harbarth S, Regli L, Porchet F, Reverdin A, Rilliet B, de Tribolet N, Pannatier A, Pittet D, Zanetti G (2008) Cefuroxime prophylaxis is effective in noninstrumented spine surgery: a double-blind, placebo-controlled study. Spine (Phila Pa 1976) 33(18):1919–1924
65. Venkatesh M, Bairavi VG, Sasikumar KC (2011) Generic antibiotic industries: challenges and implied strategies with regulatory perspectives. J Pharm Bioallied Sci 3(1):101–108
66. Vesga O, Agudelo M, Salazar BE, Rodriguez CA, Zuluaga AF (2010) Generic vancomycin products fail in vivo despite being pharmaceutical equivalents of the innovator. Antimicrob Agents Chemother 54(8):3271–3279

DISCUSSION

This thesis aimed to investigate and apply modern innovations in health research methodology to the field of spine surgery. It consists of seven chapters that addressed topics related to randomized controlled trials, observational studies, and systematic reviews and meta-analyses. By applying the findings of each chapter, clinicians, researchers, and other evidence users can advance the credibility of future research and enhance the care of patients with spinal disorders.

Chapter 1 presented a systematic survey to determine the robustness of statistically significant results from RCTs of spine surgery [17]. The Fragility Index is a novel metric to inform about the robustness of statistically significant results. This chapter was the first to report Fragility Index values after Walsh et al.'s original paper and the first to consider fragility of RCTs from a surgical specialty. Our findings will guide clinicians to exercise caution when interpreting RCTs with low Fragility Index values, but widespread adoption of the Fragility Index is necessary to practically help identify trials with less robust results and inform the potential importance of missing data. Future studies could focus on translating our findings to clinicians, researchers, reviewers, and editors.

We encountered several challenges understanding the practical utility of the Fragility Index. Clinicians can be confident that trials with increasing Fragility Index values are more robust than trials with lower values and that Fragility Index values ought to exceed numbers of participants lost to follow-up, but further work is necessary to determine whether certain thresholds of acceptability could be routinely adopted. Likewise, we discussed that investigators should recognize that both numbers of participants and numbers of events are relevant to power of their trials, but their ability to guard against low Fragility Index values in advance is limited because even well powered trials are at risk. Therefore, we concluded that the concept of fragility comes into play only at the point of examining results, rather than at the stage of sample size calculation.

Chapter 2 presented a Users' Guide to randomized controlled trials that address surgical or other non-pharmacological therapies [15]. The chapter is applicable to spine surgery even many of our examples though most of examples came from fields other because remote randomization systems, blinding, sham-controlled trials, split-body trials, expertise-based trials, and mechanistic versus practical trials are all critical to the design, conduct, and interpretation of trials of spine surgery interventions.

Chapter 3 presented a propensity score-matched cohort study to evaluate methylprednisolone for the treatment of patients with acute spinal cord injuries [18]. Our regression analyses may have been underpowered to detect significant associations due to high numbers of variables in each model and a relatively small sample size. Conventional rules of thumb recommend at least 10 observations per variable for linear regression and at least 10 events per variable for logistic regression [21]. We attempted to overcome this limitation by including multiple sensitivity analyses, the results of which were consistent with our primary analysis. We could have used iterative methods of model development to reduce numbers of variables, but these approaches risked ignoring our underlying clinical rationale. We were unable to control for potential treatment biases at each site because more than half of the patients who received methylprednisolone did so before arriving at a participating site.

Chapter 4 presented a scoping review on the surgical management of scoliosis [13]. In their framework for scoping review methodology, Arksey and O'Malley suggested engaging diverse groups of multi-disciplinary knowledge users in order to understand practical needs and guide knowledge translation [22,23]. However, we developed our thematic framework *ad hoc* according to the expertise of two reviewers and our conclusions followed from only a descriptive analysis of the literature.

Although this study effectively identified important knowledge gaps and provides direction for future research, additional work is needed to directly involve knowledge users in the design and conduct of future studies.

Chapter 5 presented strategies to improve the credibility of meta-analyses in spine surgery, most of which were applied successfully in chapters 6 and 7 [12,19,20]. However, chapter 6 did not adequately present pooled estimates in terms that were ready for clinical application and chapter 7 did not adequately explore heterogeneity among the observational studies. Neither limitation substantially affected confidence in the pooled effect estimates or materially affected our conclusions. Chapters 6 and 7 both advanced current understanding by considering estimates of effect from observational studies and randomized controlled trial separately, including data not previously pooled, and incorporating the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to evaluate confidence and summarize findings.

More broadly, this thesis was limited to several specific health research methodology issues in spine surgery. There are many methodological issues and clinical questions that were not addressed, including for example those that relate to diagnostic and prognostic study designs, economic evaluations and health technology assessments, and guidelines

development. These areas were beyond the scope of this thesis, but their exclusion does not limit the validity or importance of the work that was performed. In combination with the methodological issues addressed herein, I am confident that these areas will form the underpinnings of an impactful career applying health research methodology and conducting clinical research in spine surgery.

