University of Alberta

A critical evaluation of evidence-based emergency medicine

by

©

Maria Beatriz Ospina

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Master of Science

in

Medical Sciences - Public Health Sciences

Edmonton, Alberta

Fall 2004

Your file Votre référence
ISBN: 0-612-95829-9
Our file Notre référence
ISBN: 0-612-95829-9

Abstract

Objectives: An evaluation of emergency medicine (EM) research was completed by: exploring the scientific quality of systematic reviews (SR) published in EM journals since the introduction of the QUOROM guidelines and evaluating publication rates of meeting abstracts. **Methods:** SRs published from 2000-2004 were compared to reviews published from 1988-1998 in 5 key EM journals. Publication rate, time to publication and the impact of publication bias were evaluated for all randomized clinical trial abstracts presented at Society of Academic Emergency Medicine (SAEM) meetings (1995-2003). **Results:** The quality of EM SRs is unacceptably low, and reviews have not improved since the QUOROM was published. Abstracts are published at a slightly lower rate than other specialty societies; however, commonly reported biases were not identified in the EM literature. **Conclusions:** Researchers, authors and readers must be cautious when interpreting the evidence from published EM literature; efforts to improve the quality of evidence in emergency medicine should continue.

Acknowledgements

I always considered it a bit trite of thesis authors to include a lengthy list of people whose support was "invaluable". That is, of course, until I wrote this thesis and realized just how deeply I am in debt to the many people that have been a part of my graduate education, as friends, teachers, and colleagues.

I would like to gratefully acknowledge the enthusiastic supervision of Dr. Brian Rowe, the best supervisor and teacher I could have wished for. He has the patent rights on how to be an excellent human being, clinician, researcher, psychologist (in the bad times), and soccer coach. His inspiration, his great efforts to explain things both clearly and simply, as well as his grasp of evidence based medicine at all levels, helped me develop my own understanding. I would also like to acknowledge my professors at the University of Alberta who have contributed much to my education, and the members of my Thesis Committee, Dr. Terry Klassen, and Dr. Karen Kelly. Special thanks to Ms.Sandra Blitz from the Emergency Medicine Research Office, and Natasha Wiebe from the Alberta Centre for Child Health Evidence at the University of Alberta for the statistical advice.

The assistance of the editorial boards of *Academic Emergency Medicine* (Ann Arbour, Michigan) and *Emergency Medicine Journal* (London, UK) in facilitating the electronic access to their collections is graciously acknowledged.

I am very grateful to the Alberta Center for Injury Control and Research and the Evidence Practice Center at the University of Alberta for the studentship awards which supported me during my first and second years of graduate studies, respectively. My gratitude to the people from Academic Technologies for Learning at the University of Alberta, who hired me as GRA during my entire cycle of graduate studies: Stanley Vernhagen and Elizabeth Hanlis for their confidence in me and their generous support in terms of very flexible work hours.

As friendship does not know of time or distance, I have been lucky enough to have the continued support of many good friends overseas. Special thanks to my friends forever Claudia Irene, Heidi, Gloria, Carlos, Monica, Tatiana, Eliana, and Fer, for their support and encouragement through the distance. Life would not have been the same without the "pizzas team" (Rosi, Pepin, Nati, and Er). I am grateful to them for being my surrogate family during my years in Edmonton, for helping me get through the difficult times, and for all the emotional support, and camaraderie.

I want to express my gratitude to my student colleagues for providing a stimulating and fun environment in which to learn and grow. Special thanks to Barry Diner, Ritu Rani, Leyla Bagirzadeh, Chunu Mainali, Zhiwei Gao, Wensheng Zhang, and Brad Johnston.

The road to my graduate degree has been long and winding, so I would also like to thank some people from my early days here in Canada. Don Juzwishin, Christa Harstall, Bing Guo, Paula Corabian, Liza Chan, Leigh-Ann Topfler from AHFMR were among those who kept me going at the beginning of my immersion in the Canadian research culture.

Finally, I am forever indebted to those closest to me, whose presence (material and immaterial) helped make the completion of my graduate work possible. Paul Garcia, my rock, my best friend for the past 3 years, and the most gifted, reliable and loyal person I ever met. Thanks for your love, patience, and understanding (and your invaluable collaboration reading tons of abstracts and papers!!). Most of all, I would like to thank my family, Carlos E, Juanca, Luces, Ia, Lia, Dani, Alicia, Emilio and Thor, and especially my parents, Carlos y Rosita, for their endless love and absolute confidence in me. The knowledge that they will always be there to pick up the pieces is what allows me to repeatedly risk getting shattered. This thesis would certainly not have existed without them. I owe them everything I am and it is to them that I dedicate this work.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AEM	Academic Emergency Medicine
AJEM	American Journal of Emergency Medicine
AOEM	Annals of Emergency Medicine
BR	Brian Rowe
CCT	Controlled clinical trials
CI	Confidence interval
CJEM	Canadian Journal of Emergency Medicine
CONSORT	Consolidated Standards of Reporting Trials
DARE	Database of Abstracts of Reviews of Effects
df	Degrees of freedom
EBEM	Evidence-based emergency medicine
EBM	Evidence Based Medicine
EM	Emergency medicine
EMJ	Emergency Medicine Journal
HR	Hazard ratio
ICC	Intra class correlation coefficient
IF	Impact factor
IQR	Inter quartile range
JEM	Journal of Emergency Medicine
κ	Kappa
MA	Meta-analysis
MCID	Minimal clinically important difference
MO	Maria Ospina
OQAQ	Overview Quality Assessment Questionnaire
OR	Odds Ratio
PG	Paul Garcia
QUOROM	Quality of reporting of meta-analyses

RCT	Randomized controlled trials
SAEM	Society for Academic Emergency Medicine
SCI	Science Citation Index
SR	Systematic review

CHAPTER 1: Introduction

"Those who are enamoured of practice without science are like a pilot who goes into a ship without rudder or compass and never has any certainty where he is going. Practice should always be based upon a sound knowledge of theory, of which perspective is the guide and gateway, and without it nothing can be done well in any kind of painting."
Leonardo da Vinci, Notebooks, 1508-1518.

1.1. Overview

Evidence-based medicine (EBM) was first defined about twelve years ago [1] as the integration of the best research evidence with clinical expertise and patient values into clinical decision-making [2,3]. This definition of EBM reflects a *model for thinking* (a shift of scientific paradigm for many [4]) on how to teach, learn and practice medicine and the allied health sciences. It is also about the *translation* of research findings into clinical practice to more effectively make clinical decisions for patients' care [2,5]. The hallmark of the evidence-based practice is that, for any particular clinical decision, the strength of the evidence, and therefore the degree of certainty (or uncertainty) can be known. That is to say, EBM is not an exercise for academicians or researchers; it is not about learning to do research solely, but a conceptual advance to facilitate the incorporation of medical innovation into the multi-faceted clinical decision making process.

There has been considerable interest and debate about the value EBM has on medical practice and the quality of patient health care. Cohen et al.[6] compiled and analyzed the criticisms EBM has received over the last 10 years. They found that the debate tends to revolve around five recurring themes: reliance on empiricism, narrow definition of evidence, lack of evidence of efficacy, limited usefulness for individual patients, and threats to the autonomy of the doctor/patient relationship [6]. They conclude that, although imperfect in many senses, EBM is widely recognized as an undeniable advance in guiding health care decisions and facilitating the discussion of health care issues with patients, which must continue to evolve in various areas of academic medicine [6]. Newer approaches to EBM [7] incorporate an intuitive component that turns the decision-making process into a more context-sensitive reasoning process, which is independent of mere cause-and-effect logic considerations.

Emergency medicine (EM) is a relatively new medical specialty based on the knowledge and skills required for the prevention, diagnosis and management of acute and urgent aspects of illness and injury affecting patients of all age groups with a full spectrum of episodic undifferentiated physical and behavioural disorders. It further encompasses an understanding of the development of pre-hospital and in-hospital emergency medical systems and the skills necessary for their functioning [8]. Finally, it is a frenetic and uncontrolled clinical setting where patient volume and acuity are largely unpredictable and where just-in-time evidence is particularly needed.

Emergency medicine has grown rapidly and expanded within the hospital systems in Canada and worldwide, demanding the use of the strongest scientific evidence to guide clinical decisions and practice [9]. The specialty of emergency medicine has also evolved as a respected academic area. From the constitution of the first periodical scientific publication in emergency medicine in 1969 (*Emergency medicine*; Quadrant Healthcom, USA) to the establishment of the most recent emergency medicine publication in 2004 (*Emergency Medicine Australasia*; Blackwell Publications, Australia), approximately 38 indexed peer-reviewed emergency medicine journals have disseminated a dynamic body of research, educational and medical advances in the field.

Evidence-based emergency medicine is a growing movement within the discipline of emergency medicine. The adoption of this approach has, however, involved a big challenge: A call for excellence and higher standards of scientific evidence in emergency medicine research to be applied for the care of patients in pre-hospital and emergency care settings [9]. By endorsing EBM principles and methods, emergency medicine researchers are committed to improve the quality of the evidence produced, particularly when therapeutic or preventive interventions are considered.

Rules of evidence have been established to classify the grade of evidence upon which strength of recommendations can be developed [10]. According to these, randomized controlled trials (RCTs) and systematic reviews have been broadly recognised as the strongest methodological approaches for research questions related to the effects of treatments or interventions. Alternatively, the prospective cohort has been broadly recognized as the strongest methodological approach for research questions related to the diagnosis of diseases in health care.

The RCT is considered the strongest design for a clinical study on the efficacy/effectiveness of a health care intervention. Generation of a randomization sequence to assign participants to treatment groups is the first crucial element for a RCT by ensuring that the patients in each group do not differ in known and potentially important prognostic variables, except for the treatment under investigation. Implementation of the randomization sequence, while concealing it at least until participants have been assigned to the treatment groups (allocation concealment), is the important second element, without which, randomization collapses in a trial [11].

Alternatively, systematic reviews summarize, analyze and report the combined results of a number of RCTs (or cohort studies in diagnostic SRs), with the "unit of analysis" being the individual study rather than the individual patient. Meta-analyses in particular, can increase the power and precision of treatment effect estimates.

These rules of evidence for therapeutic interventions classify studies based on the empirical power of the research design *per se*. Nevertheless, the evidence produced by the individual studies still needs to be assessed for strengths and weaknesses in terms of its validity (or methodological quality). Methodological quality can be defined as the extent to which the design and conduct of a study are likely to have prevented systematic errors (bias) which would threaten the meaningfulness of the research process and distort the estimation of the effect measures and the conclusions of the study [12].

The objective of this thesis is to complete a critical evaluation of the methodological quality of emergency medicine research produced in recent years under the form of systematic reviews and RCTs. This thesis is presented in a paper format, with chapters written as stand-alone manuscripts with the intent that one publication per chapter will be submitted. For each theme, a structured review of the available scientific literature on the topic under study will be presented as an orientation to the field of study. The study methodology will then be described in terms of the research objectives, research design, selection criteria, evaluation of the methodological quality, and data collection and analysis procedures. Finally, results are presented, interpreted and discussed in light of the results of similar studies.

Chapter 2 will present an overview of methodological quality of reports of systematic reviews relevant to a variety of biomedical areas. Chapter 3 will focus on the methodological quality of reports of systematic reviews published in major emergency medicine journals. The evolution of scientific methods and quality of a cohort of systematic reviews published from January 2000 to March 2004 will be compared to a cohort of systematic reviews published during the period of 1988-1998. The relationship between methodological quality and other characteristics of the systematic reviews published in major emergency medicine journals will also be explored. Finally, an evaluation of potential prognostic variables that may predict the quality of systematic reviews in the emergency medicine literature will be presented.

Chapter 4 will focus on the issue of publication bias in emergency medicine research under the form of RCTs. Publication rates of RCTs abstracts presented at a major emergency medicine scientific annual meeting from 1995 to 2003 will be identified. Abstract characteristics associated with subsequent publication and whether different types of publication bias affected CCT/RCTs in emergency medicine research (i.e. positive results bias, time-lag bias, grey literature bias, and place of publication bias) will be also evaluated.

In Chapter 5, the results of previous chapters are summarized and discussed in the context of existing research. General conclusions are provided as well as an evaluation of the research methodology employed for this study. Implications for practice and research in emergency medicine are discussed and recommendations for future research are made. This approach provides an opportunity to review the literature of both abstract publication bias as well as reviews of reviews. It further permits an evaluation of the changes in EM systematic reviews since the publication of the QUOROM statement (Quality of Reporting of Meta-analyses) and will also provide the first evaluation of publication bias of RCTs ever reported in EM.

1.2. References

(1) Guyatt GH. Evidence-based medicine. ACP J Club 1991; 114:A-16.

(2) Sackett DL, Rosenberg WMC, Gray Jam, Haynes RB, Richardson WS. Evidence-based medicine: what it is and what it isn't. BMJ 1996; 312:71-72.

(3) Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 2nd Ed. New York: 2000.

(4) Guyatt GH, Haines B, Jaeschke D, Cook D, Greenhalgh T, Meade M et al. Introduction: The Philosophy of Evidence-Based Medicine. In: Guyatt G, Rennie D, editors. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. Chicago: AMA Press, 2002.

(5) Haynes RB, Hayward RSA, Jadad AR, Sebaldt AR. Evidence-based health informatics: an overview of the Health Information Research Unit at McMaster University. Health Serv 1996; 5:41-44.

(6) Cohen AM, Stavri PZ, Hersh WR. A categorization and analysis of the criticisms of Evidence-Based Medicine. Int J Med Inf 2004; 73:35-43.

(7) Greenhalgh T. Intuition and evidence--uneasy bedfellows? Br J Gen Pract 2002; 52:395-400.

(8) American College of Emergency Physicians. Model of the clinical practice of emergency medicine [policy statement] 2000.

(9) Wyer PC, Rowe BH, Guyatt GH, Cordell WH. The clinician and the medical literature: when can we take a shortcut? Ann Emerg Med 2000; 36:149-155.

(10) Users' guides to the medical literature: A manual for evidence-based clinical practice. 4th ed. Chicago: AMA Press, 2002.

(11) Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002; 359:614-618.

(12) Alderson P, Green S, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003]. In: The Cochrane Library, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

CHAPTER 2: Overview of reports on the quality of systematic reviews

2.1. Introduction

Evidence-based medicine is a growing movement within the discipline of emergency medicine. The adoption of this approach has involved a call for excellence and higher standards of scientific evidence in emergency medicine research [1]. Typical time constraints of the emergency medicine practitioner, however, hinder a critical review of all the rapidly expanding medical literature. As it is unlikely that the state of scientific knowledge will ever be reflected in the results of a single study, a proper appraisal of the most up-to-date evidence requires an inclusive search of the available research [2]. The use of systematic approaches to identify, assess, and synthesize the findings of contemporary research is one means of addressing this problem [3,4]. In the event of a large collection of research on a single topic, systematic reviews may be the most appropriate summary evidence available to the busy clinician. This method of aggregating different and often seemingly conflicting results from similar studies (i.e., randomized, controlled clinical trials) has achieved an important role in providing useful information to inform clinical decisions [5,6].

Systematic reviews have been defined as reviews of the evidence that, in their ideal form, address a focused formulated question by using a structured, a priori, well-defined systematic approach and explicit methods to identify, select, critically appraise and combine the results of relevant primary research [7,9]. As retrospective and observational studies, systematic reviews are inherently open to systematic and random errors [8,9]. Compared to narrative reviews, systematic reviews use a more structured and rigorous approach to minimize bias and random error in the identification, collection, appraisal, and interpretation of relevant primary studies, thus providing more reliable results upon which to draw conclusions and make decisions [10]. Among the scientific strategies to limit the effects of systematic and random errors on the results of the review, investigators develop explicit eligibility criteria in terms of the patient population, the specific intervention or exposure, the comparison groups, the outcome measures and the range of acceptable studies. A comprehensive search strategy to identify all the published and unpublished literature relevant for the research question is designed, and an evaluation of the methodological quality of studies for inclusion is provided [11]. The goal is to render the review process transparent and capable of being reproduced.

Systematic reviews may also include a quantitative synthesis of individual results (i.e., meta-analysis) to produce a pooled estimate of the outcome of interest for a particular intervention. They may also incorporate non-quantitative appraisals, evidence tables and analysis of individual studies as well as other sources of evidence [12,13].

When the process is conducted in a methodological and comprehensive manner and the results interpreted in a clinically relevant way, systematic reviews can help summarize existing information on a specific topic and clarify the limits of information, providing a basis for rational decision making in clinical practice and public health medicine [14]. They can also pinpoint areas in which future research is needed and guide allocation of resources [15].

While the discipline of systematic reviews has evolved (with a 500-fold increase in the number of published meta-analyses in the past decade [16]), it became apparent that not all systematic reviews are created using the same rigorous methods. Like any other form of research, variations in methodological quality associated with systematic reviews raise questions regarding the validity and relevance of their findings [17,18]. The value of a systematic review largely depends on the quality of the review process itself and the scientific methods that have been employed to minimize random error and bias. On the other hand, the validity of their results also depends on the quality of the trials: if the raw material is flawed, then the systematic review conclusions will be equally weak. What is important is not simply that systematic reviews are published, but, rather, how good those reviews are, and how well they inform [19]. As Oxman and colleagues have said, "the fact that a review article is published in a peer-reviewed journal, even a prestigious one, is no guarantee of scientific quality" [3].

Systematic reviews are retrospective studies that are not immune to bias. The process requires a host of judgments regarding the search of information, the selection of individual studies for evaluation, the weight assigned to each study, the methods for combining them and, the conclusions reached. Each of these judgments requires assumptions, and each assumption can introduce bias [20].

Other methodological shortcomings in conducting and reporting systematic reviews that may introduce bias and compromise conclusions include the absence of clearly defined inclusion and exclusion criteria, inadequate searches for relevant studies, lack of assessment of the methodological quality of the individual studies, and lack of investigation of sources of heterogeneity in the findings [21-24]. For example, a thorough literature search is important to avoid biases in the identification and selection of studies. In particular, studies with positive results are easier to locate and are more likely to be reported in English language journals and to appear in journals that are indexed on large bibliographic databases. Those with negative results, on the other hand, are more likely not to reach publication [25].

Systematic reviews may sometimes have little control over random errors; however, the greater the scientific quality (i.e. methodological rigor) of the review, the greater the chance of minimizing the effects of bias in the review process [26]. There are obvious advantages of improving the quality of systematic reviews. First, authors could use previous reviews as starting points for future research, rather than as points of contention. Clinical decisions, public health policies and future research would also be better informed. Finally, high quality systematic reviews make a more efficient use of existing data. These issues highlight the importance of assessing the methodological quality of systematic reviews and, where appropriate, basing practice or policy only on those that are methodologically sound.

Evaluating the quality of systematic reviews conducted in the past and the identification of their typical shortcomings is a necessary step to improve future reviews. Several publications have examined the methodological quality of reports of systematic reviews and meta-analysis in a broad variety of medical areas. The goal of this section is to review and summarize the evidence for quality in systematic reviews and/or meta-analyses in the published literature.

2.2 Methodology

This is a retrospective study of overviews of systematic reviews and/or meta-analyses published up to May 2004 in the scientific literature. A literature search in MEDLINE via Ovid (1966- May 2004) was conducted to identify all the relevant references (See Appendix A.1 for the search strategy). Reference lists of relevant articles were also examined to identify further studies. No grey literature sources were explored. Hard copies of all potentially relevant articles were retrieved. One investigator (MO) compared each study against the following selection criteria to judge their eligibility: Overviews analyzing the methodological quality of systematic reviews and meta-analysis in any area of academic medicine and health care. No language restrictions were imposed. Letters, editorials, historical reviews, and health technology assessment reports were excluded. Data on the following characteristics of the studies were recorded: Topic of study, number of systematic reviews and/or meta-analyses considered in the overview, language restrictions set in the search for reviews, years covered in the searches, median year of publication of the systematic reviews and/or meta-analysis under study, method of quality assessment, and level of quality of systematic reviews and/or meta-analysis according to author's conclusions. Data are reported using descriptive statistics.

2.3 Results:

2.3.1 Descriptive information:

Nine hundred and seventy-eight citations were identified with the search strategy (Figure 2.1). After the screening of titles and abstracts, 58 references were potentially eligible for inclusion in the review. Thirty-six overviews published in 37 publications analyzing the methodological quality of systematic reviews and meta-analyses in health care research were included in this literature review. The reports were published from 1987 to 2003 (Table 2.1) and included on average 67.64 (95%CI: 40.53 to 94.7; range: 5, 480) reviews and/or meta-analyses per report. The grouped median year of publication of these reviews and/or meta-analyses was 1995 (min. 1982, max.1999, n = 30 overviews providing data for this calculation). A considerable proportion (16, 44%) of overviews restricted the search of reviews and/or meta-analyses to the English-literature. Only five (14%) of them declared no language restrictions in the strategies they used to identify the studies for inclusion. The overviews addressed a wide variety of topics and specialties in medicine and healthcare.

2.3.2 Categories of overviews:

Overviews addressing the methodological quality of reviews and/or meta-analyses in anesthesia / pain management (7, 19%) and general medical journals (5, 14%) were most frequent. Cochrane reviews, specialty topics in internal medicine, and complementary medicine, were also the focus of attention in some overviews (4, 11% each). Reviews of screening and diagnostic tests (3, 8%), dentistry (2, 6%), surgical specialties (orthopedics, gynecology and obstetrics, one each), and emergency medicine and primary care (one each), and other areas of interest (3, 8%) such as health economics, DARE (Database of Abstracts of Reviews of Effects) and epidemiology (one each) were also represented in this sample.

2.3.3 Quality scoring of reviews:

A considerable proportion of overviews (13, 36%) used the Overview Quality Assessment Questionnaire (OQAQ) or a modified version (4, 11%) of this instrument to assess the methodological quality of the reviews and/or meta-analyses. Five overviews (14%) used the Mulrow criteria, whereas the Sacks checklist was used in 3 (8%) overviews. Other validity criteria were used less frequently such as the Barnes & Bero score, the Russell criteria, the QUOROM guidelines (one overview each) and some Cochrane guidelines (two overviews). Of note, 8 (22%) reviews failed to mention the quality system they used to critically appraise the systematic reviews and/or meta-analyses under study. Half (50%) of the overviews reported the results of the quality assessment in terms of a total score. Among the overviews that used the OQAQ for quality assessment, five reported the median score (weighted median: 3.47) and seven reported the mean total score (weighted mean: 3.96), indicating a pattern of major flaws in the methodological quality of the reviews / meta-analyses included in the overviews. The results of the OQAQ mean scores could not be plotted against the median year of publication of the systematic reviews and/or meta-analyses included in the overviews to explore the trends of the quality over time due to the paucity of data for this analysis.

2.3.4 Quality assessment:

The results of the methodological quality of the systematic reviews and/or meta-analyses were categorized as "high", "moderate" and "low" according to the numeric results and conclusions of the overviews' authors (Table 2.2). Two overviews (6%) concluded that Cochrane reviews had higher methodological rigor compared to those published in paper-based journals. Reviews and/or meta-analyses in anesthesia and pain management, complementary medicine and orthopedics were found to be of moderate quality in 8 (22%) overviews. A considerable proportion (12, 33%) of overviews were classified as having low methodological quality, whereas 14 (39%) overviews did not provide a definite interpretation of the quality of the reviews under consideration; however, 14 (100%) of these made a call for improved quality in reporting.

2.3.5 Type of recommendations:

Overall, all authors called for an improvement in the methods to conduct review papers in the future. The authors highlighted some methodological and reporting issues in the domains of the methods to avoid selection bias, the lack of report of criteria to assess the validity of included studies, heterogeneity testing and publication bias. They also concluded that improvements to certain aspects related to the search of the literature and synthesis of the results were required. While Cochrane reviews could be expected to be of higher methodological quality and less prone to bias than systematic reviews published in traditional medical journals, conclusive evidence with respect to this was lacking.

2.4 Discussion:

This review was conducted to describe the evidence for quality of reviews available in the published biomedical literature. Using a comprehensive approach, 36 overviews of reviews were identified. The evidence on the methodological quality of systematic reviews and/or meta-analyses mainly arises from overviews in the areas of anesthesia and pain management, and general medical journals with the OQAQ as the most frequently used quality tool. The quality of reviews produced in languages other than English is uncertain, as most of the studies restricted their searches by language of publication. Indeed, there are no studies addressing the frequency of publication of non-English systematic reviews and how their results and quality compare with those found in the English scientific literature.

Although different methodological criteria have been used to assess the methodological validity of systematic reviews and/or meta-analyses, quality is generally reported as low. These findings raise concerns about the validity of the conclusions and recommendations reached by SRs in these fields. Overall, the findings from this study suggest that few overviews of systematic review and/or meta-analyses revealed high quality work in specific fields of academic medicine and allied sciences. Many of them in fact provided insufficient data upon which to base any assessment and/or interpretation of the quality. Alternatively, it is important to note that the proportion of systematic reviews considered as of moderate quality may be overestimated, as three of the complementary medicine overviews [41-43] are duplicate publications from a single overall report [8]. The situation may therefore be worse than estimated here.

Recently, evidence-based initiatives to improve the methodological quality and reporting of systematic reviews and/or meta-analysis have emerged. In 1999, the QUOROM (Quality of Reports of Meta-analysis) conference [54] was convened to address general standards for improving the quality of reports of systematic reviews and/or meta-analyses of RCTs. Many, but not all, editorial boards of scientific journals have adopted the QUOROM reporting standards for SRs. In the emergency medicine field, only the *Canadian Journal of Emergency Medicine* has adopted the QUOROM statement to improve the quality of systematic reviews published into its instructions for authors.

The QUOROM group consisted of clinical epidemiologists, clinicians, statisticians, editors, and researchers that met to prepare recommendations on a structure of reporting details of systematic reviews to ensure that the transparency of the methods and the methodological quality of a systematic review are not hampered by reporting issues. The resulting QUOROM statement is a checklist organised into 21 headings and subheadings that includes 18 evidence-based items addressing the quality of the Abstract, Introduction, Methods, and Results sections of reports of systematic reviews of RCTs (Table 2.3). While the checklist was not intended to be a quality scoring tool for SRs, it has nonetheless been considered as a standard for appraisal of SR internal validity.

The checklist encourages authors to provide readers with information on searches, selection, validity assessment, data abstraction, study characteristics, quantitative data synthesis, and trial flow. Authors are asked to provide a flow diagram providing information about the number of RCTs identified, included, and excluded and the reasons for excluding them [54].

Among the goals of adopting QUOROM are to provide guidance to authors, to allow for more objective refereeing, and to produce higher quality reviews [55]. It has been expected that the publication and endorsement of the QUOROM guidelines in several leading medical journals (e.g., *BMJ*, *JAMA*, *Lancet*) would lead to improvements in the systematic reviews and meta-analyses published in the medical literature. Based on the information provided by the group of overviews identified here, no conclusions regarding an improvement in the methodological quality of systematic reviews and/or meta-analyses following the publication of the QUOROM guidelines can be made. Therefore, a study testing whether the quality of systematic reviews has improved after the publication of the QUOROM statement is warranted. In the following Chapter, a review of the progress made in the emergency literature will be explored.

This structured review had some limitations: First, the search for overviews was almost entirely based on a strategy that used a single electronic database (MEDLINE). This approach limited the comprehensiveness of the review because some reports not indexed in the scientific literature may have been missed. It is unknown how many other overviews not indexed in MEDLINE but referenced in other electronic databases such as EMBASE, LILACS or CINAHL have been produced. A simple search of EMBASE and CINAHL failed to produce additional references. Another limitation is that the coding of SRs is complex and a variety of terms are used to describe them; however, we feel the terms employed would have missed few. Finally, a single investigator conducted the selection of studies and data extraction; therefore the introduction of some type of selection bias (i.e. that studies are not representative of the population of all possible studies about this topic) cannot be disregarded.

2.5. Conclusions:

Although an important proportion of studies on the quality of systematic reviews and/or meta-analyses in a variety of medical specialties have been conducted to date, evidence from overviews shows a variable pattern of methodological quality across several medical and allied sciences. Reasons for variation may be attributed to differences in the quality assessment instruments employed or to real differences in their methods that deserve further analyses. Future studies should assess the sources of heterogeneity and variations in the level of quality of systematic reviews and/or meta-analyses. They should also evaluate the effect of the introduction of guidelines such as QUOROM on the level of methodological quality of systematic reviews and/or meta-analyses produced in a variety of areas of academic medicine and allied sciences. The use of validated instruments such as the OQAQ index in future evaluations is preferred to measure these changes in order to enable further comparisons across several medical and allied specialties.

Systematic reviews of RCTs have been considered as the “gold standard” of evidence for clinical and policy decisions in healthcare. Nevertheless, this statement cannot be confidently endorsed until a uniform and appropriate level of methodological quality is reached across the scientific literature. In the meantime, editors and authors should concentrate their efforts in adopting standardized guidelines for submission and appraisal for publication of higher quality systematic reviews and/or meta-analyses.

Table 2.1 Summary of studies analyzing the methodological quality of systematic reviews and meta-analysis in health care research.

TOPIC	STUDY	LANGUAGE RESTRICTIONS	NUMBER OF REVIEWS/ META-ANALYSIS	YEARS COVERED	MEDIAN YEAR OF REVIEWS	METHOD OF QUALITY ASSESSMENT	RESULTS AND LEVEL OF QUALITY
Anesthesia & Pain Management	Jadad 1996 [22]	No	80	1980-1992	1990	OQAQ	Median score = 4; Quality level: Moderate.
	Smith 1997 [33]	Yes	25	1995	1995	Mulrow criteria	No overall score provided; Quality level: Low.
	Choi 2001 [34]	N/S	82	1966-1999	1997	OQAQ	Median score = 4; Quality level: Moderate.
	Fishbain 2000 [35]	N/S	16	1966-1996	1992	OQAQ modified (max = 33)	Mean score (SD) = 18.63 (5.62); Quality level: Not interpreted.
	Furlan 2001 [19]	No	36	Up to 1998	1993	OQAQ	Mean (SD) = 4.19 (1.71); Quality level: Moderate.
	Assendelft 1995 [36]	Yes	51	1977-1992	1986	OQAQ modified (max = 100)	Median score = 23/100; Quality level: Low.
	Hoving 2001 [37]	No	12	1966-1998	1982	OQAQ modified (max = 18)	Mean score = 8.5/18; Quality level: Not interpreted.
General medical journals	Mulrow 1987 [4]	Yes	50	1985-1986	1986	Mulrow criteria	No overall score provided; Quality level: Low.
	Sacks 1996 [27]	Yes	164	1955-1990	N/S	Sacks checklist (14 items)	Mean (SD) score = 7.63 (2.84) for 1955-1982 (n = 40); 6.80 (3.86) for 1983-1986 (n = 40); 11.91 (4.79) for 1987-1990 (n = 58); Quality level: Not interpreted.
	McAllister 1999 [28]	Yes	158	1996	1996	Mulrow criteria	No overall score provided; Quality level: Not interpreted.
	Moher 2000 [29]	Yes	79	Up to 1996	N/S	OQAQ	Median score = 3; Quality level: Low.
	Rochon 2002 [30]	Yes	16	1998	1998	Barnes & Bero score	Mean (SD) score: medical journals = 0.94, throwaway journals = 0.23; Quality level: Not interpreted

Table 2.1 Summary of studies analyzing the methodological quality of systematic reviews and meta-analysis in health care research (Cont'd).

TOPIC	STUDY	LANGUAGE RESTRICTIONS	NUMBER OF REVIEWS/ META-ANALYSIS	YEARS COVERED	MEDIAN YEAR OF REVIEWS	METHOD OF QUALITY ASSESSMENT	RESULTS AND LEVEL OF QUALITY OF REVIEWS
Cochrane	Jadad 1998 [21]	N/S	36 Cochrane 39 paper-based	1995	1995	N/S	No overall score provided; Quality level: Not interpreted.
	Olsen 2001 [48]	Yes	53	1998	1998	Non-validated system.	No overall score provided; Quality level: Not interpreted.
	Shea 2002 [51]	No	52 Cochrane 52 paper-based	1966-1996	N/S	OQAQ Sacks checklist.	Mean score (95%CI): Cochrane: 3.42 (2.92-3.93);paper-based: 3.35 (2.83-3.87) Quality level: Low.
	Van Tulder 2003 [52]	N/S	28	1997-2002	N/S	Cochrane Back Review Group Method Guidelines.	No overall score provided; Quality level: Not interpreted.
Specialty topics in Internal Medicine	Bramwell 1997 [39]	Yes	176	1983-1995	1988	Mulrow criteria Sacks criteria	No overall score provided; Quality level: Low.
	Jadad 2000 [11]	N/S	50	1988-1998	1997	OQAQ	Median score All asthma reviews = 3 Cochrane reviews = 6 Paper-based reviews = 2; Quality level: Cochrane: High; All: Low.
	Goodwin 2002 [40]	N/S	5	1966-2000	1997	Russell criteria	Mean score = 8/14; Quality level: Moderate.
	Christiensen 2001 [5]	Yes	15	N/S	1998	QUOROM (18 points)	No overall score provided; Quality level: Low.
Complementary medicine	Moher 2002 [44]	N/S	47	N/S	1998	OQAQ	Median score = 3; Quality level: Low.
	Linde 2001 [8, 42]	N/S	39	1989-2000	1997	OQAQ	Mean score (SD): 4.6 (1.5); Quality level: Moderate.
	Linde 2001 [8, 43]	N/S	58	1989-2000	1998	OQAQ	Mean score (SD): 4.7 (1.5); Quality level: Moderate.
	Linde 2001 [8, 41]	N/S	18	1989-2000	1998	OQAQ	Mean score (SD): 4.6 (1.5); Quality level: Moderate.

Table 2.1 Summary of studies analyzing the methodological quality of systematic reviews and meta-analysis in health care research (Cont'd).

TOPIC	STUDY	LANGUAGE RESTRICTIONS	NUMBER OF REVIEWS/ META-ANALYSIS	YEARS COVERED	MEDIAN YEAR OF REVIEWS	METHOD OF QUALITY ASSESSMENT	RESULTS AND LEVEL OF QUALITY OF REVIEWS
Screening & Diagnostic tests	Walter 1999 [47]	N/S	57	1966-1997	1994	Non-validated system.	No overall score provided; Quality level: Not interpreted.
	Oosterhuis 2000 [2]	N/S	23	1985-1998	1996	Cochrane Methods Working Group Guidelines	No overall score provided; Quality level: Not interpreted.
	Irwig 1994 [49]	Yes	11	1990-1991	N/S	Guidelines developed by authors	No overall score provided; Quality level: Not interpreted.
Dentistry	Creugers 2003 [45]	Yes	138	1990-2001	1996	Individual components	No overall score provided; Quality level: Low.
	Glenny 2003 [46]	N/S	65	1994-2001	1998	Checklist developed by authors (15 items)	No overall score provided; Quality level: Low.
Surgical specialties	Bhandari 2001 [38]	N/S	40	1969-1999	1993	OQAQ	Mean score (SEM): 4.2 (1.78); Quality level: Moderate.
	Peipert 1997 [6]	Yes	46	1986-1995	N/S	N/S	No overall score provided; Quality level: Not interpreted.
Emergency & Primary Care	Silagy 1993 [32]	Yes	28	1991	1991	Mulrow criteria	No overall score provided; Quality level: Not interpreted.
	Kelly 2001 [53]	Yes	29	1988-1998	1995	OQAQ	Mean score (95%CI) = 2.7 (2.1-3.2); Quality level: Low
Other	Breslow 1998 [31]	Yes	29	1995	1995	OQAQ	No overall score provided; Quality level: Not interpreted
	Jefferson 2002 [50]	No	39	1990- 2001	1999	OQAQ & Mulrow modified.	No overall score provided; Quality level: Low.
	Petticrew 1999 [23]	N/S	480	1994-1998	1996	Individual quality components.	No overall score provided; Quality level: Not interpreted.

Table 2.2: Quality of systematic reviews and/or meta-analyses by topic published in health care research literature.

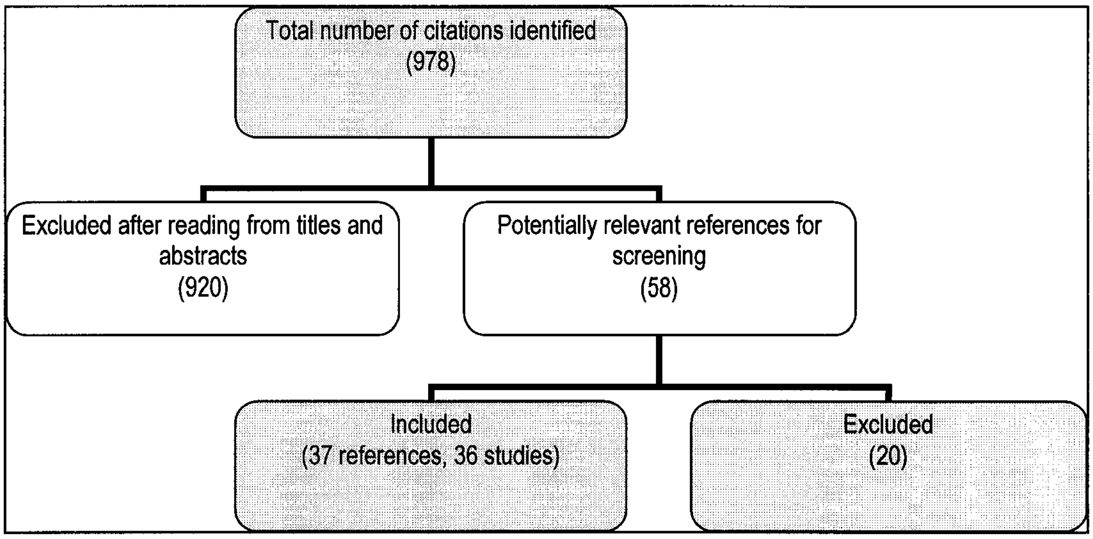
Topic of study	Quality of systematic reviews and / or meta-analyses ‡ n (%)				
	High	Moderate	Low	Not interpreted	TOTAL
Anesthesia and Pain Management	0	3	2	2	7
General Medical journals	0	0	2	3	5
Cochrane	1	0	1	2	4
Specialty topics in Internal Medicine	1	1	2	0	4
Complementary Medicine	0	3	1	0	4
Screening and diagnostic tests	0	0	0	3	3
Dentistry	0	0	2	0	2
Surgical specialties	0	1	0	1	2
Emergency & Primary Care	0	0	1	1	2
Other	0	0	1	2	3
TOTAL	**2 (5.5%)**	**8 (22.2%)**	**12 (33.3%)**	**14 (38.8%)**	**36 (100%)**

‡ Based on the conclusions/interpretation of overviews authors

Table 2.3: The QUOROM statement checklist [54].

Heading	Subheading	Descriptor
Title		Identify the report as a meta-analysis (or systematic review) of RCT.
Abstract		Use a structured format.
		Describe
	Objectives	The clinical question explicitly.
	Data sources	The databases (i.e., list) and other information sources.
	Review methods	The selection criteria (i.e. population, intervention, outcome, and study design); methods for validity assessment, data abstraction, and study characteristics, and quantitative data synthesis in sufficient detail to permit replication.
	Results	Characteristics of the RCTs included and excluded; qualitative and quantitative findings (i.e. point estimates and confidence intervals); and subgroup analyses.
	Conclusion	The main results.
		Describe
Introduction		The explicit clinical problem, biological rationale for the intervention, and rationale for review.
Methods	**Searching**	The information sources, in detail (e.g. databases, registers, personal files, expert informants, agencies, hand searching), and any restrictions (years considered, publication status, language of publication).
	Selection	The inclusion and exclusion criteria (defining population, intervention, principal outcomes, and study design).
	Validity assessment	The criteria and process used (e.g. masked conditions, quality assessment, and their findings).
	Data abstraction	The process or processes used (e.g. completed independently, in duplicate).
	Study characteristics	The type of study design, participants' characteristics, details of intervention, outcome definitions, and how clinical heterogeneity was assessed.
	Quantitative data synthesis	The principal measures of effect (e.g. relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data, how statistical heterogeneity was assessed, a rational for any a-priori sensitivity and subgroup analyses, and any assessment of publication bias.
Results	**Trial flow**	Provide a meta-analysis profile summarizing trial flow.
	Study characteristics	Present descriptive data for each trial (e.g. age, sample size, intervention, dose, duration, follow-up period).
	Quantitative data synthesis	Report agreement on the selection and validity assessment, present simple summary results (for each treatment group in each trial, for each primary outcome), present data needed to calculate effect sizes and confidence intervals in intention-to-treat analyses (e.g. 2x2 tables of counts, means and standard deviations, proportions).
Discussion		Summarize key findings, discuss clinical inferences based on internal and external validity, interpret the results in light of the totality of available evidence, describe potential biases in the review process (e.g. publication bias) and suggest a future agenda.

Figure 2.1 Flow-diagram for selection of studies: Overviews of systematic reviews and/or meta-analyses published in health care research literature.



2.6. References

(1) Wyer PC, Rowe BH, Guyatt GH, Cordell WH. The clinician and the medical literature: when can we take a shortcut? Ann Emerg Med 2000; 36:149-155.

(2) Oosterhuis WP, Niessen RWLM, Bossuyt PMM. The science of systematic reviewing studies of diagnostic tests. Clin Chem Lab Med 2000; 38:577-588.

(3) Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991; 44:91-98.

(4) Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485-488.

(5) Christensen E. Quality of reporting of meta-analyses: the QUOROM statement. Will it help? J Hepatol 2001; 34:342-345.

(6) Peipert JF, Bracken MB. Systematic review of medical evidence: the use of meta-analysis in obstetrics and gynecology. Obstet Gynecol 1997; 89:628-633.

(7) NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. Khan KS, ter Riet G, Glanville J, Sowden AJ, Kleijnen J, editors. 4. York: NHS Centre for Reviews and Dissemination, University of York. CRD Reports, 2001.

(8) Linde K, ter Riet G, Hondras M, Melchart D, Willich SN. Characteristics and quality of systematic reviews of acupuncture, herbal medicines, and homeopathy. Forsch Komplementar med Klass Naturheilkd 2003; 10:88-94.

(9) Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for health care decisions. Philadelphia: American College of Physicians, 1998.

(10) Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci 1993; 703:125-133.

(11) Jadad AR, Moher M, Browman GP, Lynda Booker, Sigouin C, Fuentes M et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000; 320:537-540.

(12) Egger M, Davey Smith G. Meta-analysis: potentials and promise. BMJ 1997; 315:1371-1374.

(13) Chalmers I, Altman DG. Foreword. In: Chalmers I, Altman DG, editors. Systematic reviews. London: BMJ Publishing, 1995.

(14) Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309:597-599.

(15) Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002; 31:115-123.

(16) Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews of the effects of health care. In: Chalmers I, Altman DG, editors. Systematic reviews. London: BMJ, 1995.

(17) Rowe BH, Wyer PC, Cordell WH. Improving the dissemination of systematic reviews in emergency medicine. Ann Emerg Med 2002; 39:293-295.

(18) Hopayian K. The need for caution in interpreting high quality systematic reviews. BMJ 2001; 323:681-684.

(19) Furlan A, Clarke J, Esmail R, Sinclair S, Irvin E, Bombardier C. A critical review of reviews on the treatment of chronic low back pain. Spine 2001; 26:E155-E162.

(20) Schringer DL. Finding truths in clinical medicine: through the looking glass-cracked. Ann Emerg Med 2001; 38:566-569.

(21) Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M et al. Methodology and reports of systematic reviews and meta-analyses: comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998; 280:278-280.

(22) Jadad AR, McQuay H. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. J Clin Epidemiol 1996; 49:235-243.

(23) Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). Int J Technol Assess Health Care 1999; 15:671-678.

(24) 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; 273:408-412.

(25) Egger M, Smith GD. Bias in location and selection of studies. BMJ 1998; 316:61-66.

(26) Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44:1271-1278.

(27) Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: An update. Mt Sinai J Med 1996; 63:216-224.

(28) McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FME, Moher D et al. The medical review article revisited: has the science improved? Ann Intern Med 1999; 131:947-951.

(29) Moher D, Pham B, Klassen T, Schulz KF, Berlin JA, Jadad AR et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol 2000; 53:964-972.

(30) Rochon PA, Bero LA, Bay AM, Gold JL, Dergal JM, Binns MA et al. Comparison of review articles published in peer-reviewed and throwaway journals. JAMA 2002; 287:2853-2856.

(31) Breslow RA, Ross SA, Weed DL. Quality of reviews in epidemiology. Am J Public Health 1998; 88:475-477.

(32) Silagy CA. An analysis of review articles published in primary care journals. Fam Pract- An International Journal 1993; 10:337-340.

(33) Smith AF. An analysis of review articles published in four anaesthesia journals. Can J Anaesth 1997; 44:405-409.

(34) Choi PT, Halpern SH, Malik N, Jadad AR, Tramer MR, Walder B. Examining the evidence in anesthesia literature: a critical appraisal of systematic reviews. Anesth Analg 2001; 92:700-709.

(35) Fishbain D, Cutler RB, Rosomoff HL, Rosomoff RS. What is the quality of the implemented meta-analytic procedures in chronic pain treatment meta-analyses? Clin J Pain 2000; 16:73-85.

(36) Assendelft WJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. JAMA 1995; 274:1942-1948.

(37) Hoving JL, Gross AR, Gasner D, Kay T, Kennedy C, Hondras MA et al. A critical appraisal of review articles on the effectiveness of conservative treatment for neck pain. Spine 2001; 26:196-205.

(38) Bhandari M, Morrow F, Kulkarni AV, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. J Bone Joint Surg Am 2001; 83-A:15-24.

(39) Bramwell VHC, Williams CJ. Do authors of review articles use systematic methods to identify, assess and synthesize information? Ann Oncol 1997; 8:1185-1195.

(40) Goodwin DM, Higginson IJ, Edwards AG, Finlay IG, Cook AM, Hood K et al. An evaluation of systematic reviews of palliative care services. J Palliat Care 2002; 18:77-83.

(41) Linde K, Hondras M, Vickers A, ter Riet G, Melchart D. Systematic reviews of complementary therapies - an annotated bibliography. Part 3: homeopathy. BMC Complement Altern 2001; 1:4.

(42) Linde K, Vickers A, Hondras M, ter Riet G, Thormahlen J, Berman B et al. Systematic reviews of complementary therapies - an annotated bibliography. Part 1: acupuncture. BMC Complement Altern Med 2001; 1:3.

(43) Linde K, ter Riet G, Hondras M, Vickers A, Saller R, Melchart D. Systematic reviews of complementary therapies - an annotated bibliography. Part 2: herbal medicine. BMC Complement Altern Med 2001; 1:5.

(44) Moher D, Soeken K, Sampson M, Ben-Porat L, Berman B. Assesing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. BMC Pediatr 2002; 2:3.

(45) Creugers NH, Kreulen CM. Systematic review of 10 years of systematic reviews in prosthodontics. Int J Prosthodont 2003; 16:123-127.

(46) Glenny AM, Esposito M, Coulthard P, Wotyhington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci 2003; 111:85-92.

(47) Walter SD, Jadad AR. Meta-analysis of screening data: a survey of the literature. Stat Med 1999; 18:3409-3424.

(48) Olsen O, Middleton P, Ezzo J, Gotzsche P, Hadhazy V, Herxheimer S et al. Quality of Cochrane reviews: assessment of a sample from 1998. BMJ 2001; 323:829-832.

(49) Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120:667-676.

(50) Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. JAMA 2002; 287:2809-2812.

(51) Shea B, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. Eval Health Prof 2002; 25:116-129.

(52) van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine 2003; 28:1290-1299.

(53) Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH. Evaluating the quality of systematic reviews in the emergency medicine literature. Ann Emerg Med 2001; 38:518-526.

(54) Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of meta-analyses of randomized controlled trials: The QUOROM statement. Quality of Reporting of Meta-analysis. Lancet 1999; 354:1896-1900.

(55) Needleman I, Grace M, Sloan P. QUOROM and systematic reviews. Br Dent J 2002; 192:605.

CHAPTER 3: Assessing the quality of reports of systematic reviews in emergency medicine: are they improving?

3.1. Introduction

Although different criteria have been used to assess the methodological validity of systematic reviews (SR) and/or meta-analyses (MA) in the medical literature, overall, the quality has been reported to be "low", thereby raising concerns about the validity of their conclusions and recommendations. Consequently, the Quality of reporting of meta-analyses (QUOROM) statement was developed to assist authors and journals in the review of systematic reviews. There is a lack of information on how the quality of systematic reviews and/or meta-analyses published in the scientific literature has improved after the publication of these QUOROM guidelines.

In the field of Emergency Medicine (EM), Kelly et al [1] examined the scientific quality of 29 systematic reviews published in 5 leading EM journals from 1988 to 1998. Results using the Overview Quality Assessment Questionnaire (OQAQ) total score (mean score: 2.7; 95% confidence interval (CI): 2.1, 3.2; Median: 2; Inter Quartile Range (IQR): 2, 3) indicated that the overall scientific quality was low. These authors concluded that the considerable confusion in the design, reporting and description of systematic reviews might limit the validity of the reported results.

This study is intended to expand on the work that Kelly et al [1] initiated by critically appraising the methodological quality of systematic reviews published from 1988 to 1998 in major emergency medicine journals. It evaluates the methodological quality of systematic reviews published in these journals from January 2000 to March 2004 and compares these results to those published from 1988 to 1998 (before the publication of the QUOROM guidelines). The study also attempts to identify factors associated with methodological quality of systematic reviews in emergency medicine.

This research will be useful to illustrate how the methodology of design, conduct and reporting of systematic reviews has changed over time in EM as a discipline. It will also allow researchers undertaking reviews in the emergency medicine area, journal editors and readers to be aware of the main flaws and the issues that should be addressed to improve the quality of systematic reviews published in EM journals in the near future.

3.2. Objectives

The objectives of this study were:

1) To evaluate the scientific quality (clinical, methodological, and reporting aspects) of systematic reviews published in 7 leading emergency medicine journals from January 2000 to March 2004;
2) To compare how the scientific methods and quality of systematic reviews published from January 2000 to March 2004 have evolved in relation to those published during the period of 1988-1998 in major emergency medicine journals;
3) To describe the association between the methodological quality and other characteristics of systematic reviews published in major emergency medicine journals from January 2000 to March 2004.
4) To explore potential factors associated with the quality of systematic reviews in the emergency medicine literature.

3.3. Methodology of the study

3.3.1. Study design:

Retrospective comparative study of all systematic reviews published from January 2000 to March 2004 compared to January 1988- December 1998 in major emergency medicine journals.

3.3.2. Eligibility criteria:

Systematic reviews were defined here as studies that clearly formulate a research question, utilize a search strategy to identify studies for inclusion, and that attempt to summarize and analyze data from the primary studies [2]. According to the technique of summarizing the results, they may be classified as qualitative or quantitative systematic reviews. A qualitative systematic review synthesizes primary studies; however, data are not statistically combined. A quantitative systematic review may also be referred to as a meta-analysis, i.e. a review that uses statistical methods to combine data from primary studies to arrive at conclusions about a body of research [3].

Articles had to be full-reported studies including the words "review", "overview", "systematic review" or "meta-analysis" in the title or abstract; or indicating in the text that the intention was to review or summarize the literature about a clinical topic. General or narrative reviews of topic areas and case reports with appended literature reviews were not eligible. Editorials, correspondence, evidence-based EM shortcut reviews, clinical practice guidelines, economic evaluations and conference summaries were excluded. Updates of reviews were considered as separate publications. No restrictions on type of patients, type of interventions and/or type of outcomes under review were applied.

Articles describing a systematic review or a meta-analysis and published in any of the following leading emergency medicine journals were considered for inclusion: *Academic Emergency Medicine* (AEM), *American Journal of Emergency Medicine* (AJEM), *Annals of Emergency Medicine* (AOEM), *Canadian Journal of Emergency Medicine* (CJEM), *Emergency Medicine* (EM), *Emergency Medicine Journal* (EMJ) (former *Journal of Accident and Emergency Medicine*), and *Journal of Emergency Medicine* (JEM).

There were several reasons for selecting these seven emergency medicine journals for inclusion in this study. First, they represent a convenience sample of peer-reviewed journals in emergency medicine selected according to the Science Citation Index (SCI) for 2002 [4]. The Impact Factor (IF) is a measure of the frequency with which articles are cited in other journals. It is considered a measure of a journal's influence; for example, *Annals of Emergency Medicine*, the highest ranking EM Journal, is ranked approximately 2.9 while *New England Journal of Medicine* is ranked >25. The seven journals selected for this study are ranked among the 10 most influential journals in the emergency medicine specialized literature (Appendix A3). Although the CJEM is not an indexed journal, it was included because it is the first and only periodical publication focused on emergency medicine published in Canada. Second, in order to enable comparisons with the cohort of systematic reviews published in emergency medicine journals from 1988 to 1998 [1], we decided to include five of the journals selected for the previous Kelly study (AEM , AJEM, AOEM, JAEM, and JEM) adding two further publications (EM and CJEM). It is important to note that CJEM is a spin-off of the JEM.