LIST OF REFERENCES

[1] Rampersaud YR, Lewis SJ, Davey JR, Gandhi R, Mahomed NN. Comparative outcomes and cost-utility after surgical treatment of focal lumbar spinal stenosis compared with osteoarthritis of the hip or knee--part 1: long-term change in health-related quality of life. Spine J 2014;14:234–43.

[2] Singh A, Tetreault L, Kalsi-Ryan S, Nouri A, Fehlings MG. Global prevalence and incidence of traumatic spinal cord injury. Clin Epidemiol 2014;6:309–31.

[3] Pannell WC, Savin DD, Scott TP, Wang JC, Daubs MD. Trends in the surgical treatment of lumbar spine disease in the United States. Spine J 2015;15:1719–27.

[4] Yoshihara H, Yoneoka D. National trends in the surgical treatment for lumbar degenerative disc disease: United States, 2000 to 2009. Spine J 2015;15:265–71.

[5] Noonan VK, Fingas M, Farry A, Baxter D, Singh A, Fehlings MG, et al. Incidence and prevalence of spinal cord injury in Canada: a national perspective. Neuroepidemiology 2012;38:219–26.

[6] Guyatt G, Rennie D, Meade MO, Cook DJ. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 3E. 3 edition. New York: McGraw-Hill Professional; 2014.

[7] Bhandari M, Tornetta P, Schemitsch EH. Randomized trials in orthopaedic surgery: one step closer. J Bone Joint Surg Am 2012;94 Suppl 1:1.

[8] Fisher CG, Wood KB. Introduction to and techniques of evidence-based medicine. Spine 2007;32:S66–72.

[9] Street J, Lenehan B, Fisher C. The quality of quality of life publications in the spinal literature: are we getting any better? J Neurosurg Spine 2009;11:512–7.

[10] Jacobs WCH, Kruyt MC, Verbout AJ, Oner FC. Spine surgery research: on and beyond current strategies. Spine J 2012;12:706–13.

[11] Dijkman BG, Abouali JAK, Kooistra BW, Conter HJ, Poolman RW, Kulkarni AV, et al. Twenty Years of Meta-Analyses in Orthopaedic Surgery: Has Quality Kept Up with Quantity? J Bone Joint Surg Am 2010;92:48–57.

[12] Evaniew N, van der Watt L, Bhandari M, Ghert M, Aleem I, Drew B, et al. Strategies to improve the credibility of meta-analyses in spine surgery: A systematic survey. Spine J 2015 15(9):2066-76.

[13] Evaniew N, Devji T, Drew B, Peterson D, Ghert M, Bhandari M. The surgical management of scoliosis: a scoping review of the literature. Scoliosis 2015;10:1.

[14] Aebi M, Szpalski M. The growth and increase of the number of journals focusing on Spine: a scientific blessing? Eur Spine J 2014;23:491–2.
[15] Evaniew N, Carrasco-Labra A, Devereaux PJ, Tikkinen KAO, Fei Y, Bhandari M, et al. How to use a randomized controlled trial addressing a surgical or other non-pharmacological therapy: Users' Guides to the medical literature. Submitted June 5, 2015 to JAMA
[16] Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. J Bone Joint Surg Am 2001;83-A:916–26.
[17] Evaniew N, Files C, Smith C, Bhandari M, Ghert M, Walsh M, et al. The fragility of statistically significant findings from randomized trials in spine surgery: A systematic survey. Spine J 2015;15:2188–97.
[18] Evaniew N, Noonan V, Fallah N, Kwon BK, Rivers CS, Ahn H, et al. Methylprednisolone for the treatment of patients with acute spinal cord injuries: A propensity score-matched cohort study from a Canadian multicenter spinal cord injury registry. J Neurotrauma 2015;32:1674–83.
[19] Evaniew N, Khan M, Drew B, Peterson D, Bhandari M, Ghert M. Intrawound vancomycin to prevent infections after spine surgery: a systematic review and meta-analysis. Eur Spine J 2015;24:533–42.
[20] Evaniew N, Belley-Côté E, Fallah N, Noonan VK, Rivers CS, Dvorak MF. Methylprednisolone for the treatment of patients with acute spinal cord injuries: A systematic review and meta-analysis. Journal of Neurotrauma 2015 Nov 3 [epub ahead of print].
[21] Kleinbaum PDG, Kupper PLL, Nizam APA, Muller PKE. Applied Regression Analysis and Other Multivariable Methods. 4 edition. Australia ; Belmont, CA: Duxbury Press; 2007.
[22] Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci IS 2010;5:69.
[23] Slobogean GP, Johal H, Lefaivre KA, MacIntyre NJ, Sprague S, Scott T, et al. A scoping review of the proximal humerus fracture literature. BMC Musculoskelet Disord 2015;16:112.