3.3.3. Search strategy for identification of studies:

MEDLINE and EMBASE databases were electronically searched from January 2000 to March 2004 using three different search strategies (See Appendix A2 for search strategies). All the editions of the seven emergency medicine journals published between January 2000 and March 2004 were hand-searched by one investigator (MO) to identify potential studies for inclusion.

3.3.4. Selection of studies:

Hard copies of all potentially relevant articles identified through the search strategies were retrieved. Two investigators (MO, BR) independently compared each study against the selection criteria to judge their eligibility. Each reviewer independently applied the selection criteria to the studies, assigning them to the following categories: "excluded", "included", "unclear/unsure". Identity of the primary authors and their affiliations and journal name was not masked. In case of disagreement, the final decision on inclusion of studies was made by consensus.

3.3.5. Assessment of methodological quality:

Two independent investigators (MO, BR) assessed the methodological quality of included articles using the Overview Quality Assessment Questionnaire (OQAQ) developed by Oxman et al[5] (See Appendix B1). The OQAQ is a 10-item scale where the first 9 questions are rated on a yes/partially/no format to assess how the review was designed and reported. A final question on the overall scientific quality of the review is rated on a 7-point Likert scale. The tool consists of questions on how the review is designed and reported; and it does not require knowledge about the included trials themselves. This scale was selected to assess the methodological quality of systematic reviews because its psychometric properties have been thoroughly tested and clearly validated using a number of different measures [5-7]. Raw rating scores for each criterion, for each systematic review were tabulated. Articles were classified as having serious or extensive flaws if they received a score of 1 to 3, and as having minimal or minor flaws if they received scores from 4 to 7.

The level of agreement between the two reviewers on the individual scores of the OQAQ was measured by calculating a Kappa (κ) coefficient [8]. A κ score in the range of 0.0 to 0.40 was considered poor agreement, from 0.41 to 0.60 moderate agreement, and a κ in the range of 0.61 to 0.80 was considered to represent substantial agreement [9]. Inter-rater agreement between the

OQAQ overall scores was estimated with the Intra Class Correlation Coefficient (ICC) [10] and 95% confidence intervals (95%CI). For this study, moderate agreement was defined as ICC > 0.6 and good agreement as ICC > 0.8 [5].

3.3.6. Data extraction:

One reviewer (MO) extracted data from each article using a standard format. Data on the following characteristics of the studies were obtained: number of authors, journal of publication, publication year, country of corresponding author, topic of study (intervention, diagnosis, other), sources of studies (bibliographic databases, hand search, reference list, contact with primary authors, contact with experts in the area, contact with appropriate industry sector, conference proceedings, theses, technology reports, unpublished studies), publication restrictions, language restrictions, design of primary studies included in the review, quality assessment criteria, type of analysis (qualitative and/or quantitative data synthesis), type of pooled measure for primary outcome (if applicable), main study results (positive, negative/uncertain), and sources of funding.

3.3.7. Data analysis

General information on the clinical, methodological and reporting aspects of the systematic reviews published during the 2000-2004 period is presented in frequency tables. Categorical data are reported as proportions and percentages with 95% CIs. Continuous data are reported using means with standard deviations (SD) or medians with interquartile ranges (IQR), where appropriate. The data were separated in two cohorts: articles published from 1988 to 1998 (Group 1, before the publication of the QUOROM statement) [1], and those published from 2000 to 2004 (Group 2, after the publication of the QUOROM statement). Systematic reviews published in theCJEM and EM journals were excluded from between-cohort comparisons, as these journals were not included in the original cohort of publications.

The mean overall quality scores (OQAQ) of the two cohorts of systematic reviews (1988-1998 and 2000-2004) were compared using the t-test statistic and the difference in means statistics. Chi-square (χ^2) tests were used to compare the proportion of articles in Group 1 and Group 2 that meet each of the 9 individual items of the OQAQ. Fisher's exact test was used when any expected cell count was less than 5 for any of these comparisons. Values of $p < 0.05$ were regarded as statistically significant for all the comparisons.

In order to describe the relation between the methodological quality and other characteristics of systematic reviews published in major emergency medicine journals, a univariate regression analysis was used. The goal was to identify important factors influencing the methodological quality of the systematic reviews published in the selected emergency medicine journals. The association of the following independent variables with the OQAQ overall score was examined: number of authors, journal of publication (above and below the median of the IF for the selected journals), study group (publication before and after the QUOROM statement), country of corresponding author (North America, other regions), topic of study (therapy, diagnosis, other), type of analysis (qualitative, meta-analysis), source of funding (declared, not declared), and main conclusions (positive, negative/uncertain). This set of independent variables was chosen *a priori*, based on data from the available literature for this topic, and from consensus among investigators regarding their potential relevance. Individual components of the OQAQ were not included in the regression analyses due to the issue of collinearity (or lack of independence) with the OQAQ total score. Those variables revealing a significant association with the quality of the systematic reviews in the univariate analysis were explored in a multiple regression model using a stepwise- selection procedure with the OQAQ overall score as the dependent variable. The results from this analysis were reported as regression coefficients (β) with 95% confidence intervals. Again, values of $p < 0.05$ were regarded as statistically significant for all the comparisons. All data were entered and analyzed in SPSS (Statistical Package for Social Sciences, Version 12.0 for Windows©).

3.3.8. Ethics

This study is exempt from ethics review by the Health Research Ethics at the University of Alberta, Edmonton, Canada as it did not involve the participation of patients or their families.

3.4. Results

3.4.1. The search:

One hundred and ninety citations were identified for potential inclusion in this study. Ninety-six potentially relevant citations were identified from computerized searches (see Appendix 1). The search strategy from the Evidence Based Informatics Project at McMaster University for MEDLINE yielded 26 records of potential review articles published from January 2000 to March

2004. The strategy for MEDLINE adapted from Kelly et al [1] identified 33 records, and the EMBASE search based on the McMaster strategy resulted in 37 hits. Manual searches on the seven selected emergency medicine journals resulted in 94 potentially relevant articles. After deleting duplicates (28 records) and excluding references published in other journals than the seven emergency medicine journals (27), the titles and abstracts of 135 references were retrieved for further examination (Figure 3.1).

After reading from titles and abstracts, 41 further references were excluded resulting in a group of 94 potential articles that were assessed independently by two reviewers (MO and BR) to decide their inclusion in this study.

3.4.2. Selection of studies:

Agreement between the two researchers (MO, BR) for selection of studies was good (N = 94;κ = 0.64; p = 0.0006). All differences were resolved by consensus (100% agreement after discussions). Forty-nine articles were excluded (Table 3.1). Most of them were excluded because no identifiable question was clearly formulated (23, 46.9%), they were narrative reviews (9, 18.4%), clinical practice guidelines (9, 18.4%), Evidence-based EM shortcut reviews (5, 10.2%), case reports with non-systematic literature reviews (2, 4.1%) or editorials (1, 2.0%). Forty-five reviews published in these seven major emergency medicine journals from January 2000 to March 2004 met all the criteria for selection (Table 3.2).

3.4.3. Characteristics of the included studies:

a. Publication aspects:

The distribution of reviews by the 7 journals is as follows: 17 (37.8%) in EMJ, 13 (28.9%) in AOEM, 6 (13.3%) in AEM, 3 (6.7%) in JEM and 2 (4.4%) in AJEM, CJEM and EM each. The years with the highest rates for publications of systematic reviews were 2001 and 2003 with 12 (26.7%) each (Table 3.3). The mean number of authors per review was 3.04 (SD: 1.77; median: 3, IQR: 2, 4) ranging from 1 to 9 authors. The United States (16, 35.6%) and the United Kingdom (15, 33.3%) were the most common countries of origin for corresponding authors of systematic reviews followed by Canada (9, 20.0%), and Australia (2, 4.4%). Other countries were less represented (3, 6.7%) in this sample of reviews. Twenty-one (46.7%) systematic reviews addressed topics related with therapeutic interventions, whereas ten reviews (22.2%) investigated diagnostic interventions.

One (2.2%) review examined therapeutic and diagnostic interventions simultaneously, and the remaining 13 (28.9%) reviews evaluated other topics such as health services evaluation, education and management in the emergency medicine setting.

b. Methodological aspects:

Most of the reviews (32, 71.1%) addressed a single research question while 13 (28.9%) examined multiple research questions. All the reviews used electronic databases to identify potential studies for inclusion. MEDLINE was the most commonly used single database (43, 95.6%) followed by EMBASE (17, 37.8%). Thirty-two (71.1%) reviews reported the years of coverage of the search, 33 (73.3%) provided the search terms, and only six (13.3%) reproduced the entire search strategy. Manual search on list of references (38, 84.4%) was the most frequent "non-electronic" source to identify potentially relevant primary studies. Contact with experts and colleagues in the area (13, 28.9%) and hand searches of scientific journals (12, 26.7%) were less often used to identify further studies for inclusion (Figure 3.2). Only five (11.1%) reviews reported the search for unpublished material. Ten (22.2%) reviews set restrictions for publication status of the primary studies and 15 (33.3%) did so for the language of publication (limited to English-language reports).

Less than half of the reviews (20, 44.4%) specified *a priori* the type of study designs considered for inclusion. Ten reviews (22.2%) restricted the inclusion to RCTs; approximately half of the reviews (21, 46.7%) allowed the inclusion of observational studies. Thirty-two reviews (71.1%) pre-specified the type of population under study and 34 (75.6%) reviews did so for the type of intervention. The type of outcome was pre-specified in 29 (64.4%) reviews. Less than a half of the reviews (21, 46.7%) reported they assessed the methodological quality of the primary studies. Six (13.3%) made an independent evaluation of the validity of the evidence (e.g. involving at least two independent reviewers). An overall score for quality assessment of primary studies was reported in nine reviews (20.0%). Among those reviews that reported assessing quality, the Jadad scale was the most commonly used approach (five reviews) followed by the Schultz's allocation concealment (three reviews) and the Greenhalgh checklist (two reviews). Other quality assessment approaches included the American College of Physicians Guidelines (ACP), the CONSORT guidelines, Sackett checklist, Crombie checklist, Deeks checklist, American Thoracic Society guidelines (ATS), National Health and Medical Research (NHMR) guidelines, and Council of Australia guidelines on quality of evidence ratings (one review each).

c. Type of analysis:

Twenty (44.4%) reviews statistically pooled the results from primary studies (meta-analysis). The other 25 (55.6%) reviews did not combine the results; however, they provided a qualitative/narrative analysis or, evidence-based tables of the included studies. Among the group of meta-analyses, pooled diagnostic measures (pooled sensitivity/specificity, likelihood ratios) were the commonest summary measures used to calculate the effect of the intervention under study (9 meta-analyses). Other combined measures were pooled odds ratios (4 meta-analyses), effect sizes (2 meta-analyses) and pooled relative risk and weighted mean differences (1 meta-analysis each). Two further meta-analyses used what they reported as "median pooled" measures. One meta-analysis did not clearly report the type of outcome measure. Estimation of 95% confidence intervals around the pooled effect estimate was frequent, with 17 out of 20 meta-analyses reporting this measure of precision.

Half of the meta-analyses calculated a measure of heterogeneity between the individual studies; however, not all of these meta-analyses explored or discussed the potential reasons for heterogeneity: Five of 20 meta-analysis planned subgroup analyses according to certain study-level variables, whereas seven included a sensitivity analysis to explore the robustness of the main findings. Finally, four (8.9%) reviews formally evaluated publication bias.

d. Type of results:

Thirteen of 20 meta-analyses reported a dichotomous primary outcome, which was reported as a statistically significant result in eight of them. Seven of 20 meta-analyses reported a continuous primary outcome, with only three being statistically significant. Overall, 16 (35.6%) reviews clearly concluded the results favoured the intervention under study. Four (8.9%) reviews concluded the intervention under study had negative effects. Fifteen reviews (33.3%) stated the evidence was insufficient to draw conclusions about the effect of the intervention. No clear conclusions were found in the remaining 10 reviews (22.2%).

e. Sources of funding:

No source of funding was declared in the report of 21 (46.7%) reviews. Research agencies (10 reviews) and government grants (9 reviews) were the most frequent source of funding. Internal funds were declared in two reviews, whereas seven reviews declared they did not receive any funding. Industry funds were not declared in any of the reviews. It is unknown whether reviews which did not mention any source of funding were self-funded or simply failed to report it in the manuscript.

3.4.4. Methodological quality of systematic reviews in emergency medicine:

The level of agreement among reviewers regarding both the OQAQ total scores (ICC: 0.83; 95% CI: 0.72, 0.90; n = 45) and the individual OQAQ quality items was good (Table 3.4) with kappa values ranging from 0.55 (OQAQ item # 3) to 0.76 (OQAQ item #7). Consensus discussions regarding for the overall SR quality score did not reveal major divergences in the application of the quality criteria among reviewers. The OQAQ scores obtained by consensus discussions between reviewers were used for the analysis of the SR quality.

A Kolmogorov-Smirnov test was used to assess whether the sample of OQAQ total scores was consistent with a Gaussian population (i.e. test of normality for the distribution of scores). The results showed that data were consistent with a Gaussian distribution (Ho: the data follow a normal distribution; Kolmogorov-Smirnov Z: 1.48, p = 0.25), normality of OQAQ total scores can be assumed; therefore, the mean can be used to describe these data and the OQAQ total scores can be used as a continuous dependent variable in the univariate and multivariate regression analyses.

The mean OQAQ total score for the 2000 – 2004 cohort of SRs was 2.96 points (95% CI: 2.44, 3.48; Median: 2, IQR: 2, 4; n = 45) indicating the reviews have major methodological flaws. When compared to the group of SR reported in the Kelly study [1] (1988-1998, 2.66; 95%CI: 2.09, 3.22; Median: 2; IQR: 2,3; n = 29), no statistically or clinically significant differences in the overall methodological quality of systematic reviews were found (F = 0.58, df = 73; p = 0.44). The direction and magnitude of this result did not significantly change after removal of four reviews published in journals not included in the 1988-1998 cohort of systematic reviews (CJEM [11,12] and EM [13, 14]: F = 1.39, df = 69; p = 0.24). There was no change in the methodological quality of the reviews *after* the introduction of the QUOROM statement (Δ: -0.30; 95%CI: -1.08, 0.48). The overall OQAQ total score for all the systematic reviews published in selected EM journals from 1988 to 2004 was 2.84 (95%CI: 2.46, 3.22; Median = 2, IQR: 2,4; n = 74) indicating a pattern of major methodological flaws.

When individual quality components were analyzed (Table 3.5), most common flaws among 2000 – 2004 SRs were related to failure to document methods to avoid selection bias for the primary studies, and the comprehensiveness of the methods to locate the evidence. Reviews also commonly lacked clear evidence of validity assessment of primary studies using appropriate criteria. Less than half of the reviews reported the methods they used to combine the evidence, supported their conclusions on the data included in the review, and described the criteria to evaluate the validity of primary studies. Authors were marginally better at reporting their search methods and the eligibility criteria for the review. There were no statistically significant differences in the individual OQAQ quality criteria between the two cohorts of reviews, except that SRs published from 2000 to 2004 had almost 4 times higher likelihood of reporting the use of any criteria system to assess the methodological quality of the primary studies (OR: 3.75; 95%CI: 1.19, 11.79).

The univariate analysis (Table 3.6) revealed that the following variables were significantly associated with the OQAQ total quality score of the systematic reviews: Type of analysis, journal of publication, region of corresponding author, and number of authors. The type of main study conclusions were in the limit of statistical significance. No statistically significant associations between the OQAQ total quality score and other study factors, such as declaration of funding, type of research question, topic of study and publication before and after the QUOROM statement, were found. The results can be interpreted as follows:

a) *There is a statistically significant linear association between the OQAQ total score and* the type of journal of publication (IF below and above the median IF). Seventeen percent of variation in the OQAQ total scores is accounted for by the journal of publication. The regression coefficient (β) suggests that reviews published in journals with higher IF ratings also score 1.51 points higher on the OQAQ than those reviews published in journals with lower IFs.

b) There is a statistically significant linear association between the OQAQ total score and the type of analysis used in the review to summarize the results. The type of analysis accounted for 37% of the variation in the OQAQ total scores. Compared to qualitative reviews, the OQAQ total score increased 2 points when a meta-analysis was evaluated (regression coefficient: 2.0; 95%CI: 1.43, 2.67).

c) There is a statistically significant association between the OQAQ total score and the region of the corresponding author. Nine percent of the variation in the OQAQ total score is accounted for by region of the corresponding author. The OQAQ total score was 1.11 points lower when the corresponding author resided in a region other than North America.

d) The number of authors demonstrated a statistically significant association with the OQAQ total score. The number of authors accounted for nine percent of variation in the OQAQ total scores. The OQAQ total score increases by 0.29 points for each further author collaborating on the review.

The results of the multiple regression analysis using a stepwise procedure for modelling showed that the only factor that remained in the model was the type of analysis for data summary. The meta-analytic approach was significantly associated with the overall quality of systematic reviews in the emergency medicine literature after controlling for all other factors in the model (e.g., journal of publication, the region of corresponding author, and the number of authors). In summary, the OQAQ total score increased 2.05 points when the review included a meta-analysis ($p < 0.0001$). There was no need to test for interactions among the independent variables as only a single predictor was found to be significant in the multiple regression model.

3.5. Discussion

This study analyzed the methodological quality of systematic reviews and/or meta-analyses published in major emergency medicine journals between January 2000 and March 2004, and compared the results with the quality of reviews of a similar cohort of reviews in emergency medicine journals published before the QUOROM guidelines launch in 1999. The rate of SR publications is clearly increasing (45 in 2000-2004 vs. 29 in the preceding 10 years), and while almost half of the reviews addressed questions regarding therapy (22.2%), the number of reviews addressing diagnostic interventions has increased (14%). Most of the reviews focused on a single research question; however, a considerable proportion of them examined multiple questions in the same review. Almost half of the reviews did not declare the source of funding.

3.5.1. Methodological quality of the reviews:

There is evidence to suggest that many of the reviews evaluated in this study exhibit major methodological flaws, limiting the validity of their conclusions. Moreover, there is limited evidence of an improvement in the conduct and reporting of these reviews following the release of the

QUOROM statement. Only four (8.8%) of the 2000-2004 reviews satisfied all the criteria of the OQAQ, whereas 17.7% were given the lowest possible score. These results are similar to the findings for the 1988-1998 cohort of systematic reviews and/or meta-analyses in emergency medicine [1], where none of the reviews satisfied all the OQAQ criteria and 20.6% were given the lowest score. The findings are consistent with those of Jadad et al. [15] who reported that 90% of 80 meta-analyses of analgesic interventions exhibited flaws that limited their claims. They are also similar to the results of Bhandari et al [16] where only 10% of 40 meta-analyses in orthopedics satisfied all the categories of the OQAQ, with 13% given the lowest possible score.

Although there appears to be a trend of improvement in the level of quality since the QUOROM statement was released, no statistically or clinically significant differences in the overall methodological quality of systematic reviews were found between both cohorts of systematic reviews and/or meta-analyses in emergency medicine. The change in the methodological quality of reviews from 1988-1998 to 2000-2004 (Δ: -0.30; 95%CI: -1.08, 0.48) indicate that neither clinically or statistically significant improvements in methodological quality occurred following the introduction of the QUOROM guidelines. Moreover, the upper limit of change (+0.48) is below the *post-hoc* MCID of 1.0. To our knowledge, this review is the first to examine the effect of the QUOROM statement on the quality of reporting of SRs, albeit limited to the emergency medicine literature. Despite the development of this important document on standards of reporting, emergency medicine journals appear to have ignored it.

The assessment of individual quality components identified areas where improvements should be performed. The aim of a systematic review is to provide a comprehensive, unbiased summary of current research evidence. In order to achieve this, a transparent and inclusive search should be done. One of the major flaws of the systematic reviews under study was the lack of comprehensive methods to locate the evidence. Although searches in MEDLINE were part of the search strategies of almost the entire sample of reviews (95.6%), searches in other electronic and non-electronic sources were less frequent. For example, the search for unpublished material was barely reported and a considerable proportion of reviews (33.3%) restricted the search for language of publication (limited to English-language reports). These findings are similar to those from evaluations of the quality of systematic reviews in other research areas [17-19]. By excluding unpublished and non-English language articles, the results of a systematic review are susceptible to publication biases and may present a misleading picture of the effects of an intervention [20].

There is good evidence that research findings showing statistically significant results are more likely to be submitted and accepted for publication [21], and more likely to be published in English language journals [22]. It has also been reported that the exclusion of grey literature may exaggerate the estimates of the intervention's effects by 15%-38% depending on the type of grey literature [23]. It was disappointing to find that only 11.1% of the systematic reviews included in this evaluation searched for both published and unpublished data with no language restrictions.

Emergency medicine systematic reviews frequently failed to document how bias in the selection of primary studies was avoided. Similar findings have been reported in other overviews of reviews [1,17,24-27]. Reviews also regularly failed to assess the validity of primary studies using appropriate criteria. Less than half of the reviews reported they assessed the methodological quality of the primary studies; the Jadad scale was the most commonly used quality tool. This finding is comparable to those from overviews in a variety of medical and allied research areas [1,17-19,24-33].

Less than half of the reviews reported the methods they used to combine the evidence. Similar results have been found in other studies [18,27,30-32], where deficiencies were frequent for the summary of the results of primary studies in the reviews. Exceptions are the studies of Bramwell et al. [17] in a group of systematic reviews in oncology and Moher et al. [26] in the field of paediatric complementary and alternative medicine, where authors were particularly good at combining studies for qualitative and/or quantitative analysis and basing their conclusions on the data included in the reviews.

The univariate analysis revealed a number of potentially important factors that might contribute to the quality of reviews. The number of authors was one of these variables. Bramwell et al [17] did not find a relationship between quality score (using the Sacks checklist) and the number of authors in the review. Differences between the findings from the univariate analysis and the findings of the Bramwell et al. [17] study may be related to differences in the instruments used for quality assessment

The association between the OQAQ total score and type of conclusions (positive, negative/uncertain) was in the limit of statistical significance. This latter null result is in conflict with the results from similar studies. For example, Jadad et al. [15], Bhandari et al. [16] and Furlan et al. [34] have reported that the distribution of the OQAQ overall scores between systematic reviews and/or meta-analyses with positive conclusions and those with negative results is different and those of high quality were also less likely to produce positive results, whereas reviews with lower OQAQ overall scores tended to report positive findings. On the contrary, Assendelfelt et al. [35] assessed 51 reviews using a modified version of the OQAQ and found that reviews with relatively high methodological quality had a positive conclusion. Based on these current results and those contradictory ones from similar studies, no definite conclusions can be reached regarding the direction of the association between the level of quality and the type of conclusions of the reviews and/or meta-analyses.

The results of the multiple regression analysis also showed that the type of analysis for data summary significantly helps to predict the overall quality of systematic reviews in the emergency medicine literature. The multiple regression analysis of Bhandari et al. [16] showed that the most important predictors of the quality of a meta-analysis were the affiliation with an epidemiology department and journal type.

There may be two plausible interpretations for the importance of the "type of analysis" identified in this study (that is to say, reviews including a meta-analysis for data summary are more likely to be of higher methodological quality). First, it may be a true finding and the nature of meta-analytic procedures may explain the reduction in the likelihood of bias in the reviews [30]. On the other hand, the result could be an artefact related to the type of instrument that was used to evaluate the methodological quality of the reviews. More precisely, a meta-analysis could score higher due to problems of interpretation of items #7 and #8 of the OQAQ if the "methods to combine the findings" are interpreted as equivalent of the numerical pooling of results. Therefore, the superiority of meta-analyses would be spurious and their quality over-estimated when the OQAQ is applied. As mentioned above, the problem could be more a matter of interpretation of the items in question than a construct weakness of the items by themselves. As Oxman [5] recognizes, "it does not imply that a narrative review cannot be scientifically rigorous and that a reviewer might, for example, choose not to use quantitative techniques to combine study findings for sound reasons and still be explicit about what was done. Nor does it imply that meta-analyses are always scientifically sound".

3.5.2. Strengths and weaknesses of this study:

This is the first study to compare how the methodological quality of systematic reviews in the medical and allied literature has evolved before and after the publication of the QUOROM statement to improve the reporting and methodological quality of systematic reviews and/or meta-analyses. A comprehensive search of the literature was performed and it is unlikely that potentially relevant systematic reviews and/or meta-analyses were omitted. Although the findings cannot be generalized to all reviews published by emergency physicians in any journal, it was felt that the evaluation undertaken reflects the quality of systematic reviews that are easily accessible to most emergency physicians. The evaluation, however, was confined to English-language publications in a sample of EM journals and it is unknown whether there may be a publication bias against systematic reviews in emergency medicine that do not have significant findings.

The methods used to evaluate the systematic reviews and/or meta-analyses could be criticized; however, there is no consensus regarding what is the best scoring system to assess the methodological quality of reviews. The OQAQ index was chosen due to its simplicity and extensive clinimetric development and validation. It has been the most widely used review of reviews tool, and recommended as the first-choice instrument to assess the methodological quality of systematic reviews and/or meta-analyses [30]. Furthermore, the OQAQ accords well with the QUOROM statement. We do, however, recognize its limitation and bias against qualitative reviews.

Another potential limitation is that one of the investigators involved in the quality assessment was the primary or collaborative author on some of the systematic reviews under study. However, efforts were made to minimize the influence of personal bias when assessing the quality of systematic reviews through a process that involved an independent evaluation by two investigators and the need to reach a consensus regarding both the individual components and overall quality score. Moreover, the level of agreement was high between reviewers and no systematic error was detected when the reviewer assessed his own work.

Another limitation of this study is common to other evaluations of the quality of scientific research: The quality of the systematic reviews was assessed by examining a published report of the review, instead of evaluating the conduct of the review itself. It must be accepted that a review's report may reveal less about how the review was conducted than it does about the reviewers' ability to write well. The evidence from sparse research addressing the quality of reporting versus the quality of conduct, however, have found a reasonably good correlation between how investigators conduct their research and how it is subsequently reported [36,37].

3.6. Conclusions and recommendations

Systematic reviews have an important role to play in evidence-based emergency medicine, and the preparation of a larger number of systematic reviews is an important challenge for the emergency medicine research community. Their usefulness, however, is entirely dependent upon their validity. This study showed that the quality of systematic reviews and/or meta-analyses in emergency medicine continue to be unacceptably low despite repeated calls for improvement by EBEM groups [1,38-43] and the dissemination of the QUOROM guidelines. Although systematic reviews appear easy superficially, those who have been involved in their production recognize that a high-quality systematic review is extremely difficult to complete and requires resources, time and dedication on the part of many. Most of the methodological deficiencies of systematic reviews and/or meta-analyses published in emergency medicine research could be corrected easily in the future by ensuring that investigators address sharply defined clinical questions and adhere to the issues raised by each item in the OQAQ scoring system and in the QUOROM statement.

Quality of reviews may also improve if emergency medicine journals provide authors with explicit instructions to report systematic reviews and/or meta-analyses. Finally, readers of emergency medicine journals should also be aware of the methodological limitations of systematic reviews and/or meta-analysis and critically appraise their results before implementing their results in clinical practice.

Table 3.1: List of reviews excluded from the study of the quality of systematic reviews in the emergency medicine literature.

STUDY ID	REASONS FOR EXCLUSION
Abu-Laban RB, 2001 [44]	Narrative review.
Af Geijerstam J-L, 2004 [45]	No identifiable question was formulated.
American College of Emergency Physicians 2001 [46]	Clinical Practice Guidelines.
American College of Emergency Physicians, 2000 [47]	Clinical Practice Guidelines.
American College of Emergency Physicians, 2000 [48]	Clinical Practice Guidelines.
American College of Emergency Physicians, 2002 [49]	Clinical Practice Guidelines.
American College of Emergency Physicians, 2003 [50]	Clinical Practice Guidelines.
American College of Emergency Physicians, 2002 [51]	Clinical Practice Guidelines.
Aminzadeh F, 2002 [52]	No identifiable question was formulated.
Babcock ICh, 2000 [53]	EBEM shortcut reviews.
Bernstein SL, 2002 [54]	No identifiable question was formulated.
Blackman K, 2001 [55]	No identifiable question was formulated.
Bond GR, 2002 [56]	Narrative review.
Borland ML, 2002 [57]	No identifiable question was formulated.
Boudreaux ED, 2002 [58]	No identifiable question was formulated.
Braksiek RJ, 2002 [59]	No identifiable question was formulated.
Dart DC, 2000 [60]	Clinical Practice Guidelines.
D'Onofrio G, 2002 [61]	Narrative review.
Edlow JA 2000 [62]	EBEM shortcut review.
Fermann GJ, 2002 [63]	No identifiable question was formulated.
Franc-Law JM, 2000 [64]	Case report + Literature review.
Frank JR, 2002 [65]	EBEM shortcut review.
Gallagher EJ, 2002 [66]	Editorial.
Graham CA 2004 [67]	No identifiable question was formulated.
Hsieh M, 2002 [68]	EBEM shortcut review.
Inamasu J, 2002 [69]	No identifiable question was formulated.
Jagoda AS, 2002 [70]	Clinical Practice Guidelines.
Kao LW, 2003 [71]	Narrative review.
Kuhn G 2001 [72]	No identifiable question was formulated.
Kuhn GJ, 2002 [73]	No identifiable question was formulated.
Lew M 2003 [74]	No identifiable question was formulated.
Li SF, 2002 [75]	Case report + Literature review.
Maas Cortes L, 2001 [76]	No identifiable question was formulated.
Maio RF, 2002 [77]	No identifiable question was formulated.
May CR, 2001 [78]	No identifiable question was formulated.
McClune T, 2002 [79]	Narrative review.
McClune T, 2003 [80]	No identifiable question was formulated.
Monaghan M, 2000 [81]	Narrative review.
Nunnink L, 2002 [82]	No identifiable question was formulated.

Table 3.1: List of reviews excluded from the study of the quality of systematic reviews in the emergency medicine literature (Cont').

STUDY ID	REASONS FOR EXCLUSION
Rhine DJ, 2000 [83]	Narrative review.
Rothman RE, 2003 [84]	Narrative review.
Schull MJ, 2002 [85]	No identifiable question was formulated.
Tran T, 2000 [86]	No identifiable question was formulated.
Warden CR, 2003 [87]	Clinical Practice Guidelines.
Weaver CS, 2003 [88]	EBEM shortcut review.
Weigand JV, 2001 [89]	No identifiable question was formulated.
Wilber ST, 2003 [90]	No identifiable question was formulated.
Yeung JK, 2002 [91]	No identifiable question was formulated.

Table 3.2 Characteristics of reviews included in the study of the quality of systematic reviews in the emergency medicine literature.
MA = Meta-analysis; Dx: Diagnosis; Tx: Therapy; HSE: Health Services Evaluation; EMJ: Emergency Medicine Journal; AOEM: Annals of Emergency Medicine; JEM: Journal of Emergency Medicine; AEM: Academic Emergency Medicine; CJEM: Canadian Journal of Emergency Medicine; EM: Emergency Medicine; AJEM: American Journal of Emergency Medicine

REFERENCE	COUNTRY	TOPIC	JOURNAL	TYPE OF ANALYSIS	NUMBER OF RESEARCH QUESTIONS	AUTHORS CONCLUSIONS	FUNDING	OVERALL QUALITY (OQAQ)
Ahmad 2004 [92]	UK	Dx	EMJ	MA	Single	Positive	None	4
Alter 2000 [93]	USA	Tx	AOEM	MA	Single	Positive	Unknown	4
Arnold 2004 [94]	USA	Epidemiology	AOEM	MA	Single	Neutral	None	3
Balk 2001 [95]	USA	Dx	AOEM	MA	Multiple	Negative	None	4
Batchelor 2002 [96]	UK	Risk factors	EMJ	MA	Single	Positive	Unknown	2
Boudreaux 2004 [97]	USA	HSE	JEM	Qualitative	Multiple	Uncertain	Unknown	4
Brown 2002 [98]	USA	Dx	AOEM	MA	Single	Positive	Unknown	7
Bush 2002 [99]	UK	Tx	EMJ	Qualitative	Single	Uncertain	Unknown	2
Cooke 2003 [100]	UK	HSE	EMJ	Qualitative	Single	Uncertain	Unknown	1
Cueto 2001 [101]	USA	Dx	JEM	MA	Single	Positive	Unknown	2
Davies 2003 [13]	Australia	Tx	EM	Qualitative	Single	Positive	Unknown	1
Eddleston 2003 [102]	Sri Lanka	Tx	AOEM	Qualitative	Single	Uncertain	None	3
Edmonds 2002 [103]	Canada	Tx	AOEM	MA	Single	Positive	Declared	4
Farion 2003 [104]	Canada	Tx	AEM	Qualitative	Single	Positive	None	7
Fredriksson 2003 [105]	Canada	HSE	AJEM	Qualitative	Single	Uncertain	Unknown	2
Gabbe 2003 [14]	Australia	HSE	EM	Qualitative	Multiple	Uncertain	Declared	1
Glavan 2003 [106]	USA	Measurement	AEM	MA	Single	Uncertain	Unknown	1
Goodacre 2000 [107]	UK	HSE	EMJ	Qualitative	Single	Positive	Unknown	3
Graham 2004 [67]	UK	Education	EMJ	Qualitative	Multiple	Uncertain	Declared	2
Ionnadis 2001 [108]	USA	Tx & Dx	AOEM	MA	Single	Positive	None	4
Ionnadis 2001 [109]	USA	Dx	AOEM	MA	Single	Positive	None	4
Kline 2000 [110]	USA	Dx	AOEM	MA	Multiple	Positive	None	4
Lau 2001 [111]	USA	Dx	AOEM	MA	Multiple	Uncertain	None	4

Table 3.2 Characteristics of reviews included in the study of the quality of systematic reviews in the emergency medicine literature (Cont').

REFERENCE	COUNTRY	TOPIC	JOURNAL	TYPE OF ANALYSIS	NUMBER OF RESEARCH QUESTIONS	AUTHORS CONCLUSIONS	FUNDING	OVERALL QUALITY (OQAQ)
Lavine 2001 (11)	Canada	Tx	CJEM	Qualitative	Single	Uncertain	Unknown	1
Leary 2003 (112)	USA	Tx	AOEM	Qualitative	Multiple	Uncertain	Declared	1
Li 2001 (113)	USA	Dx	JEM	MA	Single	Neutral	Unknown	2
Marill 2001 (114)	USA	Tx	AEM	MA	Single	Uncertain	Unknown	3
Mattick 2002 (115)	UK	Tx	EMJ	Qualitative	Single	Positive	None	1
McCusker 2003 (116)	Canada	HSE	AEM	Qualitative	Single	Uncertain	Declared	2
Murphy 2001 (117)	UK	Tx	EMJ	Qualitative	Multiple	Uncertain	Unknown	1
Perry 2002 (118)	Canada	Tx	AEM	MA	Single	Negative	None	7
Quin 2000 (119)	UK	Tx	EMJ	Qualitative	Multiple	Uncertain	None	2
Roberts 2003 (120)	UK	Tx	EMJ	Qualitative	Single	Uncertain	Declared	1
Robinson 2000 (121)	UK	Dx	EMJ	Qualitative	Single	Positive	Unknown	3
Robinson 2001 (122)	UK	Tx	EMJ	Qualitative	Single	Uncertain	Unknown	3
Rodrigo 2000 (123)	Uruguay	Tx	AJEM	MA	Single	Negative	Unknown	5
Rowe 2000 (124)	Canada	Tx	AOEM	MA	Single	Uncertain	Declared	6
Smith 2003 (125)	UK	Tx	EMJ	Qualitative	Single	Uncertain	Declared	2
Snooks 2004 (126)	UK	HSE	EMJ	Qualitative	Multiple	Uncertain	Declared	1
Trout 2000 (127)	USA	Measurement	AEM	Qualitative	Single	Positive	Unknown	3
Trzeciak 2003 (128)	USA	HSE	EMJ	Qualitative	Multiple	Uncertain	None	1
Vickery 2001 (129)	UK	Tx	EMJ	Qualitative	Multiple	Negative	Unknown	2
Wilbur 2001 (12)	Canada	Tx	CJEM	Qualitative	Single	Uncertain	Unknown	1
Worster 2002 (130)	Canada	Dx	AOEM	MA	Single	Positive	None	7
Yildiz 2003 (131)	Turkey	Tx	EMJ	Qualitative	Multiple	Uncertain	None	2

Table 3.3: Publication of systematic reviews in emergency medicine journals by year (2000-2004).

Year	Frequency	Percent (%)	Cumulative percent
2000	8	17.8	17.8
2001	12	26.7	44.4
2002	7	15.6	60.0
2003	12	26.7	86.7
2004	6	13.3	100
TOTAL	45	100%	

Table 3.4: Inter-observer agreement for individual components of the OQAQ.

OQAQ Item	Inter-observer agreement [Kappa (95% CI)]
1	0.56 (0.31 – 0.81)
2	0.69 (0.51 – 0.87)
3	0.55 (0.36 – 0.74)
4	0.64 (0.44 – 0.84)
5	0.67 (0.48 – 0.86)
6	0.74 (0.55 – 0.93)
7	0.76 (0.59 – 0.93)
8	0.58 (0.39 – 0.77)
9	0.67 (0.49 – 0.85)

Table 3.5: Quality of reports of systematic reviews published in emergency medicine journals from 1988-1998 and 2000-2004.

Question	1988 – 1998 n (%)	2000-2004 ‡ n (%)		Total n (%)‡	P-value
	n = 29	n = 41	n = 45	n = 70	
1. Search method stated	16 (55.2)	28 (62.2)	26 (63.4)	42 (60.0)	0.488
2. Search comprehensive	5 (17.2)	12 (26.7)	12 (29.3)	17 (24.3)	0.248
3. Inclusion criteria reported	20 (69.0)	20 (44.4)	19 (46.3)	39 (55.7)	0.060
4. Selection bias avoided	6 (20.7)	8 (17.8)	8 (19.5)	14 (20.0)	0.903
5. Validity criteria reported	**5 (17.2)**	**18 (40.0)**	**18 (43.9)**	**23 (32.9)**	**0.019***
6. Validity assessed appropriately	7 (24.1)	13 (28.9)	13 (31.7)	20 (28.6)	0.490
7. Combining methods reported	14 (48.3)	17 (37.8)	17 (41.5)	31 (44.3)	0.572
8. Finding combined appropriately	11 (37.9)	18 (40.0)	18 (43.9	29 (41.4)	0.617
9. Conclusions supported by data	15 (51.7)	17 (37.8)	16 (39.0)	31 (44.3)	0.292

‡: Left column of 2000-2004 Group exclude four reviews published in journals not included in the 1988-1998 cohort of systematic reviews. They were excluded from the 1988 to 1998 – 2000 to 2004 comparison.
* Statistically significant at 5% level.

Table 3.6: Predictors of quality of systematic reviews / meta-analyses in emergency medicine.

Variable	Unadjusted regression coefficient (95%CI)	Adjusted regression coefficient (95%CI)
Type of analysis [D] (qualitative, meta-analysis) *	2.05 (1.43, 2.67)	**2.05 (1.43 – 2.67)**
Journal of publication [D] (above/below the IF median: 0.743) *	-1.51 (-2.29, -0.74)	‡
Region of corresponding author [D] (North America, other) *	-1.11 (-1.93, -0.29)	‡
Number of authors [C]*	0.29 (0.07, 0.51) *	
Main study conclusions [D] (positive, negative/uncertain)	-0.81 (-1.6, 0.00)	NA
Funding [D] (declared, not declared/unknown)	-0.62 (-1.46, 0.21)	NA
Type of research question [D] (single, multiple)	-0.59 (-1.47, 0.28)	NA
Publication before and after QUOROM statement[D]	0.21 (-5.95, 1.01)	NA
Topic of study [D] (intervention, diagnosis/other)	-0.04 (-0.84, 0.74)	NA

[C] = continuous variable; [D] = Dichotomous variable
*: Variables included in the multivariate analysis ($p<0.2$).
‡: Removed from the multivariate model.
NA = Not applicable as not included in the multivariate analysis

Figure 3.1 Flow diagram for selection of systematic reviews and/or meta-analyses published in major emergency medicine journals from 2000 to 2004.

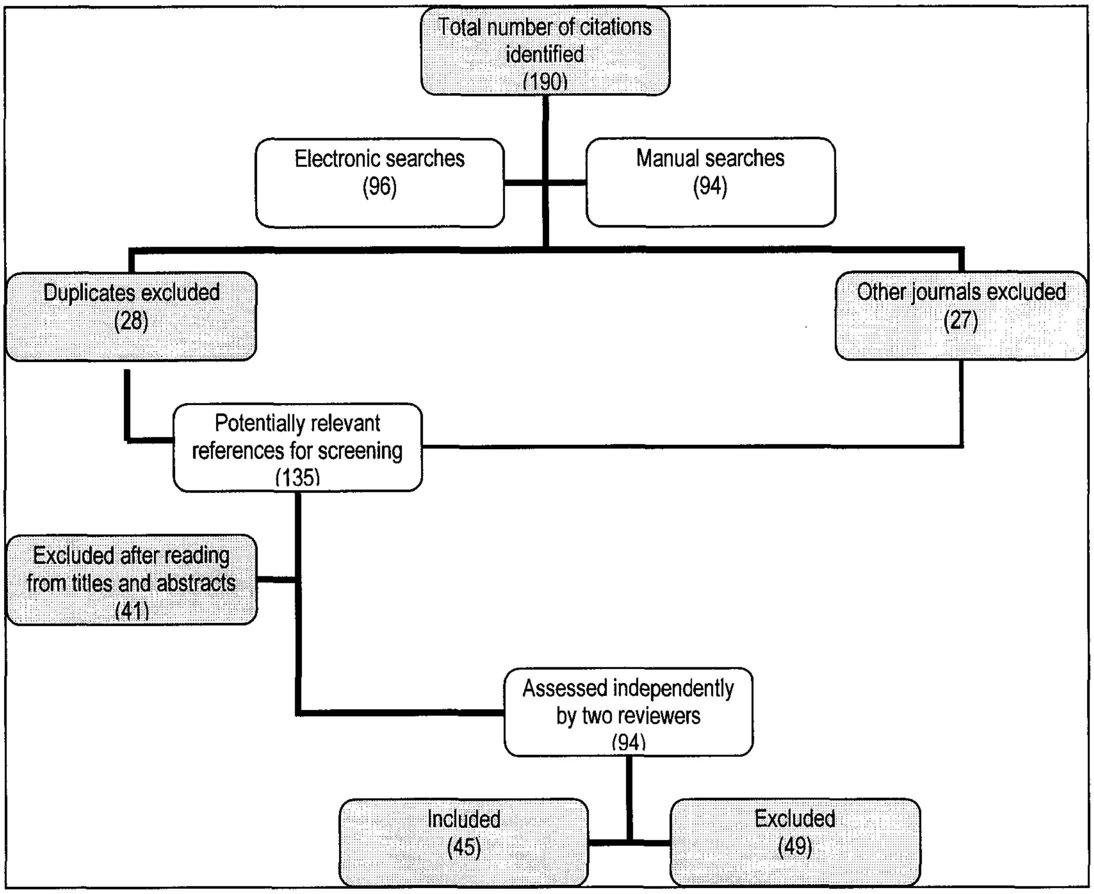
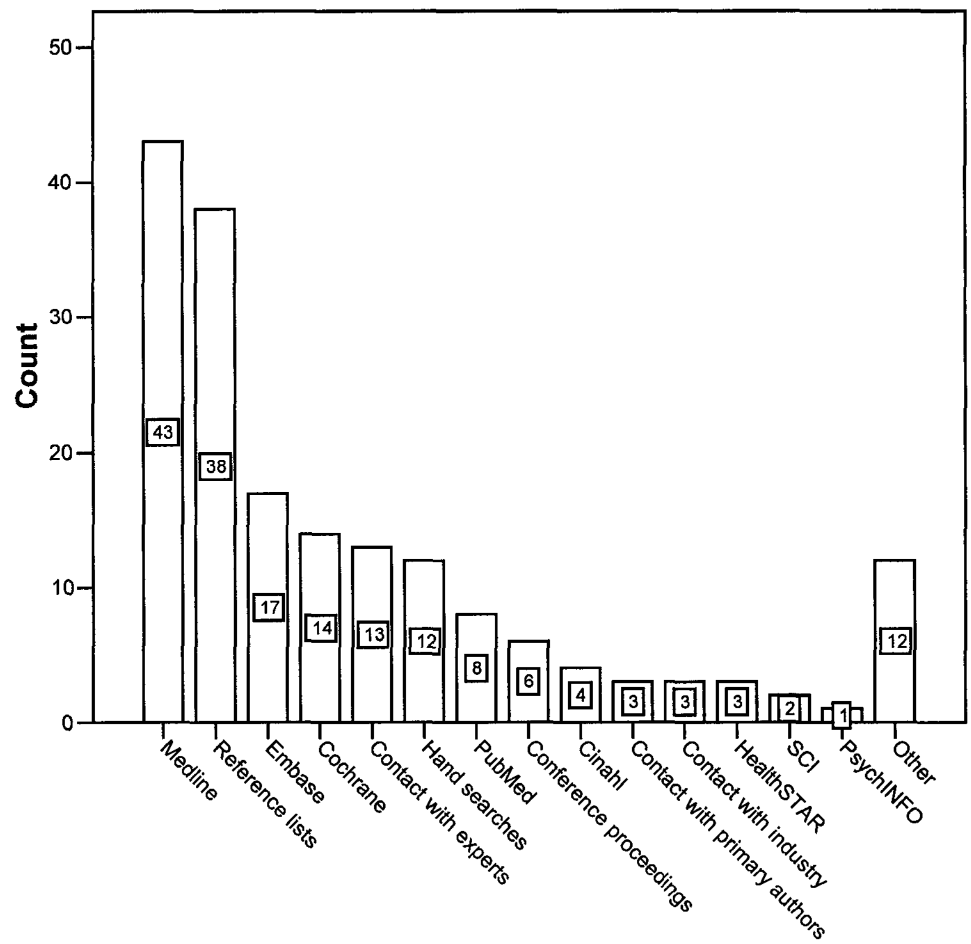


Figure 3.2: Sources of information used to identify potential studies for inclusion in the emergency medicine reviews.



Other include: *National Institutes of Health database, Lilacs, Current Contents, Register of the Medical Editors' Trial Amnesty, Best Evidence database, Doctors Net UK, Dogpile, Google, ClinPSYCH, AMED, PILOTS, Effective Health Care Bulletins, Effective Matters, Health Evidence Bulletins Wales, BIDS, Healthplan, Helmis, DARE.*

3.7. References

(1) Kelly KD, Travers A, Dorgan M, Slater L, Rowe BH. Evaluating the quality of systematic reviews in the emergency medicine literature. Ann Emerg Med 2001; 38:518-526.

(2) Pettiti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis methods for quantitative synthesis in medicine. New York: Oxford University Press, 1994.

(3) Egger M, Smith GD, O'Rourke K. Systematic reviews in health care: meta-analysis in context. In: Egger M, Smith GD, Altman DG, editors. London: BMJ Publishing Group, 2001.

(4) ISI. 2002 JCR Science Edition. ISI Web of Science. 2004. Accessed: 7-5-2004.

(5) Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991; 44:91-98.

(6) Oxman AD, Cook DJ, Guyatt GH. User's Guide to the Medical Literature: How to Use an Overview. JAMA 1994; 272:1367-1371.

(7) Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44:1271-1278.

(8) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174.

(9) Seigel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. Am J Epidemiol 1992; 135:571-578.

(10) Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford, New York, Tokyo: Oxford University Press, 1989.

(11) Lavine E, Scolnik D. Lack of efficacy of humidification in the treatment of croup: why do physicians persist in using an unproven modality? CJEM 2001; 3: 209-212.

(12) Wilbur K, Zed PJ. Is propofol an optimal agent for procedural sedation and rapid sequence intubation in the emergency department? CJEM 2001; 3:302-310.

(13) Davies RJ. Buffering the pain of local anaesthetics: a systematic review. Emerg Med 2003; 15:81-88.

(14) Gabbe BJ, Cameron PA, Finch CF. The status of the Glasgow Coma Scale. Emerg Med 2003; 15:353-360.

(15) Jadad AR, McQuay H. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. J Clin Epidemiol 1996; 49:235-243.

(16) Bhandari M, Morrow F, Kulkarni AV, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. J Bone Joint Surg Am 2001; 83-A:15-24.

(17) Bramwell VHC, Williams CJ. Do authors of review articles use systematic methods to identify, assess and synthesize information? Ann Oncol 1997; 8:1185-1195.

(18) Glenny AM, Esposito M, Coulthard P, Wotyhington HV. The assessment of systematic reviews in dentistry. Eur J Oral Sci 2003; 111:85-92.

(19) Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. JAMA 2002; 287:2809-2812.

(20) Moher D, Schachter HM. Potential solutions to the problem of conducting systematic reviews of new health technologies. CMAJ 2002; 25:1674-1675.

(21) Scherer R, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts: a meta-analysis. JAMA 1994; 272:158-162.

(22) Egger M, Zellweger-Zahner T, Scheider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997; 350:326-329.

(23) McAuley L, Pham B. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet 2000; 356:1228-1231.

(24) Christensen E. Quality of reporting of meta-analyses: the QUOROM statement. Will it help? J Hepatol 2001; 34:342-345.

(25) Jadad AR, Moher M, Browman GP, Lynda Booker, Sigouin C, Fuentes M et al. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000; 320:537-540.

(26) Moher D, Soeken K, Sampson M, Ben-Porat L, Berman B. Assesing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. BMC Pediatr 2002; 2:3.

(27) Oosterhuis WP, Niessen RWLM, Bossuyt PMM. The science of systematic reviewing studies of diagnostic tests. Clin Chem Lab Med 2000; 38:577-588.

(28) Walter SD, Jadad AR. Meta-analysis of screening data: a survey of the literature. Stat Med 1999; 18:3409-3424.

(29) Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120:667-676.

(30) Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M et al. Methodology and reports of systematic reviews and meta-analyses: comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998; 280:278-280.

(31) Olsen O, Middleton P, Ezzo J, Gotzsche P, Hadhazy V, Herxheimer S et al. Quality of Cochrane reviews: assessment of a sample from 1998. BMJ 2001; 323:829-832.

(32) van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine 2003; 28:1290-1299.

(33) Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). Int J Technol Assess Health Care 1999; 15:671-678.

(34) Furlan A, Clarke J, Esmail R, Sinclair S, Irvin E, Bombardier C. A critical review of reviews on the treatment of chronic low back pain. Spine 2001; 26:E155-E162.

(35) Assendelft WJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. JAMA 1995; 274:1942-1948.

(36) Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. J Clin Oncol 1986; 4:942-951.

(37) Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309:597-599.

(38) Dick WF. [Evidence-based emergency medicine][Article in German]. Anaesthesist 1998; 47:957-967.

(39) Emond SD, Wyer PC, Brown MD, Cordell WH, Spooner CH, Rowe BH. How relevant are the systematic reviews in the cochrane library to emergency medical practice? Ann Emerg Med 2002; 39:153-158.

(40) Wyatt J, Bellis F. The cochrane library is a relevant source for emergency medicine. Emerg Med J 2002; 19:436.

(41) Wyer PC, Rowe BH, Guyatt GH. Evidence-based emergency medicine. The clinician and the medical literature: when can we take a shortcut? Ann Emerg Med 2000; 36:149-155.

(42) Rowe BH, Klassen T, Wyer PC. Evidence-based emergency medicine/perspectives. One is the only number that you'll ever need! Ann Emerg Med 2000; 36:520-523.

(43) Schriger DL. Evidence-based emergency medicine/perspectives. One is the loneliest number: be skeptical of evidence summaries based on limited literature reviews. Ann Emerg Med 2000; 36:517-519.

(44) Abu-Laban RB, Zed PJ, Roy A, Evans KG. Severe methemoglobinemia from topical anesthetic spray: report, discussion and qualitative systematic review. CJEM 2001; 3:51-56.

(45) Af Geijerstam J-L, Britton M, Marke L-A. Mild head injury: observation or computed tomography? Economic aspects by literature review and decision analysis. Emerg Med J 2004; 21:54-58.

(46) American College of Emergency Physicians. Clinical policy for the management and risk stratification of community-acquired pneumonia in adults in the emergency department. Ann Emerg Med 2001; 38:107-113.

(47) American College of Emergency Physicians. Clinical policy: critical issues for the initial evaluation and management of patients presenting with a chief complaint of nontraumatic acute abdominal pain. Ann Emerg Med 2000; 36:406-415.

(48) American College of Emergency Physicians. Clinical policy: critical issues in the evaluation and management of adult patients presenting with suspected acute myocardial infarction or unstable angina. Ann Emerg Med 2000; 35:521-544.

(49) American College of Emergency Physicians. Clinical policy: critical issues in the evaluation and management of patients presenting to the emergency department with acute headache. Ann Emerg Med 2002; 39:108-122.

(50) American College of Emergency Physicians. Clinical policy: critical issues in the initial evaluation and management of patients presenting to the emergency department in early pregnancy. Ann Emerg Med 2003; 41:123-133.

(51) American College of Emergency Physicians. Clinical policy: critical issues in the evaluation and management of adult patients presenting with suspected pulmonary embolism. Ann Emerg Med 2003; 41:257-270.

(52) Aminzadeh F, Dalziel WB. Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. Ann Emerg Med 2002; 39:238-247.

(53) Babcock Irvin Ch, Wyer PC, Gerson LW, for the Society for Acad Emerg Med Public Health and Education Task Force Preventive Services Work Group. Clinical preventive services - an emergency medicine evidence-based review. Acad Emerg Med 2000; 7:1042-1054.

(54) Bernstein SL, Becker BM. Diagnosis and management of smoking and smoking-related illness in the emergency department: a systematic review. Acad Emerg Med 2002; 9:720-729.

(55) Blackman K, Brown SGA, Wilkes GJ. Plasma alkalinization for tryciclic antidepressant toxicity: a systematic review. Emerg Med 2001; 13:204-210.

(56) Bond GR. The role of activated charcoal and gastric emptying in gastrointestinal decontamination: a state-of-the-art review. Ann Emerg Med 2002; 39:273-286.

(57) Borland ML, Jacobs I, Rogers IR. Options in prehospital analgesia. Emerg Med 2002; 14:77-84.

(58) Boudreaux ED, Francis JL, Loyacano T. Family presence during invasive procedures and resuscitations in the emergency department: a critical review and suggestions for future research. Ann Emerg Med 2002; 40:193-205.

(59) Braksiek RJ, Roberts DJ. Amusement park injuries and deaths. Ann Emerg Med 2002; 39:65-72.

(60) Dart DC, Goldfrank LR, Chyka PA, Lotzer D, Woolf AD, McNally J et al. Combined evidence-based literature analysis and consensus guidelines for stocking of emergency antidotes in the United States. Ann Emerg Med 2000; 36:126-132.

(61) D'Onofrio G, Degutis LC. Screening and brief intervention for alcohol problems in the emergency department: a systematic review. Acad Emerg Med 2002; 9:627-638.

(62) Edlow JA, Wyer PC. How good is negative cranial computed tomographic scan result in excluding subarachnoid hemorrhage. Ann Emerg Med 2000; 36:507-516.

(63) Fermann GJ, Suyama J. Point of care testing in the emergency department. J Emerg Med 2002; 22:393-404.

(64) Franc-Law JM, Rossignol M, Vernec A, Somogyi D, Shrier I. Poisoning-induced acute atraumatic compartment syndrome. Am J Emerg Med 2000; 18:616-621.

(65) Frank JR, Ovens H. Shiftwork and emergency medical practice. CJEM 2002; 4:421-428.

(66) Gallagher EJ. How well do clinical practice guidelines guide clinical practice? Ann Emerg Med 2002; 40:394-398.

(67) Graham CA. Advanced airway management in the emergency department: what are the training and skills maintenance needs for UK emergency physicians? Emerg Med J 2004; 21:14-19.

(68) Hsieh M, Auble TE, Yealy DM. Predicting the future: can this patient with acute congestive heart failure be safely discharged from the emergency department? Ann Emerg Med 2002; 39:181-189.

(69) Inamasu J, Ichikizaki K. Mild hypothermia in neurologic emergency: an update. Ann Emerg Med 2002; 40:220-230.

(70) Jagoda AS, Cantrill SV, Wears RL, Valadka A, Gallagher EJ, Gottesfeld SH et al. Clinical policy: neuroimaging and decisionmaking in adult mild traumatic brain injury in the acute setting. Ann Emerg Med 2002; 40:231-249.

(71) Kao LW, Kirk MA, Evers SJ, Rosenfeld SH. Droperidol, QT prolongation, and sudden death: what is the evidence? Ann Emerg Med 2003; 41:546-558.

(72) Kuhn G. Circadian rythm, shift work, and emergency medicine. Ann Emerg Med 2001; 37:88-98.

(73) Kuhn GJ. Diagnostic errors. Acad Emerg Med 2002; 9:740-750.

(74) Lew M, Klonis E. Emergency management of eclampsia and severe pre-eclampsia. Emerg Med 2003; 15:361-368.

(75) Li SF, Ender K. Toothpick injury mimicking renal colic: case report and systematic review. J Emerg Med 2002; 23:35-38.

(76) Maas Cortes L, Hargarten SW. A systematic literature review on emergency department-based interventions that address smoke detectors in home. Acad Emerg Med 2001; 8:925-929.

(77) Maio RF, Garrison HG, Spaite DW, Desmond JS, Gregor MA, Stiell IG et al. Emergency Medical Services Outcomes Project (EMSOP) IV: pain measurement in out-of-hospital outcomes research. Ann Emerg Med 2002; 40:172-179.

(78) May CR. Management of venous thromboembolic disease in the lower limb. Emerg Med 2001; 13:211-223.

(79) McClune T, Burton AK, Waddell G. Whiplash associated disorders: a review of the literature to guide patient information and advice. Emerg Med J 2002; 19:499-506.

(80) McClune T, Burton AK, Waddell G. Evaluation of an evidence based patient educational booklet for management of whiplash associated disorders. Emerg Med J 2003; 20:514-517.

(81) Monaghan M. Intravenous lignocaine as pretreatment for intubation of patients with acute severe head injury. Emerg Med 2000; 12:337-343.

(82) Nunnink L. Blunt carotid artery injury. Emerg Med 2002; 14:412-421.

(83) Rhine DJ. Hyperbaric oxygen therapy in carbon monoxide poisoning: on neurological sequelae. CJEM 2000; 2:22-24.

(84) Rothman RE, Ketlogetswe KS, Dolan T, Wyer PC, Kelen GD. Should emergency departments conduct routine HIV screening? a systematic review. Acad Emerg Med 2003; 10:278-285.

(85) Schull MJ, Slaughter PM, Redelmeier DA. Urban emergency department overcrowding: defining the problem and eliminating misconceptions. CJEM 2002; 4:76-83.

(86) Tran T, Wax JR, Philput C, Steinfeld JD, Ingardia CJ. Intentional iron overdose in pregnancy-management and outcome. J Emerg Med 2000; 18:225-228.

(87) Warden CR, Zibulewsky J, Mace S, Gold C, Gausche-Hill M. Evaluation and management of febrile seizures in the out-of-hospital and emergency department settings. Ann Emerg Med 2003; 41:215-222.

(88) Weaver CS, Terrell KM. Update: do opthtalmic nonsteroidal anti-inflammatory drugs reduce the pain associated with simple corneal abrasion without delaying healing? Ann Emerg Med 2003; 41:134-140.

(89) Weigand JV, Gerson LW. Should emergency departments institute a falls prevention program for elder patients? a systematic review. Acad Emerg Med 2001; 8:823-826.

(90) Wilber ST, Gerson LW. A research agenda for geriatric emergency medicine. Acad Emerg Med 2003; 10:251-260.

(91) Yeung JK, Zed PJ. A review of etomidate for rapid sequence intubation in the emergency department. CJEM 2002; 4:194-198.

(92) Ahmad S, Beckett MW. Value of serum prolactin in the management of syncope. Emerg Med J 2004; 21:E3.

(93) Alter HJ, Koepsell TD, Hilty WM. Intravenous magnesium as an adjuvant in acute bronchospasm: a meta-analysis. Ann Emerg Med 2000; 36:191-197.

(94) Arnold JL, Halpern P, Tsai MC, Smithline H. Mass casualty terrorist bombings: A comparison of outcomes by bombing type. Ann Emerg Med 2004; 43:263-273.

(95) Balk EM, Ioannidis JPA, Salem D, Chew PW, Lau J. Accuracy of biomarkers to diagnose acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med 2001; 37:478-494.

(96) Batchelor J, McGuiness A. A meta-analysis of GCS 15 head injured patients with loss of consciousness or post-traumatic amnesia. Emerg Med J 2002; 19:515-519.

(97) Boudreaux ED, O'Hea EL. Patient satisfaction in the Emergency Department: a review of the literature and implications for practice. J Emerg Med 2004; 26:13-26.

(98) Brown MD, Rowe BH, Reeves MJ, Bermingham JM, Goldhaber SZ. The accuracy of the enzyme-linked immunosorbent assay D-Dimer Test in the diagnosis of primary embolism: a meta-analysis. Ann Emerg Med 2002; 40:133-144.

(99) Bush S. Is cocaine needed in topical anaesthesia? Emerg Med J 2002; 19:418-422.

(100) Cooke MW, Higgins J, Kidd P. Use of emergency observation and assessment wards: a systematic literature review. Emerg Med J 2003; 20:138-142.

(101) Cueto SM, Cavanaugh SH, Benenson RS. Computed tomography scan versus ventilation-perfusion lung scan in the detection of pulmonary embolism. J Emerg Med 2001; 21:155-164.

(102) Eddleston M, Juszczak E, Buckley N. Does gastric lavage really push poisons beyond the pylorus? A systematic review of the evidence. Ann Emerg Med 2003; 42:359-364.

(103) Edmonds ML, Camargo CA, Pollack CV, Rowe BH. The effectiveness of inhaled corticosteroids in the emergency department treatment of acute asthma: a meta-analysis. Ann Emerg Med 2002; 40:145-154.

(104) Farion KJ, Osmond MH, Hartling L, Russell KF, Klassen TP, Crumley E et al. Tissue adhesives for traumatic ulcerations: a systematic review of randomized controlled trials. Acad Emerg Med 2003; 10:110-118.

(105) Fredriksson M, Herlitz J, Nichol G. Variation in outcome in studies of out-of-hospital cardiac arrest: a review of studies conforming to the Utstein guidelines. Am J Emerg Med 2003; 21:276-281.

(106) Glavan B, Shewakramani S, Hollander JE. Incomplete data reporting in studies of emergency department patients with potential acute coronary syndromes using troponins. Acad Emerg Med 2003; 10:943-948.

(107) Goodacre SW. Should we establish chest pain observation units in the UK? A systematic review and critical appraisal of the literature. J Accid Emerg Med 2000; 17:1-6.

(108) Ioannidis JPA, Salem D, Chew PW, Lau J. Accuracy and clinical effect of out-of-hospital electrocadiography in the diagnosis of acute cardiac ischemia: a meta-analysis. Ann Emerg Med 2001; 37:461-470.

(109) Ioannidis JPA, Salem D, Chew PW, Lau J. Accuracy of imaging technologies in the diagnosis of acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med 2001; 37:471-477.

(110) Kline JA, Johns KL, Colucciello SA, Israel EG. New diagnostic tests for pulmonary embolism. Ann Emerg Med 2000; 35:168-180.

(111) Lau J, Ioannidis JPA, Balk EM, Milch C, Terrin N, Chew PW et al. Diagnosing acute cardiac ischemia in the emergency department: a systematic review of the accuracy and clinical effect of current technologies. Ann Emerg Med 2001; 37:453-460.

(112) Leary MC, Saver JL, Gobin YP, Jahan R, Duckwiler GR, Vinuela F et al. Beyond tissue plasminogen activator: mechanical intervention in acute stroke. Ann Emerg Med 2003; 41:838-846.

(113) Li J. Capnography alone is imperfect for endotracheal tube placement confirmation during emergency intubation. J Emerg Med 2001; 20:223-229.

(114) Marill KA, Runge T. Meta-analysis of the risk of torsades de pointes in patients treated with intravenous racemic sotalol. Acad Emerg Med 2001; 8:117-124.

(115) Mattick A. Use of tissue adhesives in the management of paediatric lacerations. Emerg Med J 2002; 19:382-385.

(116) McCusker J, Karp I, Cardin S, Durand P, Morin J. Determinants of emergency department visits by older adults: a systematic review. Acad Emerg Med 2003; 10:1362-1370.

(117) Murphy R, Driscoll P, O'Driscoll R. Emergency oxygen therapy for the COPD patient. Emerg Med J 2001; 18:333-339.

(118) Perry JJ, Lee J, Wells G. Are intubation conditions using rocuronium equivalent to those using succinylcholine? Acad Emerg Med 2002; 9:813-823.

(119) Quin G. Chest pain evaluation units. J Accid Emerg Med 2000; 17:237-240.

(120) Roberts K, Smith A. Outcome of diabetic patients treated in the prehospital arena after a hypoglycaemic episode, and an exploration of treat and release protocols: a review of the literature. Emerg Med J 2003; 20:274-276.

(121) Robinson N. The focused trauma ultrasound examination. Can, and should, accident and emergency physicians in the UK acquire this skill? J Accid Emerg Med 2000; 17:330-333.

(122) Robinson N, Clancy M. In patients with head injury undergoing rapid sequence intubation, does pretreatment with intravenous lignocaine/lidocaine lead to an improved neurological outcome? A review of the literature. Emerg Med J 2001; 18:453-457.

(123) Rodrigo G, Rodrigo C, Burschtin O. Efficacy of magnesium sulfate in acute adult asthma: a meta-analysis of randomized trials. Am J Emerg Med 2000; 18:216-221.

(124) Rowe BH, Bretzlaff JA, Bourdon C, Bota GW, Camargo CA. Intravenous magnesium sulfate treatment for acute asthma in the emergency department: a systematic review of the literature. Ann Emerg Med 2000; 36:181-190.

(125) Smith A, Roberts K. Interventions for post-traumatic stress disorder and psychological distress in emergency ambulance personnel: a review of the literature. Emerg Med J 2003; 20:75-78.

(126) Snooks HA, Dale J, Hartley-Sharpe C, Halter M. On-scene alternatives for emergency ambulance crews attending patients who do not need to travel to the accident and emergency department: a review of the literature. Emerg Med J 2004; 21:212-215.

(127) Trout A, Magnusson AR, Hedges JR. Patient satisfaction investigations and the emergency department: what does the literature say? Acad Emerg Med 2000; 7:695-709.

(128) Trzeciak S, Rivers EP. Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. Emerg Med J 2003; 20:402-405.

(129) Vickery D. The use of the spinal board after the pre-hospital phase of trauma management. Emerg Med J 2001; 18:51-54.

(130) Worster A, Preyra I, Weaver B, Haines T. The accuracy of noncontrast helical computed tomography versus intravenous pyelography in the diagnosis of suspected acute urolithiasis: a meta-analysis. Ann Emerg Med 2002; 40:280-286.

(131) Yildiz A, Sachs GS, Turgay A. Pharmacological management of agitation in emergency settings. Emerg Med J 2003; 20:339-346.

CHAPTER 4: Randomized Controlled Trials in Emergency Medicine: Where do they all go?

"We too commonly see references of "so many successful cases", with a certain inevitable emphasis on the word "successful". Such papers have their value and also their manifest dangers...There is unquestionably a false emphasis in all such publications, tending to increase the reputation of the writer, but not to render the public more secure. We have no proper balance to this very natural tendency to publish our successes except through the more frequent publication of our errors and failures which likewise mark the path of every successful practitioner. Such papers, written by men of experience and standing, would do much toward overcoming the tendency to over-security and would certainly serve an educational purpose which the ordinary publication so often fails to attain"
[Anonymus] The reporting of unsuccessful cases (editorial).
Boston Medical and Surgical Journal 1909; 263-264 [1].

4.1. Introduction

Presentation and subsequent publication of biomedical studies in the peer-reviewed literature are natural and expected outcomes that represent the completion of the research pathway. Publication of a scientific manuscript in a peer-reviewed journal reaffirms the work's scientific validity, adds new knowledge, and updates the evidence for better patient care. It helps enrich a permanent forum of informed discussion and encourages the sharing of experience among clinicians and researchers.

Underreporting of research -particularly of clinical trials- is a well-recognized problem, the size of which is difficult to estimate. There are two main aspects to this problem: publication deficit (also called "the iceberg phenomenon" [2] or "the file-drawer problem" [3]) and publication bias. The former refers to the fact that many studies are never published, with the deficit representing the difference between the proportion of studies eventually reaching full journal publication and the absolute (100%) rate of publication [4]. On the other hand, specific types of publication bias have been characterized and documented: *Positive results bias* (i.e. the publication of research results depends on the nature, direction or strength of the study findings [5,6], whereby manuscripts with statistically significant "positive" results are more likely to reach publication than studies reporting non-significant, "negative" results [7-17]), *time-lag bias* (i.e. the speed of publication depends on the direction and/or strength of the study results [18,19,]), *grey literature bias* (i.e. the results reported in journal articles are systematically different from those presented in working papers, dissertations or conference abstracts[6,20]), *full publication bias* (i.e. the full publication of studies initially presented at conferences or in other informal formats is dependent on the direction and/or strength of their findings [10,21]) and finally *place of publication bias* (i.e. studies with positive results may be more likely to be published in widely circulated journals than studies with negative results [10,22,23]).

The selective publication of randomized controlled trials (RCTs) based on the direction or strength of the results has been emphasized as a serious threat to validly assess the effectiveness of health care interventions [8]. This situation has important implications for clinicians who want to find good evidence of the effectiveness of interventions and for researchers synthesizing the evidence under the form of systematic reviews. Despite the efforts of the Cochrane Collaboration to identify and facilitate unlimited access to the results of published and unpublished RCTs, conclusions drawn from these reviews may be imprecise or biased if all trials cannot be included [24,25]. Therefore, publication bias poses considerable problems when pooled analyses of the effect of interventions (i.e. meta-analysis) are conducted, as it leads to an overestimation of the effect size [23], and a reduction of the power of systematic reviews to detect moderate but clinically important treatment effects [4,26]. Failure to publish results from RCTs, therefore, may have important negative consequences for health policy, clinical decision making and the clinical outcomes of patient care [6,27]. The problem has also been described as a form of scientific misconduct that can lead those caring for patients to make inappropriate treatment decisions, either providing ineffective or dangerous forms of care, or delaying the provision of other beneficial health care interventions [4, 28,29]. No publication of RCTs has other ethical implications such as breaking up the contract between investigators, study participants and funding agencies [29].

Presentation of results under abstract format at scientific meetings constitutes an intermediate stage in the process of disseminating research findings. One of the most important functions of scientific meetings is to provide a forum for communicating the results of novel research. Although subsequent publication of a full-text paper in a peer-reviewed journal is the natural and expected outcome of such presentations, it is not uncommon for abstract to be the only format of publication for biomedical research studies [9,11]. It is thus important to determine the extent to which conference abstracts from scientific meetings result in publications in peer-reviewed journals.

Considerable evidence from a variety of clinical specialties have shown that a substantial proportion of abstracts presented at scientific meetings remain unpublished in the scientific literature as full-length manuscripts. Systematic reviews examining the follow-up of meeting abstracts for various clinical specialties have found that more than half were not fully published after presentation at the meetings [9,30] and that some characteristics of the abstracts such as "positive results" and sample size may be associated with subsequent manuscript publication.

The Society for Academic Emergency Medicine (SAEM) scientific annual meeting is an important forum for communication of novel biomedical research in the field of emergency medicine. The SAEM meeting is comparable with the meetings of other scientific specialties in terms of attendance and number of abstracts submitted [11,12]. Over the past thirty years, the number of abstracts presented at the SAEM annual meetings has exponentially increased from 18 in 1974 to 563 in 2004, reflecting the development, evolution, and continued growth of research in emergency medicine. Particularly, there is evidence of an increasing trend for more randomized controlled clinical trials (RCTs) and controlled clinical trials (CCTs) (from 0% in 1974, 12% in 1983, 12% in 1989, 15% of all submissions in 1997) and more blinded studies (from 0% in 1974, 7% in 1983, 5% in 1989, to 11% in 1997) presented as abstracts in the SAEM meetings [31]. The publication rates of abstracts submitted at SAEM meetings have been studied in the past for individual SAEM meetings. For example, Callaham et al. [11] found that 62% of *all* abstracts accepted for presentation at the 1991 SAEM meeting were subsequently published. Little is known about the factors leading to subsequent full publication of RCT abstracts submitted to annual meetings like SAEM and whether there is any indication of publication bias in relation to RCTs in emergency medicine. The present study is the first to examine what proportion of abstracts of RCTs presented at a sample of SAEM annual meetings eventually will be published as full-length articles, and whether publication bias has any effect in emergency medicine research.

4.2. Objectives

The aims of this study were:

1) To determine the rate of subsequent full publication of CCT/RCTs presented at 1995-2003 SAEM meetings in peer-reviewed biomedical journals.
2) To identify abstract characteristics associated with subsequent publication and whether different types of publication bias affected CCT/RCTs in emergency medicine research (i.e. positive results bias, time-lag bias, grey literature bias, and place of publication bias).

4.3. Methodology of the Study

4.3.1. Study design:

Retrospective cohort study based on CCT/RCT abstracts accepted for presentation at the SAEM annual scientific meetings from 1995 to 2003.

4.3.2. Eligibility criteria:

All abstracts described either as CCT or RCTs and accepted for oral or poster presentation at the 1995-2003 annual scientific meetings of SAEM were eligible. A CCT was defined as any form of prospective comparative study where participants were concurrently enrolled and assigned to one of two or more treatments or interventions. A RCT was defined as a prospective comparative study where participants were allocated by a random process to receive one of two or more clinical interventions [32,33]. Abstracts and manuscripts were required to report the results from CCTs/RCTs involving human participants. Abstracts reporting meta-analyses, cohort studies, case-control studies, case reports, case series, incidence/prevalence studies, qualitative studies, laboratory studies, and descriptions of procedures and instrument validations were excluded.

4.3.3. Selection of abstracts and manuscripts

One reviewer (MO) hand-searched all 4399 abstracts of oral and poster presentations accepted for the SAEM annual meetings from 1995 to 2003 to identify all the CCT and RCT abstracts. A second reviewer (BR) independently reviewed a 10% random sample [34] of the same sample of abstracts. The abstracts were published in the special May issues of the *Academic Emergency Medicine* journal, the official journal of the SAEM. If the inter-evaluator agreement on the 10% sample of abstracts was high, it was assumed that the reviewers used the same criteria to select the CCT/RCT abstracts and that the eligibility criteria were similarly applied. Therefore, it was decided that one investigator could reliably continue selecting the CCT/RCT abstracts from the remaining 1995 - 2003 sample of abstracts. Disagreements on eligibility were resolved by discussion until consensus was reached.

A computer-based search using MEDLINE (via PubMed), EMBASE and CINAHL (via Ovid) databases from 1995 to April 2004 was completed to identify publications in peer-reviewed journals resulting from any trial reported in the SAEM abstracts (updated in May 2004 in the Cochrane

Register of Controlled Trials). The databases were searched by one of the investigators (MO) using both authors and key text words within the title and the abstract. Only full manuscripts published in peer reviewed journals were considered to represent abstract publications. No efforts were made to check for multiple articles stemming from a single abstract or vice-versa. Once a match was found, the search process was concluded.

The abstract and possible resultant manuscript were examined to ensure that the following criteria were satisfied: 1) The first author of the SAEM abstract should be one of the authors of the full publication, 2) at least one outcome of the SAEM abstract was an outcome of the manuscript, and 3) the topic of evaluation and the interventions were identical. An outcome was defined as a finding stated in the Results section of the SAEM abstract that was also stated in the Results section of the manuscript. If two SAEM abstracts were combined into one final manuscript, both abstracts were considered published. Abstracts were considered as unpublished if they could not be retrieved as full reports, including papers that were still in the publishing process (in press), not easily accessible (grey literature), or not published at all. No attempts were made to contact the authors to ascertain the outcome of the SAEM abstracts.

4.3.4. Assessment of methodological quality:

The methodological quality of CCT/RCT abstracts was independently assessed by two independent reviewers (MO, PG) blinded to their publication status. Full-published manuscripts were also independently evaluated by the same reviewers. Trial quality was assessed using the criteria for concealment of allocation [35,36] described in the Cochrane Handbook and the Jadad Scale [37] (see Appendix B2 and B3). The former is based on the evidence of a strong relationship between the potential for bias in the results and the allocation concealment [38]. The Jadad scale [37] is a validated 5-point scale that includes three items rated on a yes/no format that are directly related to bias reduction: randomization, double-blinding and description of drop-outs and withdrawals. Points for the first two items (randomization and double-blinding) may be added or deducted depending on the appropriateness of the methods used to randomize and blind the trial. This scale was selected to assess the methodological quality of CCTs and RCTs because its psychometric properties have been thoroughly tested, providing rigorous evidence to support its use [37,39]. The level of agreement between the two reviewers on both the individual scores of the Jadad scale and the allocation concealment components was measured by calculating Kappa (κ) coefficients and 95% CI [40]. A κ score in the range of 0.0 to 0.40 was considered poor agreement, from 0.41 to 0.60 moderate agreement, and a κ in the range of 0.61 to 0.80 was considered to represent substantial agreement [41]. Inter-rater agreement for the Jadad scale overall scores was estimated with the Intra Class Correlation Coefficient (ICC) [42] and 95% confidence intervals (95%CI). For this study, moderate agreement was defined as ICC > 0.6 and good agreement as ICC > 0.8 [43].

4.3.5. Data extraction:

One reviewer (MO) extracted data from both abstracts and manuscripts using a standard format. Abstracts were assigned to 17 broad research categories based on standard SAEM abstract classification. Placement of each abstract into discrete categories was often difficult as abstracts frequently had several components and could fall into a number of categories. The following information was collected for abstracts and manuscripts: title, year of meeting (abstract only), publication status (abstract only), journal name, date of publication (manuscript only), number of authors, country of the first author, report of university affiliation of any of the authors, declaration of funding, number of centers participating (multicenter, single study), design (parallel, cross-over), sample size, and whether the primary outcome was clear or unclear.

Study results were classified as positive in three different ways: First, a positive outcome was defined when statistical significance (defined by a p-value<0.05 or by a 95%CI excluding the unity) was achieved for the primary outcome. Second, a positive outcome was defined when the direction of the result favoured the experimental intervention. Third, a positive outcome was defined when the authors' overall conclusions endorsed the experimental intervention. No distinction between efficacy and equivalence trials was made as it was expected that, considering the amount of information provided in the abstracts, the accuracy to distinguish among the two categories would be low. Every CCT/RCT resulting in a publication was assigned a journal citation impact factor (IF) [44] derived from the Science Citation Index (SCI) for the year of publication of the manuscript. The IF is a measure of the frequency with which articles are cited in other journals. This was used as an indicator of a journal's relative influence in terms of the level of impact in the medical field.

4.3.6. Data analysis

Dichotomous and categorical data were reported as proportions and percentages with 95% CIs. Continuous data were reported using means with standard deviations (SD) or medians with interquartile ranges (IQR) where appropriate. Publication rate was defined as the proportion of subsequent published CCT/RCT abstracts to the total number of CCT/RCT SAEM abstracts. To determine the likelihood of positive results bias (dependent variable: published vs. non-published) associations between abstract characteristics and the dependent variable were calculated by univariate and multivariate logistic regression analyses (stepwise forward logistic procedure based on the likelihood ratio). Unadjusted and adjusted odds ratios (OR) and their corresponding 95%CI were calculated. The regression models tested the following independent variables: country of origin, number of authors, report of university affiliation, study design (parallel, crossover), number of study centers (single, multicenter), sample size, primary outcome (clear, unclear), type of primary outcome (dichotomous, continuous), direction of the results (favouring the treatment or not), and authors' conclusions (endorsing a treatment or not). The variables were selected based on previous evidence indicating that these variables may have an important effect on the status of subsequent publication outcome.

To assess the probability of grey literature bias in this sample of SAEM abstracts, changes from abstract to manuscript in terms abstract data were explored. The likelihood of changes in variables from abstract to manuscript was calculated using odds ratios (OR) and their associated 95%CI. McNemar tests were used for paired comparisons.

Kaplan-Meier time to event analysis was used to investigate the time to publication to enable inclusion of abstracts where time since presentation at the SAEM meeting was insufficient for publication to have occurred. The time to publication was defined, in months, as the median time between the month of abstract presentation and the month of journal publication. Full publication was defined as the event; abstracts that remained unpublished were censored at the last follow-up time period (May 2004). Differences in subsequent full publication according to abstract characteristics were investigated using Peto's log-rank tests. To determine predictors of the time to full-manuscript publication of CCT/RCT abstracts, Cox proportional hazard models were constructed after ensuring that assumptions of the Cox proportional hazard models were not violated.

To assess the effect of place of publication bias (dependent variable: IF above and below the median IF for the sample of published abstracts), univariate and multivariate logistic regression analyses (stepwise forward logistic procedure based on the likelihood ratio) were undertaken for the sample of CCT/RCT manuscripts from SAEM abstracts. Unadjusted and adjusted OR and their corresponding 95%CIs were calculated. The regression models tested the same variables that were included in the other regression analyses. Statistical analyses were performed by using the SPSS (Statistical Package for Social Sciences) software version 12.0 for Windows©. A 2-sided type I error rate of 0.05 was used for all statistical tests.

4.3.7. Ethics

This study is exempt from ethics review by the Health Research Ethics at the University of Alberta, Edmonton, Canada as it did not involve the participation of patients or their families

4.4. Results

4.4.1. General characteristics of abstracts and manuscripts

Selection of Abstracts: The level of agreement among reviewers for the selection of CCT/RCT abstracts was excellent (κ: 0.92; 95% confidence interval (CI): 0.86, 0.97; n = 440). Figure 4.1 shows the study selection flow. Of the 4399 abstracts presented at the 1995-2003 SAEM annual scientific meetings, 383 were identified as CCT/RCTs (8.7%).

Publication Outcomes: One hundred and ninety four CCT/RCTs (50.7%) were subsequently published in peer-reviewed journals and 189 remained unpublished up to April 2004 (Table 4.1). The majority of CCT/RCT abstracts addressed miscellaneous emergency medicine research topics (19.1%), pediatrics (15.1%), injury/trauma (13.3%), and airway/respiratory (12.0%) topics. The remaining 13 categories were used for less than 10% of the abstracts (Figure 4.2).

Characteristics of CCT/RCTs: Baseline characteristics of the abstracts are shown in Table 4.2. Characteristics of the manuscripts are reported in Table 4.3. Of note, none of the CCT/RCT abstracts declared the funding source for their studies. Manuscripts were published in 33 journals: 64.9% in any of six emergency medicine journals (*Academic Emergency Medicine* [55], *Annals of Emergency Medicine* [46], *American Journal of Emergency Medicine* [16], *Pediatric Emergency Care* [5], *Emergency Medicine Journal,* and *Prehospital Emergency Care* [one each]) and the remaining 35.1% in a variety of medical journals.

4.4.2. Methodological quality of CCT/RCT abstracts and manuscripts

The inter-rater agreement for the Jadad overall scores, the individual components of the Jadad scale, and the allocation concealment for both the CCT/RCT abstracts and the subsequent publications was good (Table 4.4) with κ coefficients ranging from 0.60 (double blinding bonus item for abstracts) to 1.00 (randomization deduction and double blinding deduction items for manuscripts). No differences in methodological quality measures were found between unpublished abstracts and those that were subsequently published. The changes in methodological quality from abstract to manuscript form were statistically significant for the median Jadad scores, for some individual components of the Jadad scale such as randomization bonus, double blinding bonus, and withdrawals/dropouts) and for the allocation concealment components (Table 4.5). The methodological quality of the abstracts was not significantly associated with their publication status (OR Jadad scores: 1.13; 95%CI: 0.87, 1.46; χ^2: 0.91, df = 1, p = 0.34)

4.4.3. Predictors of publication

Table 4.6 provides the results of the univariate and multivariate logistic regression analyses assessing the effect of predictors for publication. From the univariate analysis, there were no statistically significant differences between the published and the unpublished CCT/RCT abstracts in any of the potential predictors of publication, except that authors of abstracts reporting university affiliations were more likely to reach full-publication compared to those with no report of university affiliations (OR: 1.58, 95%CI: 1.03, 2.42; χ^2: 4.18, df = 1, p = 0.04). Abstracts were not more likely to be published when they reported significant results (OR = 1.11; 95%CI: 0.74, 1.66; χ^2: 0.15, df = 1, p = 0.69) or endorsed any of the treatments in the conclusions (OR = 1.07; 95%CI: 0.71, 1.60; χ^2: 0.5, df = 1, p = 0.80).

None of the variables included in the multivariate logistic regression model were statistically significant at 5% level.

4.4.4. Time to publication

One abstract was excluded from the survival analyses as it had been published before the 1995 meeting. The overall median time of publication was 32 months (95%CI: 23, 42). After 5 years, 41% of the CCT/RCT SAEM abstracts remained unpublished (Figure 4.3). To evaluate whether the speed of subsequent publication depended on the direction and/or strength of the study results (time-lag bias), Peto's log rank test was used to compare the time of publication by overall study findings. There were no statistically significant differences in the median time to publication of CCT/RCT abstracts reporting significant results (30 months; 95%CI: 17.2, 42.8) compared to those that did not (32 months; 95%CI: 17.9, 46.2; Peto's log rank test = 0.01, p = 0.91). Abstracts reporting significant results favouring the experimental intervention were not published faster (32 months; 95%CI: 18.7, 45.3) than those favouring control interventions or reporting null results (29 months, 95%CI: 6.0, 52.0; Peto's log rank test = 0.00, p = 0.94). Statistical differences were not found when comparisons were made between the median publication time of abstracts endorsing a treatment (30.06 months; 95%CI: 18.2, 41.8) versus the median publication time of those that did not (33.08 months, 95%CI: 21.47, 44.7; Peto's log rank test = 0.02, p = 0.89).

Results from the Cox proportional hazard model analysis confirmed the findings of the univariate logistic regression model: after controlling for factors such as number of authors, and country of corresponding author, CCT/RCT abstracts reporting university affiliation of any of the authors were 1.42 times more likely to reach publication compared to those abstracts that did not report such academic affiliation. Report of university affiliation was a significant predictor of the instantaneous relative risk of subsequent publication after controlling for other factors (Table 4.7).

4.4.5. Changes from abstract to publication

To evaluate the probability of grey literature bias, 194 abstracts were paired with their corresponding manuscripts. No significant changes from abstract to manuscript were found in the number of authors (Median [IQR]: abstracts: 4 (3, 6), manuscripts: 5 (3, 6); Wilcoxon Signed Ranks Test: Z = -1.29, p = 0.19), (OR: 0.92, 95%CI: 0.31, 1.46; χ^2: 0.59, df = 1, p = 0.44), the number of participating centers (OR: 3.66; 95%CI: 1.0, 12.9; 0.09, 0.69; Mc Nemar χ^2:: 3.5,df = 1 p = 0.99), the type of primary outcome measure (OR: 1.5, 95%CI: 0.53, 4.17; χ^2: 0.26, df = 1 p = 0.6) or the reported significance of the primary outcome (OR: 1.54, 95%CI: 0.72, 3.26; χ^2: 0.89, df = 1, p = 0.345). Statistically significant changes, however, were observed in the reported affiliation with a university (OR: 9.3, 95%CI: 2.83, 30.2; χ^2:: 18.581, df = 1, p = 0.0001), in the definition of the primary outcome (OR: 2.05; 95%CI: 1.20, 3.49; Mc Nemar χ^2: 6.55, df = 1, p = 0.01), and the authors' conclusions (OR: 0.26; 95%CI: 0.098, 0.69; Mc Nemar χ^2: 7.842, df = 1, p = 0.007). There is a 9-fold increase in report of university affiliation in published manuscripts versus abstracts. Manuscripts were also 2 times more likely to define the primary outcome under study in a clear way. Author's conclusions also changed from abstract to manuscript form. Compared to abstracts, manuscripts were 0.26 times less likely to endorse the treatment under study. Abstracts were more often to support the intervention under study in the conclusions than their corresponding manuscripts (see Table 4.8).

4.4.6. Place of publication bias:

The median IF of the journals that published the CCT/RCT manuscripts was 1.535 (IQR: 1.05, 2.46). Results from the logistic multivariate regression (Table 4.9) indicated that, compared to manuscripts reporting dichotomous primary outcomes, continuous primary outcomes have 0.4 chance to get published in journals with higher levels of circulation after controlling for other factors (OR: 0.42; 95% CI: 0.20, 0.86; Wald test: 5.58, df = 1, p = 0.01). Similarly, manuscripts that received Jadad scores > 3, were more than twice as likely to get published in journals with higher IF compared to those with Jadad scores lower than 3 (OR: 2.26; 95%CI: 1.17, 4.33; Wald test: 6.0, df = 1, p = 0.14). No evidence of place of publication bias was found in this sample of CCT/RCT manuscripts. Studies with positive results (e.g. statistically significant, or either favouring or endorsing the experimental intervention) were not more likely to be published in higher rated journals than studies with negative results.

4.5. Discussion

4.5.1. General remarks

This study examined the percent publication, time to publication and determinants of publication of a cohort of 383 CCT/RCT abstracts accepted for presentation at the 1995 – 2003 annual SAEM meetings. In particular, the occurrences of several types of publication bias (positive outcome bias, time-lag bias, grey literature bias, and place of publication bias) were explored. A publication rate of 50.4% (95% CI: 45.3, 55.5) for CCT/RCT abstracts was identified through electronic database searches. This finding is in accordance with figures reported for all research abstracts (not solely CCT/RCTs) in other scientific meetings [9,45]. A meta-analysis on the publication rate of CCT/RCT abstracts [30] calculated a weighted mean rate of full publication of 55.8% (95%CI: 51.8, 59.9) which is higher than the proportion reported in this study. With respect to other studies in the emergency medicine field, the publication percentage is comparable to the findings of Callaham et al. [11] for *all* the abstracts presented at the 1991 SAEM meeting (44%), Walby et al. [46] for all research abstracts presented at the scientific meetings of the Australasian College for Emergency Medicine from 1995 to 1998 (35%), and Korn et al. [47] for prospective studies presented at four emergency medicine research forums (43%).

Publication rates of abstracts reporting either statistically significant results or endorsing any of the treatments under study were not different from those with non-significant results (OR: 1.08; 95%CI: 0.72, 1.62) or those that did not endorse any treatment (OR: 1.05; 95%CI: 0.70, 1.57), respectively. This finding is surprising, since the association of positive results and publication has been noted previously by many authors [5,8,10,11,15-17,21,30,48]. In the meta-analysis of Scherer et al. [30] a strong association between "positive" results and full publication (RR = 1.51; CI 1.27 to 1.79) was identified when authors defined "positive" results as a "significant" result favouring any treatment arm over another. RCTs in emergency medicine tend to be published no matter what the outcome of the trial, and this situation may depend on the rate of RCTs produced in emergency medicine. Compared to other biomedical areas, emergency medicine produces a lower number of RCTs. Therefore, once a RCT is completed, it is unlikely that these results (either positive or negative) will not be submitted or accepted for publications in the emergency medicine journals.

This study also failed to demonstrate a positive outcome bias in this sample of CCT/RCT SAEM abstracts and supports the findings of other studies that have not found a statistically significant association between study outcome and full publication[10,12,23,28,49,50]. There are a number of possible explanations for the differences between these results and research reporting a positive association between study outcome and status of publication. First, it is possible that studies differed in the definitions of what constituted a "positive" finding; therefore, a positive association between results and publication may depend on the definition employed. Some of the definitions in other studies placed emphasis on the presence or absence of statistical significance (p-values), the direction of the study results (positive versus negative or neutral), or on subjective classifications based on the author's conclusions. Furthermore, in other previously reported studies, a dichotomous classification was used with negative outcome trials comprising any study without statistically significant results [9,30,51]. In this study, using three different definitions of a positive outcome (statistical significance, direction of results, and study conclusions) we failed to identify an association between outcome and publication status.

Second, the methodological quality of research may be an issue to explain our results. The methodological quality of a research project is a factor that has been hypothesized to influence the likelihood of subsequent full publication [28]. The congruence between the quality of a project reported in abstract form compared to the quality assessed from a subsequent full publication is questionable. Considerable inconsistencies between abstracts and subsequent published reports have been pointed out in previous evaluations[16,52]. No significant association between the methodological quality of the abstracts and their subsequent publication status was found in this study (OR Jadad scores: 1.133; 95%CI: 0.87, 1.46). Moreover, this finding is supported by the results of a systematic review on the fate of biomedical meeting abstracts, which could not draw any conclusions on the relationship between study quality and the likelihood of abstract acceptance and subsequent full publication [45]. Similarly, in a follow-up study on summary reports of controlled trials, Chalmers et al. [28] failed to identify abstract quality as a predictor of full publication. Alternatively, the absence of a relationship between methodological quality and the likelihood of full publication may be also be explained as an effect of the method selected to assess the methodological quality of the abstracts or simply the fact that the abstracts may not have enough space to present all the information needed for an adequate evaluation of the study quality. Finally, the utility of the Jadad scale has not been evaluated for application at the meeting abstract level, so the validity of this method may partially explain the difficulty identifying the association.

On the other hand, differences between the methodological quality from abstract to manuscript were detected in this study and have been described in similar studies [23,52], particularly regarding some individual quality components, such as the adequacy of randomization and double blinding procedures, the accounting for study withdrawals and dropouts, and the allocation concealment. This finding may suggest a type of publication bias that has not been described before and perhaps could be described as "space bias": differences in the reporting of methodological characteristics depend on the space available to report the study results. This bias should be evaluated in other settings.

Predictors of publication were evaluated using both logistic regression models and proportional hazard model techniques. Some variables described as predictors of publication in other studies, such as sample size [9,10,12,28,35,53,54] and multicenter status [23,] were not statistically significantly associated with successful manuscript publication in this study. The lone factor found to be positively associated with publication in this sample of CCT/RCT abstracts was the documentation of a university affiliation. This supports the results of other studies [10,23] that have shown a similar association. Such a finding may reflect the urgency of publication in academic locations where tenure and promotion are tightly linked to productivity compared to those authors who are not academically linked.

The median time to publication of the CCT/RCT SAEM abstracts was 32 months (95%CI: 22.97, 41.10). This result is similar to the findings reported by Evers [4], Timmer et al. [55], Krzyzanowska et al.[56], and Ioannidis [18] in other medical specialties; however, it was longer than times reported by other groups[30,50,57]. Explanations for these divergences include differences in the measure used to report the outcome (medians, means) and the methods to calculate the time to publication. Time of publication was calculated here using a Kaplan Meier time-to-event approach. If a time-to-event analysis was not used, the crude median time to publication in this study would be 24 months, a figure that is in accordance to the 19.7 months [30] and 18 months [50,57] reported elsewhere. One meta-analysis [9] found that 32% to 66% of abstracts reported at scientific meetings of 6 different subspecialties were published in full manuscript format 3 years later and only a minority of the studies were published thereafter. In the present study, 59% of the CCT/RCT abstracts were published in full within 5 years of presentation. Previous reports have suggested that the majority of abstracts are published in indexed journals within four years of presentation [58,60], therefore, the fate of 41% of the studies that were not published afterwards is unknown.

The discussion around the so called "time lag bias" showed conflicting results with the evidence provided to date. Abstracts reporting results favouring the experimental intervention (i.e. statistically significant, positive findings, positive conclusions) were not published faster than those favouring control interventions or reporting neutral results. These findings support Dickersin et al [61] results on the association between time interval to publication and statistical significance from a sample of 133 controlled trials. Time to publication was not associated with statistical significance or with any other quality or study characteristics in this study. Conversely, Ionnadis [18] reported that randomized efficacy trials were published more rapidly when results reached traditional levels of statistical significance. Again, the differences in the findings among the group of studies exploring the association between positive results and time to publication may be confounded by the definitions used to classify the studies as "positive" or "negative".

Significant changes in magnitude or direction of the association between the set of independent variables and the status of publication and time to publication were not observed when a sub-sample of abstracts published from 1995-2000 was evaluated using the same parameters. Allowing for additional time to publish did not alter the results from this review.

The findings from this study suggest that a grey literature bias is present in this sample of CCT/RCT SAEM abstracts. That is to say, author's conclusions in the scientific abstract differ from those in the corresponding publications (OR: 0.26; 95%CI: 0.098, 0.69). Differences were also found in other study characteristics, such as the reported affiliation with a university (OR: 9.3, 95%CI: 2.83, 30.2) and the clear or unclear definition of the primary outcome (OR: 2.05; 95%CI: 1.20, 3.49). Studies comparing the information from meeting abstracts to those in the subsequent full paper shed some doubts on the reliability of results reported in abstracts [10,20]. Weintraub et al. [62] examined a random sample of papers published in the *Journal of Pediatric Surgery* together with their original scientific abstracts. Changes in the conclusions from abstract to manuscript were found in 30% of the comparisons. They also stated that the manuscript included numbers that were often mathematically inconsistent with the abstract. On the other hand, Bellefeuille et al. [16] found "good to excellent correlation" between the conclusions of manuscripts and meeting abstracts in 15 out of 18 phase III trial reports (a smaller sample size than that collected in the present study).

It could be argued that differences found here may be due to the fact that meeting abstracts are "works in progress" rather than a correct summary of the underlying completed research. If true, this would explain why changes in sample size and shifts in the reported results have been found in other studies. Therefore, changes in the direction of the result would be a natural consequence of the increased power to detect statistically significant differences. Nevertheless, no significant changes in the sample size between the abstract and its corresponding publication were found in this study, suggesting that the change in the direction of the conclusions was not the result of increased power to detect the difference. By virtue of the relative paucity of details about study characteristics in abstracts, some authors have suggested that research available in abstract form is unreliable and incomplete [63] and that the validity of the information presented in abstract form is limited to allow for a formal quality analysis [64]. Furthermore, the use of meeting abstracts as citations in research publications is discouraged by many researchers whom consider them misleading or inappropriate [58,65].

The grey literature bias reported here may be an important finding that should be considered by those including the results from scientific abstracts in meta-analyses of the efficacy and effectiveness of clinical interventions. It has been reported in the past that exclusion of grey literature from systematic reviews may exaggerate efficacy estimates by 15%-38% [66]. How their inclusion in a meta-analysis might bias the efficacy estimates should also be considered, particularly when abstract constraints of space usually lead to a less detailed description of methods and results, which can hamper a critical appraisal of their validity. No definite answer to this debate has been provided to present. Moher et al. [67] suggest that, although it is unknown whether there would be more or less bias if the unpublished studies were in the form of full reports rather than abstracts, their methods of reporting certainly need to improve[67].

Finally, no evidence of place of publication bias [22] was found; however, manuscripts with higher Jadad scores were 2 times more likely to be published in journals with higher IF. This does not necessarily mean that higher quality CCT/RCTs are published in higher quality journals. The interpretation of the impact factor must be undertaken with an understanding of its true meaning: the number of times, on average, that a source article in a particular journal is likely to be cited [68]. This distinction between the rate of citation and journal quality is crucial when interpreting these results.

4.5.2. Strengths and weaknesses of this study

The approach used here to evaluate the publication rate and the effects of publication bias on a sample of CCT/RCT abstracts have some strengths and weaknesses. It could be argued that the abstracts evaluated may not be representative of emergency medicine research in general as they were presented at one individual emergency meeting; however, SAEM is the principal forum for presenting novel research, particularly RCTs in the North American academic EM research context. The research evaluated in the present study, however, only reflects abstracts that were *accepted* at the scientific meetings and does not necessarily reflect the state of all research in the field of emergency medicine. For example, only accepted RCT abstracts were evaluated instead of including all submissions. It should thus be kept in mind that this study can only provide an impression of part of the publication bias in emergency medicine, as the fate of those studies that were submitted and rejected is unknown. Nevertheless, as few RCTs are conducted in the EM field compared to other biomedical areas, it is unlikely that EM researchers would not attempt to disseminate the findings of studies that demand a lot of resources and effort. Future studies should compare whether a difference exists among rejected and accepted manuscripts for presentation at scientific meetings in terms of their subsequent rate of publication and the nature of their results.

Secondly, the description of the publication experience may have been affected by the search methods that were used in this study. The restriction of the searches to electronic databases can contribute to an outcome misclassification because certain studies that were classified as "not published" may have indeed been published elsewhere and simply not indexed in the electronic databases. Moreover, not all relevant journals are reviewed by the Index Medicus, thus the proportion of the true rate of publication of CCT/RCT abstracts may be underestimated. On the contrary, considering each abstract as an independent submission may lead to an overestimate of the proportion of the research that is eventually published since a single report might account for more than one abstract. Future studies in this area should use complementary approaches, such as survey methodology, to complement the information on the status of publication of CCT/RCT abstracts and to further understand why a considerable proportion of trials remain unpublished.

The quality of the underlying research was one of the most difficult variables to assess in the prediction of publication analysis, especially as the information available from the abstracts was limited. Furthermore, the use of the Jadad scale may be questioned given its inherent bias against therapies that could not be adequately double-blinded, and the paucity of information on the validity of its use for abstract evaluation. Conversely and despite their weaknesses, the Jadad criteria are the most widely quoted and evidence-based criteria available at the present time.

The use of time-to-event analysis to evaluate the questions under study strengthened the results by providing more sensitivity to the logistic regression analysis. When the outcome of a study is the time between one event (i.e. SAEM meeting) and another (publication), a number of problems can occur: the times are unlikely to be normally distributed and the time interval may vary during the study period. In addition, the study has a finite duration (up to April 2004), during which not all abstracts would have experienced the outcome of interest (publication). Therefore, all these factors can be controlled for only when a time-to-event analysis approach is used.

The main concern in this study arises from the likelihood of misclassification of the abstract results as positive or negative. Although three indicators (reported significance, direction of the results, and author's endorsement) selected for this study constituted an appropriate alternative to control for confounders in the classification of the abstract results, they can only serve as an approximation to the eventual study outcome. If non differential classification is present, this may have weakened the effect, which would explain the lack of statistical evidence for the association under study. Nevertheless, if sample size calculations (80% power) were made to detect a 25% difference in full publication rate between studies with statistically significant results compared to those without, 138 abstracts of RCTs or 2760 abstracts would be needed. As the number of both abstracts and CCT/RCT abstracts were larger in this study, it is unlikely that our finding is a false negative result.

In summary, the approach selected to study publication bias here may have been subject to methodological limitations that possibly introduced some bias. It is difficult, however, to assess the degree to which the methodological problems may have influenced the study results.

4.6. Conclusions and recommendations

The proportion of emergency medicine RCT/CCT abstracts published is slightly lower than for other speciality societies; however, biases reported by others do not appear to be as common or problematic in this group of projects. Given the differences between the abstracts and manuscripts, caution is warranted with respect to employing meeting abstracts as a source of evidence for future research or systematic reviews. Information cannot be considered entirely reliable, as inconsistencies with information at the manuscript level were frequent in the direction of the primary outcome results.

The solution to these problems rests in a number of areas. First, substantial improvements are required with respect to the quality of reporting abstracts. The decision of restricting the abstract to 250 words should be revisited to allow for inclusion of more comprehensive methodological information in the body of the abstract to enhance its value and validity. Particularly, guidelines to improve structured abstracts should be developed to include a more detailed description of the methodology of the studies that considers the adequacy of the methods of randomization and blinding, the concealment of allocation to treatments and the description of the withdrawals and dropouts. A format based on the CONSORT statement would be preferred, and certainly could be accomplished without a space expansion[69].

Different methodological approaches have been used to assess the impact of publication bias, each of them having advantages and disadvantages. *Post-hoc* assessment of the effect publication bias has on meta-analysis, such as funnel plotting, modelling or sensitivity analysis, have been proposed [66,70-73]. While these methods may be useful at interpreting the findings of a meta-analysis, no information on the nature and determinants of publication bias is provided. Retrospective cohort-based studies, similar to the one reported here, identify RCTs at an early stage before full publication (e.g. abstracts submitted to scientific meetings) and examine characteristics such as publication rates, time to publication, and associations between abstract characteristics (direction and strength of the effect studied, methodological quality, sample size) and subsequent publication in peer-reviewed journals. Compared to *post-hoc* evaluations of the effect of publication bias in meta-analysis and systematic reviews, bias in this preliminary stage of the publication process has been poorly studied. Similar studies addressing this problem are thus warranted.

It can be concluded that publication bias is not likely to be of high magnitude in this set of emergency medicine abstracts; however, it must be considered when evaluating the literature in emergency medicine. The proposal for a trial registry in emergency medicine should be explored to improve the accessibility of trial information and data. Joint solutions among clinicians, researchers and journal editors will surely facilitate the implementation and dissemination of this endeavour. A call for a collaborative work in this area should therefore be a priority for the research agenda in academic emergency medicine.

Table 4.1: Publication rates for RCT/CCT abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

Year	**Number of abstracts**	**CCT/RCTs abstracts N (%)**	**Published N (%)**
1995	394	33 (8.37)	23 (69.7)
1996	466	40 (8.58)	24 (60.0)
1997	544	50 (9.19)	28 (56.0)
1998	533	42 (7.87)	27 (64.3)
1999	537	43 (8.00)	28 (65.1)
2000	502	44 (8.76)	27 (61.4)
2001	475	36 (7.57)	17 (47.2)
2002	497	52 (10.46)	17 (32.7)
2003	451	43 (9.5)	3 (7.0)
TOTAL	4399	383 (8.7)	194 (50.4)

Table 4.2: Characteristics of CCT/RCT abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

Baseline characteristics	Overall (N = 383)	Published (N = 194)	Unpublished (N = 189)
Country (N [%])			
USA	339 (88.5)	169 (87.0)	170 (90.0)
Others	44 (11.5)	25 (12.9)	19 (10.0)
Number of authors			
Median [IQR]	4.0 (3, 6)	4.0 (3, 6)	4.0 (3, 5)
Reported university affiliation (N [%])	253 (66.1)	138 (71.1) ‡	115 (60.8)
Declaration of funding (N [%])	NS	NS	NS
Study design (N [%])			
Parallel	327 (85.4)	164 (84.5)	163 (86.2)
Crossover/factorial	56 (14.6)	30 (15.5)	26 (13.8)
Number of study centers (N [%])			
Single center	322 (84.1)	165 (85.1)	157 (83.1)
Multicenter	61 (15.9)	29 (14.9)	32 (16.9)
Sample size (Median [IQR])	68 (34, 150)	68 (34, 150)	68 (31, 152)
Primary outcome (N [%])			
Clear	239 (62.4)	127 (65.5)	112 (59.3)
Unclear	144 (7.6)	67 (34.5)	77 (40.7)
Type of primary outcome (N [%])			
Dichotomous	143 (37.3)	71 (36.6)	72 (38.1)
Continuous	240 (62.7)	123 (63.4)	117 (61.9)
Statistical significance (N [%])			
Yes	161 (42.0)	84 (43.3)	77 (40.7)
No/not reported	222 (57.9)	110 (56.7)	112 (59.3)
Direction of result (N [%])	*n =161*	*n =83*	*n =77*
Favour treatment	154 (95.6)	80 (95.2)	74 (96.1)
Favour control/unclear	7 (4.3)	4 (4.8)	3 (3.9)
Authors' conclusions (N [%])			
Endorsing treatment	210 (54.8)	108 (55.7)	102 (54.0)
Not endorsing treatment	173 (45.1)	86 (44.3)	87 (46.0)

‡ Statistically significant difference at 5% level
NS = Not specified.

Table 4.3: Characteristics of the published manuscripts from RCT/CCT abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

Characteristics	**Manuscripts (N = 194)**
Country (N [%])	
USA	170 (87.6)
Canada	17 (8.8)
Others	7 (3.6)
Number of authors	
(Mean [SE]; Median [IQR])	5.08 (0.18); 5 (3, 6)
Reported university affiliation (N [%])	164 (84.5)
Declaration of funding (N [%])	104 (53.6)
Government	28 (14.4)
Industry	53 (27.3)
Others	42 (21.64)
Study design (N [%])	
Parallel	162 (83.5)
Crossover/factorial	32 (16.5)
Number of study centers (N [%])	
Single center	157 (80.9)
Multicenter	37 (19.1)
Sample size (Median [IQR])	75.5 (40, 164.7)
Primary outcome (N [%])	
Clear	149 (76.8)
Unclear	45 (23.2)
Type of primary outcome (N [%])	
Dichotomous	68 (35.1)
Continuous	126 (64.9)
Conclusions (N [%])	
Statistical significance (yes)	89 (45.9)
Direction of result (favours treatment)	80 (41.2)
Conclusion (author's endorsement)	94 (48.5)

Table 4.4: Inter-observer agreement for the evaluation of the methodological quality of CCT/RCT abstracts presented at the Society for Academic Emergency medicine meetings from 1995-2003.and their published manuscripts.

Quality measure	**Abstracts (n = 383)**	**Manuscripts (n = 194)**
Jadad score [ICC (95% CI)]	0.84 (0.78,0.88)	0.93 (0.88,0.96)
Individual components [Kappa (95% CI)]		
Randomization	0.93(0.82, 1.00)	0.96 (0.89, 1.00)
Randomization bonus	0.83 (0.63, 1.00)	0.95 (0.88, 1.00)
Randomization deduction	0.90 (0.79, 1.00)	1.00
Double blinding	0.97 (0.94, 1.00)	0.95 (0.88, 1.00)
Double blinding bonus	0.60 (0.35, 0.85)	0.86 (0.73, 0.99)
Double blinding deduction	0.95 (0.89, 0.99)	1.00
Withdrawals/dropouts	0.64 (0.49, 0.79)	0.64 (0.43, 0.85)
Allocation concealment [Kappa (95% CI)]	0.77 (0.62, 0.92)	0.87 (0.76, 0.98)

Table 4.5: Methodological quality of CCT/RCT abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.and their subsequent published manuscripts.

Quality measure	Abstracts		Manuscripts
	Unpublished (n = 189)	Published (n = 194)	(n = 194)
Jadad score [Median (IQR)]	1 (1,1)	1 (1,1)	3 (2, 2) ¥
Individual components [% (95%CI)]			
Randomization	91.0 (86.8, 95.1)	94.3 (90.9, 97.6)	93.3 (89.7, 96.8)
Randomization bonus	2.6 (0.3, 4.9)	1.0 (0.1, 2.4)	54.1 (46.9, 61.2) ¥
Randomization deduction	2.1 (0.1, 4.1)	3.6 (1.0, 6.2)	3.1 (0.6, 5.5)
Double blinding	37.6 (30.6, 44.6)	46.4 (39.2, 53.5)	46.9 (39.7, 54.0)
Double blinding bonus	4.8 (1.7, 7.9)	8.2 (4.2, 12.1)	38.1 (31.1, 45.0) ¥
Double blinding deduction	0.5	0	0
Withdrawals/dropouts	6.9 (3.7, 10.0)	4.6 (1.6, 7.6)	53.6 (46.4, 60.7) ¥
Allocation concealment [%, 95%CI)]			
Adequate	0.5	0.5	37.1 (30.1, 44.0) ¥
Unclear	97.4 (95.1, 99.7)	95.4 (92.4, 98.4)	57.2 (50.1, 64.3) ¥
Inadequate	2.1 (0.1, 4.1)	4.1 (1.2, 6.9)	5.7 (2.3, 9.0) ¥

¥ Statistically significant difference at 1% level. Published abstracts vs. manuscripts; Wilcoxon Sign Rank test for paired comparison of the median Jadad scores; McNemar test (2x2) and McNemar-Bowker test (KxK) for paired comparisons of proportions.

Table 4.6: Predictors of RCT/CCT abstract publication in full manuscript form - univariate and multivariate logistic regression results.

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Country (1 = USA, 0 = Other countries)	0.75 (0.40, 1.42)	NA
Number of authors (Continuous)	1.08 (0.99, 1.17)*	1.07 (0.98, 1.16)
Reported university affiliation (0 = No, 1 = Yes)	1.58 (1.03, 2.42)*	1.53 (0.99, 2.35)
Study design (0 = Parallel, 1 = Crossover)	1.14 (0.65, 2.02)	NA
Number of study centers (0 = Single, 1 = Multi)	0.86 (0.49, 1.49)	NA
Sample size (Continuous)	1.00 (0.99, 1.00)	NA
Primary outcome (0 = Unclear, 1 = Clear)	1.30 (0.86, 1.97) *	‡
Type of primary outcome (0 = Dichotomous, 1 = Continuous)	1.06 (0.70, 1.61)	NA
Statistical significance (0 = No/not reported, 1 = Yes)	1.11 (0.74, 1.66)	NA
Direction of result (0 = Favor control/unclear, 1 = Favor intervention)	0.81 (0.17, 3.74)	NA
Authors' conclusions (0 = Not endorsing treatment, 1 = Endorsing treatment)	1.07 (0.71, 1.60)	NA
Quality (Jadad score, continuous)	1.13 (0.87, 1.46)	NA

* Variables included in the multivariate analysis ($p < 0.2$)
NA = Not applicable as not included in the multivariate analysis
‡ Removed from the multivariate model ($0.2 < p < 0.3$)

Table 4.7: Predictors of RCT/CCT abstract publication in full manuscript form - Cox's proportional hazard model results.

Variable	Unadjusted HR (95% CI)	Adjusted HR (95% CI)
Country (0 = Other countries, 1 = USA)	0.64 (0.42, 0.99) *	0.67 (0.44, 1.03)
Number of authors (Continuous)	1.05 (0.99, 1.10) *	1.03 (0.98, 1.09)
Reported university affiliation (0 = No, 1 = Yes)	1.47 (1.07, 2.01) *	**1.42 (1.03, 1.95)**
Study design (0 = Parallel, 1 = Crossover)	1.22 (0.83, 1.79)	NA
Number of study centers (0 = Single, 1 = Multi)	¥	NA
Sample size (Continuous)	1.19 (0.89, 1.58)	NA
Primary outcome (0 = Unclear, 1 = Clear)	1.2 (0.89, 1.16)	NA
Type of primary outcome (0 = Dichotomous, 1 = Continuous)	0.97 (0.72, 1.30)	NA
Statistical significance (0 = No/not reported, 1 = Yes)	1.01 (0.76, 1.35)	NA
Direction of result (0 = Favor control/unclear, 1 = Favor intervention)	1.03 (0.37, 2.83)	NA
Authors' conclusions (0 = Not endorsing treatment, 1 = Endorsing treatment)	1.01 (0.76, 1.35)	NA
Quality (Jadad score, continuous)	1.10 (0.91, 1.33)	NA

¥ = Proportional Hazard assumption not met.
* Variables included in the multivariate analysis ($p < 0.2$)
NA = Not applicable as not included in the multivariate analysis
‡ Removed from the multivariate model ($0.2 < p < 0.3$)

Table 4.8: Changes in authors' conclusions from abstract to manuscript form for abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

		Manuscripts	
		Not endorsing treatment	Endorsing treatment
Abstracts	Not endorsing treatment	81	5
	Endorsing treatment	**19**	89

Table 4.9: Predictors of RCT/CCT publication in higher impact journals for abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

Variable	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Country *	0.34 (0.13, 0.88)	‡
Number of authors *	1.23 (1.08, 1.40)	1.11 (0.95, 1.29)
Reported university affiliation *	2.57 (1.11, 5.95)	2.03 (0.79, 5.18)
Study design *	0.28 (0.12, 0.67)	‡
Number of study centers *	3.53 (1.60, 7.79)	‡
Sample size *	3.14 (1.74, 5.64)	1.82 (0.91, 3.64)
Primary outcome	1.00 (0.51, 1.95)	NA
Type of primary outcome *	0.33 (0.18, 0.62)	**0.42 (0.20, 0.86)**
Statistical significance	0.96 (0.54, 1.69)	NA
Direction of result	0.98 (0.55, 1.74)	NA
Authors' conclusions	1.38 (0.78, 2.44)	NA
Quality (Jadad score below/above median) *	1.90 (1.07, 3.39)	**2.26 (1.17, 4.33)**
Declaration of funding *	3.04 (1.69, 5.47)	1.93 (0.96, 3.88)

* Variables included in the multivariate analysis ($p < 0.2$)
NA = Not applicable as not included in the multivariate analysis
‡ Removed from the multivariate model ($0.2 < p < 0.3$)

Figure 4.1: Flow-diagram for identification of RCT/CCT abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.

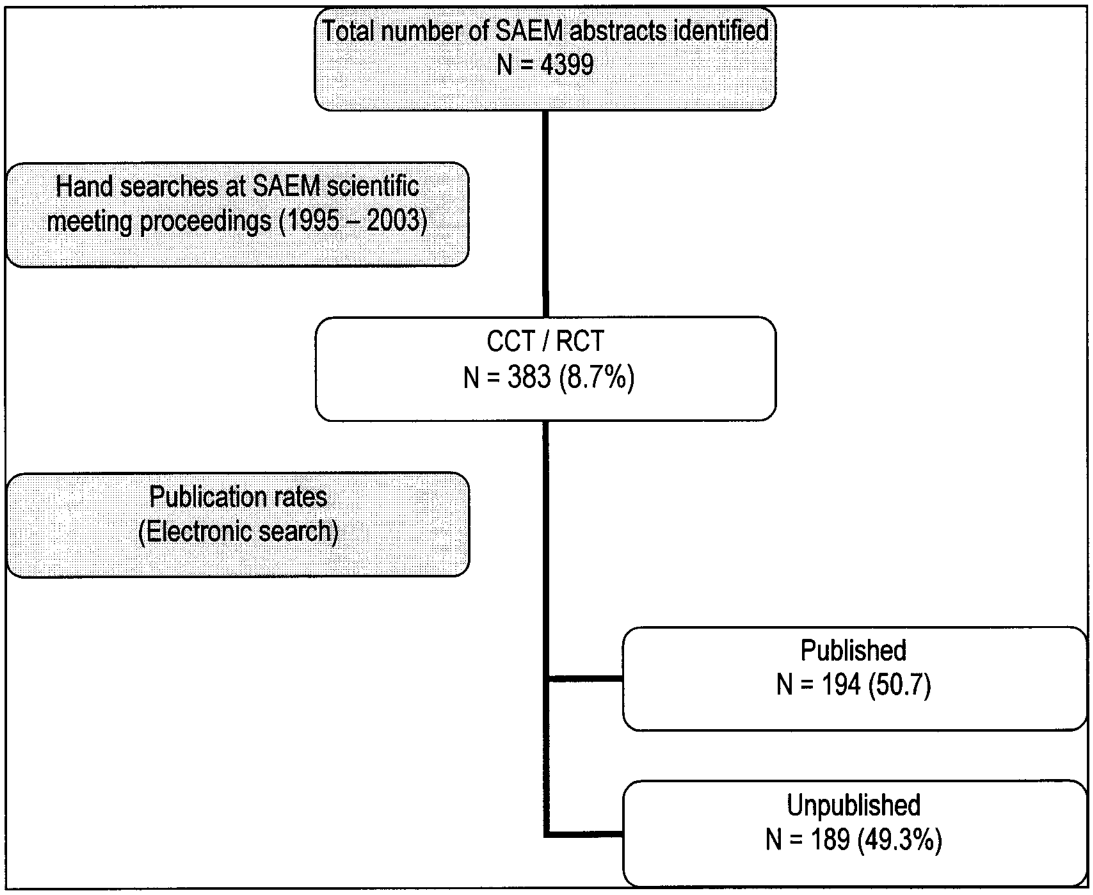


Figure 4.2: Research category of RCT/CCT abstracts presented at the Society for Academic Emergency medicine meetings from 1995-2003.

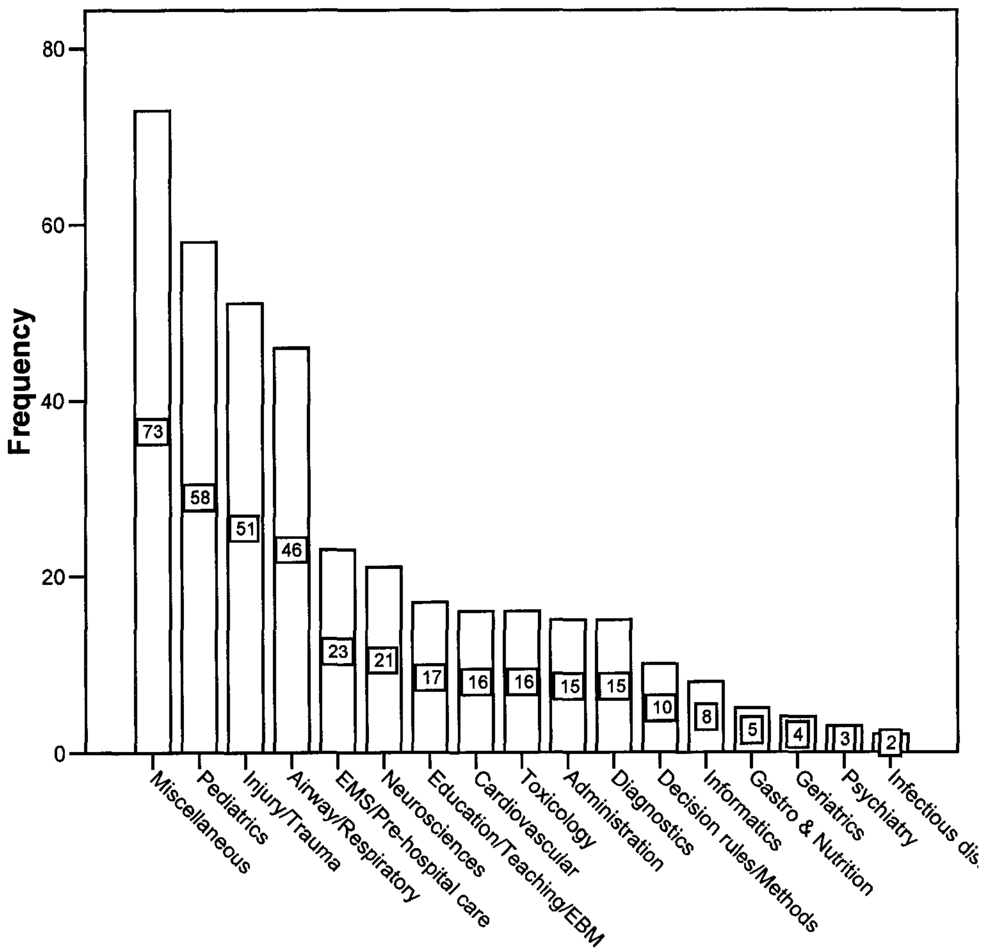


Figure 4.3: Time to publication of CCT/RCT abstracts presented at the Society for Academic Emergency medicine meetings from 1995-2003.

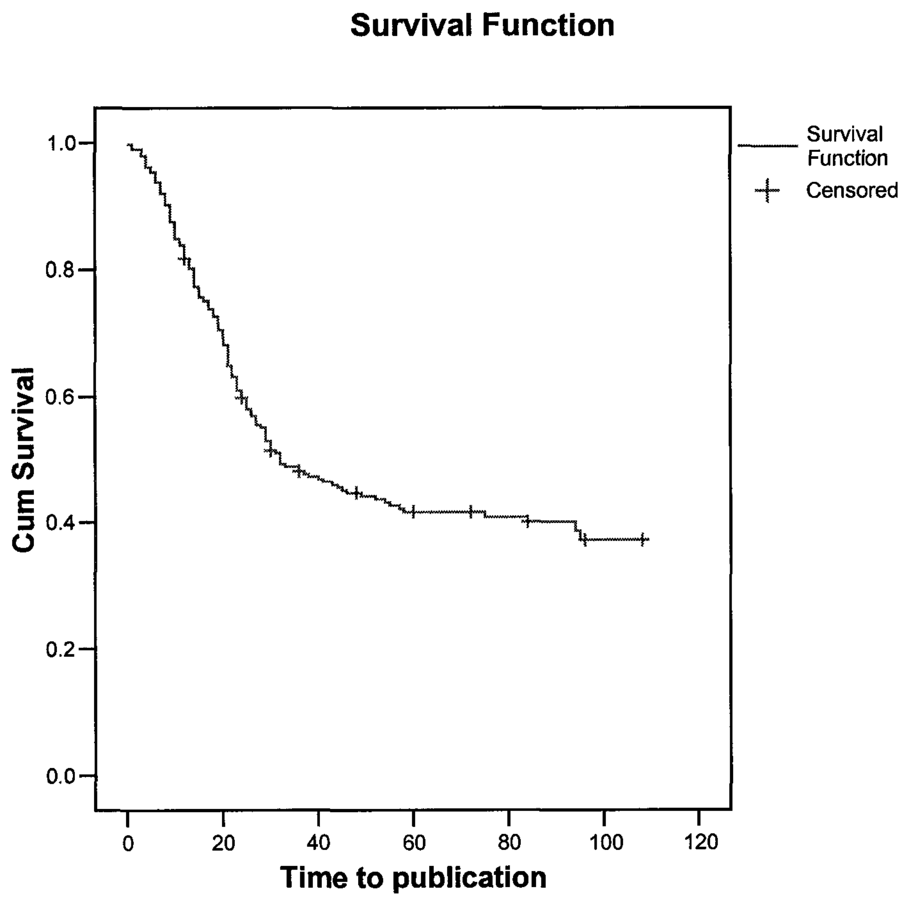
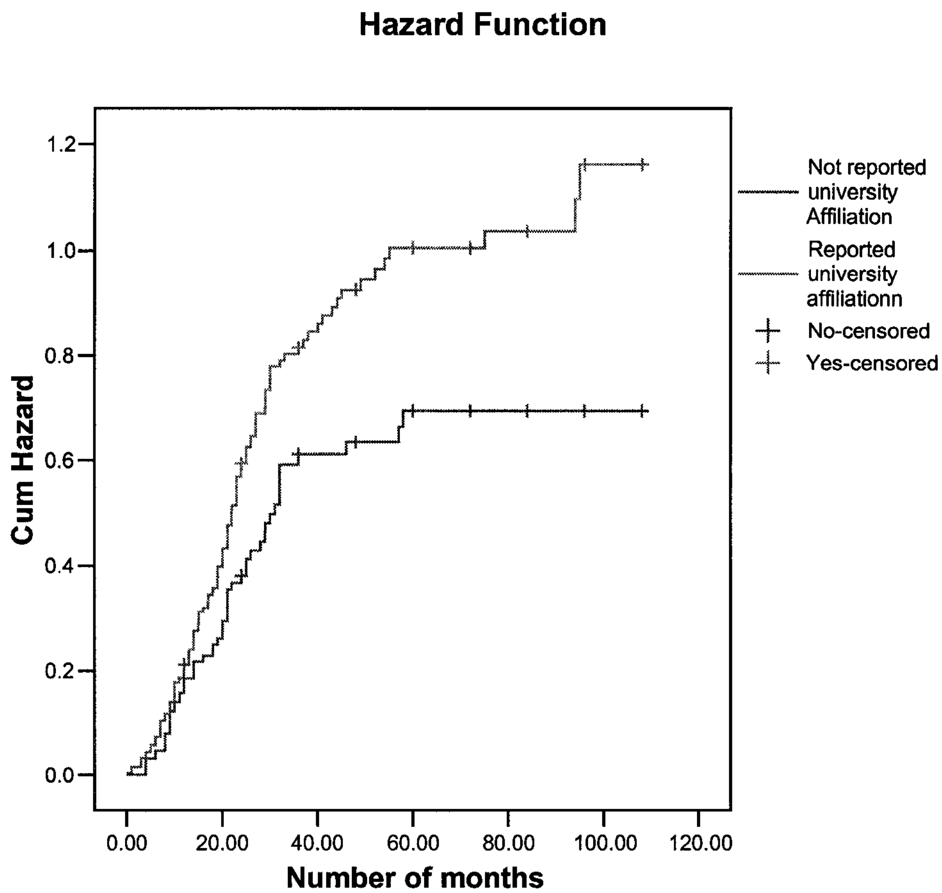


Figure 4.4: Time to publication by report of university affiliation for abstracts presented at the Society for Academic Emergency Medicine meetings from 1995-2003.



4.7. References

(1) Anonymus. The reporting of unsuccessful cases (editorial). Boston Medical and Surgery Journal 1909;263-264.

(2) Boissel JP, Haugh MC. The iceberg phenomenon and publication bias: the editors' fault? Clin Trials Metaanal 1993; 28:309-315.

(3) Rosenthal R. The "file drawer problem" and tolerance for null results. Psychol Bull 1979; 86:638-641.

(4) Evers JL. Publication bias in reproductive research. Hum Reprod 2000; 15:2063-2066.

(5) Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990; 263:1385-1389.

(6) Song F, Eastwood AJ, Gilbody S, Dulley L, Sutton AJ. Publication and related biases. Health Technol Assess 2000; 4:1-115.

(7) Rennie D, Flanagin A. Publication bias. The triumph of hope over experience. JAMA 1992; 267:411-412.

(8) Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337:867-872.

(9) Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. JAMA 1994; 272:158-162.

(10) Eloubeidi MA, Wade SB, Provenzale D. Factors associated with acceptance and full publication of GI endoscopic research originally published in abstract form. Gastrointest Endosc 2001; 53:275-282.

(11) Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. JAMA 1998; 280:254-257.

(12) Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. JAMA 1998; 280:257-259.

(13) Petticrew M, Gilbody S, Song F. Lost information? The fate of papers presented at the 40th society for Social Medicine Conference. J Epidemiol Community Health 1999; 53:442-443.

(14) Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32:51-63.

(15) Dickersin K. NIH trials and publication bias. Online J Curr Clin Trials 1993; 50.

(16) De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: Factors which influence presentation and subsequent publication. Ann Oncol 1992; 3:187-191.

(17) Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. Control Clin Trials 1987; 8:343-353.

(18) Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998; 279:281-286.

(19) Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. JAMA 1998; 279:319-320.

(20) Jasko JJ, Wood JH, Schwartz HS. Publication rates of abstracts presented at annual musculoskeletal tumor society meetings. Clin Orthop 2003;10: 98-103.

(21) Klassen TP, Wiebe N, Russell K, Stevens K, Hartling L, Craig WR et al. Abstracts of randomized controlled trials presented at the Society for Pediatric Research meeting. Arch Pediatr Adolesc Med 2002; 156:474-479.

(22) Ben-Schlomo Y, Davey-Smith G. "Place of publication" bias. BMJ 1994; 309:274.

(23) Timmer A, Hilsden RJ, Cole J, Hailey D, Sutherland LR. Publication bias in gastroenterological research - a retrospective cohort study based on abstracts submitted to a scientific meeting. BMC Med Res Methodol 2002; 2:7.

(24) Hopewell S, McDonald S. Full publication of trials initially reported as abstracts in the Australian and New Zealand Journal of Medicine 1980-2000. Intern Med J 2003; 33:192-194.

(25) Hopewell S, Clarke M. Methodologists and their methods. Do methodologists write up their conference presentations or is it just 15 minutes of fame? Int J Technol Assess Health Care 2001; 17:601-603.

(26) Simes RJ. Publication bias: The case for an international registry of clinical trials. J Clin Oncol 1986; 4:1529-1541.

(27) Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lircainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. Int J Cardiol 1993; 40:161-166.

(28) Chalmers I, Adams M, Dickersin K, Hetherington J, Tarnow-Mordi W, Meinert C et al. A cohort study of summary reports of controlled trials. JAMA 1990; 263:1401-1405.

(29) Chalmers I. Underreporting research is scientific misconduct. JAMA 1990; 263:1405-1408.

(30) Scherer RW, Langenberg P. Full publication of results initially presented in abstracts: revisited (Cochrane Methodology Review). In: *The Cochrane Library*, Issue 1, 2004. John Wiley and Sons: Chichester, UK.

(31) Singer AJ, Homan CS, Brody M, Thode HC Jr, Hollander JE. Evolution of abstracts presented at the annual scientific meetings of academic emergency medicine. Am J Emerg Med 1999; 17:638-641.

(32) Pocock SJ. Clinical trials: a practical approach. Chichester: John Wiley & Sons, 1997.

(33) Jadad A. Randomized controlled trials: a user's guide. London: BMJ Group, 1998.

(34) Fleiss JL. Statistical methods for rates and proportions. 2nd ed. John Wiley & Sons, 1981.

(35) Alderson P, Green S, Higgins JPT, editors. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003]. In: The Cochrane Library, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

(36) Altman D, Schulz K. Concealing treatment allocation in randomised trials. BMJ 2001; 323:446-447.

(37) Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996; 17:1-12.

(38) Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effect in controlled clinical trials. JAMA 1995; 273:408-412.

(39) Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998; 352:609-613.

(40) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174.

(41) Seigel DG, Podgor MJ, Remaley NA. Acceptable values of kappa for comparison of two groups. Am J Epidemiol 1992; 135:571-578.

(42) Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford, New York, Tokyo: Oxford University Press, 1989.

(43) McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996; 1:30-46.

(44) ISI. 2002 JCR Science Edition. ISI Web of Science 2004.

(45) von Elm E, Costanza MC, Walder B, Tramer MR. More insight into the fate of biomedical meeting abstracts: a systematic review. BMC Med Res Methodol 2003; 3:12.

(46) Walby A, Kelly AM, Georgakas C. Abstract to publication ratio for papers presented at scientific meetings: how does emergency medicine compare? Emerg Med (Fremantle) 2001; 13:460-464.

(47) Korn CS, Henderson R. Publication of research abstracts presented at four emergency medicine research forums. Acad Emerg Med 2000; 7:534.

(48) Castillo J, Garcia-Guasch R, Cifuentes I. [Publications derived from free communications at the XX Congress of the Anesthesiology and Resuscitation Spanish Association (Anesthesia 92)] [Spanish]. Rev Esp Anestes Reanim 2000; 47:53-56.

(49) Landry VL. The publication outcome for the papers presented at the 1990 ABA conference. J Burn Care Rehabil 1996; 17:23A-26A.

(50) Cheng K, Preston C, Ashby D, O'Hea U, Smyth RL. Time to publication as full reports of abstracts of randomized controlled trials in cystic fibrosis. Pediatr Pulmonol 1998; 26:101-105.

(51) Dickersin K, Min Y, Meinert C, Meinert CL. Factors influencing publication of research results: follow up of applications submitted to two institutional review boards. JAMA 1992; 267:867-872.

(52) Relman AS. News reports of medical meetings: how reliable are abstracts? N Engl J Med 1980; 303:277-278.

(53) Dickersin K, Min Y. Publication bias: the problem that won't go away. Ann NY Acad Sci 1993; 703:135-146.

(54) Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997; 315:640-645.

(55) Timmer A, Blum T, Lankisch PG. Publication rates following pancreas meetings. Pancreas 2001; 23:212-215.

(56) Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. JAMA 2003; 290:495-501.

(57) Diezel K, Pharoah FM, Adams CE. Abstracts of trials presented at the Vth World Congress of Psychiatry (Mexico, 1971): a cohort study. Psychol Med 1999; 29:491-494.

(58) Marx WF, Cloft HJ, Do HM, Kallmes DF. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. AJNR Am J Neuroradiol 1999; 20:1173-1177.

(59) Goldman L, Loscalzo A. Fate of cardiology research originally published in abstract form. N Engl J Med 1980; 303:255-259.

(60) Daluiski A, Kuhns CA, Jackson KR, Lieberman JR. Publication rate of abstracts presented at the annual meeting of the Orthopaedic Research Society. J Orthop Res 1998; 16:645-649.

(61) Dickersin K, Olson CM, Rennie D, Cook D, Flanagin A, Zhu Q et al. Association between time interval to publication and statistical significance. JAMA 2002; 287:2829-2831.

(62) Weintraub WH. Are published manuscripts representative of the surgical meeting abstracts? An objective appraisal. J Pediat Surg 1987; 22:11-13.

(63) Tramer M, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ 1997; 315:635-640.

(64) Davies MW, Dunster KR, East CE, Lingwood BE. Fate of abstracts published in the proceedings of the first annual Perinatal Society of Australia and New Zealand Congress in 1997. J Paediatr Child Health 2002; 38:501-506.

(65) Rubin HR, Redelmeier DA, Wu AW, Steinberg EP. How reliable is peer review of scientific abstracts? Looking back at the 1991 Annual Meeting of the Society of General Internal Medicine. J Gen Intern Med 1993; 8:255-258.

(66) McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence the estimates of intervention effectiveness reported in meta-analyses? Lancet 2000; 356:1228-1231.

(67) Moher D, Schachter HM. Potential solutions to the problem of conducting systematic reviews of new health technologies. CMAJ 2002; 166:1674-1675.

(68) Kurmis AP. Understanding the limitations of the journal impact factor. Bone Joint Surg Am 2003; 85:2449-2454.

(69) Moher D, Schulz KF, Altman D, CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA 2001; 285:1987-1991.

(70) Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000; 53:1119-1129.

(71) Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. BMJ 2000; 320:1574-1577.

(72) Egger M, Davey-Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315:629-634.

(73) Egger M, Davey-Smith G. Meta-analysis: Bias in location and selection of studies. BMJ 1998; 316:61-66.

CHAPTER 5: General discussion and conclusions

5.1. Overview

The preceding chapters have reviewed the previous research on the quality of systematic reviews in a variety of medical areas, described the methodological quality of systematic reviews published in leading emergency medicine journals, and explored the presence of publication bias on a cohort of randomized clinical trials (RCTs) abstracts presented at nine consecutive meetings of the Society of Academic Emergency Medicine (SAEM).

5.2. Results

The structured review of overviews on systematic reviews and/or meta-analyses presented in Chapter 2 revealed that a considerable proportion of studies on the quality of systematic reviews and/or meta-analyses in a variety of medical specialties have been conducted to date. Evidence from overviews showed a variable pattern of methodological quality across several medical and allied sciences.

Chapter 3 evaluated the methodological quality of systematic reviews published from 2000 to 2004 in major emergency medicine journals and compared the results with a cohort of systematic reviews in emergency medicine published before the publication of the QUOROM guidelines. Factors associated with methodological quality of systematic reviews in emergency medicine were also explored. These data suggest that while SRs are becoming more common in the EM literature, their quality remains unacceptably low. For example, many of the reviews published in major emergency medicine journals exhibit major methodological flaws limiting the validity of their conclusions. Moreover, there is limited evidence of an improvement in the conduct and reporting of these reviews following the publication of the QUOROM statement. Common flaws in the methods to locate the evidence, in documenting the measures to control for selection bias of primary studies, and in the methods to assess the validity of primary studies using appropriate criteria, were identified. No definite conclusions were reached regarding the direction of the association between the level of quality and the type of conclusions of the reviews and/or meta-analyses.

Chapter 4 examined the publication rate, time to publication and determinants of publication of a cohort of RCT abstracts accepted for presentation at nine consecutive SAEM scientific meetings. In particular, the occurrence of several types of publication bias (positive outcome bias, time-lag bias, grey literature bias, and place of publication bias) was of primary concern. The rate of publication was lower than the combined rates reported for other medical specialties; however, it was comparable to the findings reported for other emergency medicine meetings. The study failed to demonstrate evidence for a positive outcome bias in this sample of CCT/RCT SAEM abstracts. Associations between other predictors such as sample size, number of study centers, type of trial design, number of authors, among others and the publication outcome were not statistically significant. The lone variable that was statistically significantly associated with publication was the reporting of university affiliation by the authors. The methodological quality of the abstracts was not a predictor of publication; however, differences between the methodological quality from abstract to manuscript were detected in this study. This finding may suggest a type of publication bias that has not been described before and that can be described as "space bias": differences in the report of methodological characteristics of the study depend on the space available to report the study results. The median time to publication of the CCT/RCT SAEM abstracts was not associated with the study outcomes or with any other quality or study characteristics.

On the other hand, it is likely that a grey literature bias is present in this sample of CCT/RCT SAEM abstracts; that is to say, author's conclusions, either endorsing or not endorsing the treatment, changed from the scientific abstract to their corresponding publications. No evidence of place of publication bias was found for the cohort of RCT manuscripts derived from SAEM abstracts; however, manuscripts with higher methodological quality were more likely to be published in wider circulating journals.

5.3. Limitations

Potential limitations of this critical evaluation of evidence-based emergency medicine literature under the form of systematic reviews and RCTs need to be considered when interpreting the results from the evaluations.

The evaluation of the evidence from systematic reviews in emergency medicine was confined to English-language publications in a sample of emergency medicine journals and it is unknown whether there may be a publication bias against systematic reviews in emergency medicine that do not have significant findings. Second, the methods used to evaluate the systematic reviews and/or meta-analyses could be criticized; and a call for new methodological approaches to systematic review quality is advocated.

With respect to the analysis of the effect of publication bias in emergency medicine RCTs, it is important to recognize that the evidence evaluated in the present study only reflects reports of RCTs that were accepted at nine scientific meetings under abstract format and does not necessarily reflect the state of all research in the field of emergency medicine. Second, the description of the publication experience may have been affected by the search methods that were used in this study and the rate of publication in other non-indexed publications was not explored.

5.4. Recommendations and future directions

An important proportion of studies on the quality of systematic reviews and/or meta-analyses in a variety of medical specialties have been conducted to date; however, a variable pattern of methodological quality is demonstrated across several medical and allied sciences. Future studies should assess the sources of heterogeneity and variations in the level of quality of systematic reviews and/or meta-analyses. They should also evaluate the effect of the introduction of guidelines such as QUOROM on the level of methodological quality of systematic reviews and/or meta-analyses produced in a variety of areas of academic medicine and allied sciences. The use of validated instruments such as the OQAQ index in future evaluations is preferred to measure these changes in order to enable further comparisons across several medical and allied specialties. Since the OQAQ index tends to overestimate the value of meta-analyses compared to qualitative systematic reviews, new validated approaches to evaluate the methodological quality of systematic reviews should be developed.

Systematic reviews of RCTs are considered the "gold standard" for clinical and policy decisions in emergency medicine and other fields. Nevertheless, this statement cannot be confidently endorsed until a uniform and appropriate level of methodological quality is reached for the systematic reviews produced by researchers publishing their data in emergency medicine journals. In the meantime, editors and authors should concentrate their efforts in adopting standardized guidelines for submission and appraisal for publication of higher quality systematic reviews and/or meta-analyses in emergency medicine journals. Specifically, all journals should immediately endorse and adopt the QUOROM guidelines for the submission and acceptance of manuscripts describing systematic reviews and/or meta-analyses. Finally, readers of emergency medicine journals should also be aware of the methodological limitations of systematic reviews and/or meta-analysis and critically appraise their results before implementing their results in clinical practice.

The quality of reporting emergency medicine RCT/CCT abstracts must also improve if they want to be employed as a source of evidence for future research or systematic reviews. The decision of restricting the abstract to 250 words should be revisited to allow for inclusion of more comprehensive methodological information in the body of the abstract to enhance its value and validity. Particularly, guidelines to improve structured abstracts should be developed to include a more detailed description of the methodology of the studies that considers the adequacy of the methods of randomization and blinding, the concealment of allocation to treatments and the description of the withdrawals and dropouts. A format based on the CONSORT statement would be preferred, and certainly could be accomplished without a space expansion. Given the differences between emergency medicine RCT/CCT abstracts and subsequent manuscripts that were found in this evaluation, caution is warranted with respect to employing this grey literature as a source of evidence for future research or systematic reviews in emergency medicine. Information cannot be considered reliable, as inconsistencies with information at the manuscript level were frequent in the direction of the primary outcome results.

More studies exploring the nature of publication bias in emergency medicine research should be undertaken. Retrospective cohort-based studies identifying RCTs at an early stage before full publication (e.g. abstracts submitted but not necessarily accepted to scientific meetings) would be useful to have a more comprehensive perspective on how emergency medicine research is affected by a variety of publication biases. Further studies (i.e. online author's surveys and focus groups) assessing potential publication barriers and other determinants of subsequent RCT/CCT publication are warranted.

The proposal for a trial registry in emergency medicine should be explored to improve the accessibility of trial information and data. Joint solutions among clinicians, researchers and journal editors will surely facilitate the implementation and dissemination of this endeavour. A call for a collaborative work in this area should therefore be a priority for the research agenda in academic emergency medicine.

Appendix A: Search strategies for the literature reviews

A.1: Search strategy for Chapter 2: Overview of Quality of Reports of Systematic Reviews.

Medline (Ovid): 1966 – May 2004

Search terms	Yielded records
1 review literature/	1041
2 meta-analy$.tw.	10983
3 metaanal$.tw.	405
4 (systematic$ adj4 (review$ or overview$)).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]	6984
5 review.pt.	1038205
6 review.ti.	113138
7 review literature.pt.	37757
8 or/1-7	1096410
9 case report/	0
10 letter.pt	507113
11 historical article.pt.	209332
12 review of reported cases.pt.	49732
13 review,multicase.pt.	8248
14 or/9-13	768704
15 (quality adj assess$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]	3423
16 8 not 14	1028361
17 16 and 15	978

A.2: Search strategies for identification of systematic reviews in emergency medicine.

a) MEDLINE (Ovid) search strategy adapted from the Evidence Based Informatics Project at McMaster University: 1966 – March Week 1 2004.

Search terms	Yielded records
1 meta-analysis.pt,sh.	12959
2 (meta-anal: or metaanal:).tw.	10693
3 (quantitativ: review: or quantitativ: overview:).tw.	225
4 (systematic: review: or systematic: overview:).tw.	5450
5 (methodologic: review: or methodologic: overview:).tw.	147
6 (integrative research review: or research integration:).tw.	77
7 review.pt,sh. or review:.tw. or overview:.tw.	1276151
8 quantitativ: synthes:.tw.	85
9 1 or 2 or 3 or 4 or 5 or 6 or 8	20378
10 (medline or medlars).tw,sh. or embase.tw.	14529
11 (scisearch or psychinfo or psycinfo).tw.	490
12 (psychlit or psyclit).tw.	545
13 (hand search: or manual search:).tw.	1423
14 (electronic database: or bibliographic database:).tw.	1198
15 (pooling or pooled analys: or mantel haenszel).tw.	4927
16 (peto or der simonian or dersimonian or fixed effect:).tw.	1429
17 or/10-16	21146
18 17 and 7	12846
19 9 or 18	29297
20 emergency medicine.mp.	6432
21 19 and 20	51
22 limit 21 to yr=2000- 2004	**26**

A.2: Search strategies for identification of systematic reviews in emergency medicine (Cont'd).

b) MEDLINE (Ovid) search using Kelly et al {58} search strategy (via OVID): 1966 – March Week 1 2004

	Yielded records
1 meta analysis.pt.	8644
2 exp meta-analysis/	5214
3 metaanal:.tw.	394
4 (meta adj anal:).tw.	10348
5 (quantitative: review: or quantitative: overview:).tw.	255
6 (systematic: review: or methodol: overview:).tw.	5318
7 (methodol: review: or analytic$: overview:).tw.	275
8 (analytic: review: or analytic: overview:).tw.	430
9 or/1-8	20524
10 (medline or embase or index medicus).ti,ab,sh.	13870
11 (pooled or pooling).ti,ab,sh.	18468
12 (((combined or combining) adj data) or trials or studies or reviews).mp. [mp=title, abstract, name of substance, mesh subject heading]	1130372
13 or/10-12	1147971
14 13 and review.pt.	195032
15 9 or 14	209562
16 exp Emergency Medicine/	4935
17 15 and 16	101
18 limit 17 to yr=2000- 2004	**33**

A.2: Search strategies for identification of systematic reviews in emergency medicine (Cont'd).

c) EMBASE (Ovid) search strategy developed for use in the Protocol Enhancement Project from the McMaster Medline strategy: 1988 – 2004 Week 10

		Yielded records
1	exp meta analysis/	17725
2	meta-anal$.ti,ab,hw,tn,mf.	21158
3	metaanal$.ti,ab,hw,tn,mf.	624
4	quantitativ$ review$.ti,ab,hw,tn,mf.	174
5	quantitativ$ overview$.ti,ab,hw,tn,mf.	48
6	systematic$review$.ti,ab,hw,tn,mf.	0
7	systemtic$ overview$.ti,ab,hw,tn,mf.	0
8	methodologic$ review$.ti,ab,hw,tn,mf.	84
9	methodologic$ overview$.ti,ab,hw,tn,mf.	24
10	integrative research review$.ti,ab,hw,tn,mf.	5
11	research integration$.ti,ab,hw,tn,mf.	28
12	quantitativ$ synthes$.ti,ab,hw,tn,mf.	73
13	(medline or medlars or embase).ti,ab,hw,tn,mf.	11022
14	hand search$.ti,ab,hw,tn,mf.	466
15	manual search$.ti,ab,hw,tn,mf.	461
16	(pooling or pooled analy$).ti,ab,hw,tn,mf.	2829
17	mantel haenszel.ti,ab,hw,tn,mf.	894
18	(peto or der simonian or dersimonian or fixed effect$).ti,ab,hw,tn,mf.	753
19	(scisearch or psychinfo or psycinfo or psychlit or psyclit).ti,ab,hw,tn,mf.	702
20	(electronic database$ or bibliographic database$).ti,ab,hw,tn,mf.	1098
21	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	34011
22	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	34011
23	emergency medicine.mp.	8494
24	22 and 23	83
25	limit 24 to yr=2000- 2004	**37**

A3: Emergency Medicine journals for hand searching.

Journal name	SCI Impact factor for 2002 {144}
Academic Emergency Medicine (AEM)	1.535
American Journal of Emergency Medicine (AJEM)	1.208
Annals of Emergency Medicine (AOEM)	2.148
Emergency Medicine (EM)	0.591
Emergency Medicine Journal (EMJ)(former Journal of Accident and Emergency Medicine Journal)	0.565
Canadian Journal of Emergency Medicine (CJEM)‡	-------
Journal of Emergency Medicine (JEM)	0.743
Mean Impact factor (Median)	**1.131 (0.743)**

‡ Not indexed journal.

APPENDIX B: Instruments for quality assessment

B.1 Overview Quality Assessment Questionnaire (OQAQ).

	Quality features	1	2	3
1	Were the search methods used to find evidence on the primary question(s) stated?	No	Partially	Yes
2	Was the search for evidence reasonably comprehensive?	No	Can't tell	Yes
3	Were the criteria used for deciding which studies to include in the overview reported?	No	Partially	Yes
4	Was bias in the selection of studies avoided?	No	Can't tell	Yes
5	Were the criteria used for assessing the validity of the included studies reported?	No	Partially	Yes
6	Was the validity of all the studies referred to in the text assessed using appropriate criteria?	No	Can't tell	Yes
7	Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?	No	Partially	Yes
8	Were the findings of the relevant studies combined appropriately relative to the primary question of the overview?	No	Can't tell	Yes
9	Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?	No	Partially	Yes
10	How would you rate the scientific quality of this review?	1-7		

<table>
<tr><th colspan="7">Flaws</th></tr>
<tr><td colspan="2">Extensive</td><td></td><td colspan="3">Minor</td><td></td></tr>
<tr><td></td><td colspan="3">Major</td><td></td><td colspan="2">Minimal</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr>
</table>

Note: If the methods that were used are reported incompletely relative to a specific item, score that item as "partially". Similarly, if there is no information provided regarding what was done relative to a particular question, score it as "can't tell", unless there is information in the overview to suggest either that the criterion was or was not met.

For Question 8, if no attempt has been made to combine findings, and no statement is made regarding the inappropriateness of combining findings, check "no". If a summary (general) estimate is given anywhere in the abstract, the discussion, or the summary section of the paper, and it is not reported how that estimate was derived, mark "no" even if there is a statement regarding the limitations of combining the findings of the studies reviewed. If in doubt mark "can't tell".

For an overview to be scored as "yes" on Question 9, data (not just citations) must be reported that support the main conclusions regarding the primary question(s) that the overview addresses.

The score for Question 10, the overall scientific quality, should be based on your answers to the first nine questions. The following guidelines can be used to assist with deriving a summary score: If the "can't tell" option is used one or more times on the preceding questions, a review is likely to have minor flaws at best and it is difficult to rule out major flaws (i.e., a score of 4 or lower). If the "no" option is used on Questions 2, 4, 6 or 8, the review is likely to have major flaws (i.e., a score of 3 or less, depending on the number and degree of the flaws).

B.2 Jadad scale for quality assessment of randomized controlled trials.

	Quality features	Score
1.	Was the study described as randomized (this includes the use of words such as randomly, random and randomization)? Yes = 1 No = 0	
2.	Was the study described as double-blind? Yes = 1 No = 0	
3.	Was there a description of withdrawals and drop-outs? Yes = 1 No = 0	
Additional points	Add 1 point if:	
	Method to generate the sequence of randomization was described and was appropriate (e.g. table of random numbers, computer generated, coin tossing, etc.)	
	Method of double-blinding described and appropriate (identical placebo, active placebo, dummy)	
Point deduction	Subtract 1 point if:	
	Method of randomization described and it was **in**appropriate (allocated alternately, according to date of birth, hospital number, etc.)	
	Method of double-blinding described but it was **in**appropriate (comparison of tablet vs. injection with no double dummy)	
	OVERALL SCORE (Maximum 5)	

B.3 Concealment of treatment allocation.

Concealment of treatment allocation:		Adequate Inadequate Unclear
Adequate:	e.g. central randomization; numbered/coded containers; drugs prepared by pharmacy; serially numbered, opaque, sealed envelopes	
Inadequate:	e.g. alternation, use of case record numbers, dates of birth or day of week; open lists	
Unclear:	Allocation concealment approach not reported or fits neither above category	

APPENDIX C: Data collection forms

C.1: Eligibility form for Chapter 3: Systematic reviews in emergency medicine.

CITATION #			

REVIEWER	BR	MO

Please, assess whether the article meets the following criteria:

INCLUSION CRITERIA

1. Research question:			
An identifiable question was formulated	YES	NO	UNSURE/CAN'T TELL

2. Search:			
A search strategy is described to identify studies for inclusion	YES	NO	UNSURE/CAN'T TELL

3. Analysis:			
Authors attempted to summarize and analyze data from the primary studies	YES	NO	UNSURE/CAN'T TELL

EXCLUSION CRITERIA

Format of publication			
The article is NOT any of the following: editorial, correspondence, abstract, EBEM shortcut reviews	YES	NO	UNSURE/CAN'T TELL

FINAL DECISION:

	INCLUDED (meets the eligibility criteria above)
	NOT INCLUDED (fails to meet any of the eligibility criteria above)
	CAN'T TELL (needs more information from authors to make a decision)

Observations

C.2: Data collection form for Chapter 3: Systematic reviews in emergency medicine.

CITATION #			

Date: __/__/__	MO	BHR

1. Number of authors:

1	2	3	4	5+

2. Journal of Publication:

Academic Emergency Medicine	
American Journal of Emergency Medicine	
Annals of Emergency Medicine	
Emergency Medicine	
Emergency Medicine Journal	
Canadian Journal of Emergency Medicine	
Journal of Emergency Medicine	

3. Publication year:

	2000		2002
	2001		2003

4. Country of corresponding author:

	Australia		Canada
	Germany		The Netherlands
	United Kingdom		USA
	Other (specify)		

5. Topic of study:

	Intervention/Therapy
	Diagnosis
	Other (specify)

6. Sources of studies:

			Bibliographic databases	
	MEDLINE		EMBASE	
	CINAHL		HealthSTAR	
	PubMed		Science Citation Index	
	PsychLIT		CENTRAL - Cochrane	
	Other(s) (specify)			
			Hand-searches of journals	
			Reference list	
			Contact with authors	
			Contact with appropriate industry sector	
			Conference proceedings	
			Publication restrictions?__________	
			Language restrictions?__________	

7. Design of primary studies included:

	RCT		CCT
	Before & After studies		Cohort studies
	Case-control studies		Cross-sectional
	Other (specify)		

8. Quality assessment criteria used:

NO	YES	If YES, which one:

9. Type of analysis:

	Qualitative		Quantitative (Meta-analysis)

10. Main study results:

	Positive		Negative/ Uncertain

11. Source of funding

	Government grant		Internal funds
	Pharmaceutical company		Charity
	Not described		Other (specify)

OBSERVATIONS:

____
____
____
____
____

C.2: Data collection forms for Chapter 4: Randomized controlled trials in emergency medicine.

ABSTRACT

ID ABSTRACT		**Published:**	Yes (1)	No (0)	**Year of meeting**		**Number authors**	
Title of abstract					**Authors' names**			
Classification	Decision Rules/Methods (6)	Gastro & Nutrition (12)			**Country first author**		**Affiliation university**	Yes (1)
Administration (1)	Diagnostics (7)	Geriatrics (13)						No (0)
Airway & Respiratory (2)	Education/Teaching/EBM (8)	Infectious Diseases (14)			**QUALITY ASSESSMENT – Jadad scale**			
Cardiovascular (3)	EMS/Pre Hospital Care (9)	Informatics (15)			Randomization?		Double blind bonus?	
Injury/Trauma (4)	Neurosciences (10)	Pediatrics (16)			Randomization bonus?		Double blind deduction?	
Psychiatry (5)	Toxicology (11)	Other no classified (17)			Randomization deduction?		Withdrawals described?	
Type funding	**Declared**	Yes (1)	No (2)		Double blind study?		Jadad score	
Government (Yes 1, No 0)	Associations (Yes 1, No 0)				**CONCEALMENT OF TREATMENT ALLOCATION**			
Industry (Yes 1, No 0)	Other (s)	None			Adequate (1)	Inadequate (0)	Unclear (2)	
RESULTS								
Study design	Parallel (0)	Factorial (1)	Crossover (2)	Other (S) (3)	**Multicenter**	Yes (1)	No (0)	
Pilot study	Yes (1)	No (0)			**Total sample size (N)**		**# Withdrawals**	
Population					**Primary outcome**	Selected (0)	Inferred (1)	Stated (2)
Author's conclusions					**Significance**	Yes (1)	No (0)	Not reported (2)
Positive (1)	Negative (0)	Neutral/unclear (2)			**If significant**	Favour control (0)	Favour Tx (1)	Unclear (2) / N/A (3)
Type of primary outcome	Dichotomous (0)				Continuous (1)			

C.2: Data collection forms for Chapter 4: Randomized controlled trials in emergency medicine (Cont'd).

MANUSCRIPT

ID Manuscript		**Journal:**		**Year publ.**		**Month publ.**			
Date submission				**Date acceptance**				**Number authors**	
Title of manuscript				**Authors' names**					
Classification	Decision Rules/Methods (6)	Gastro & Nutrition (12)	**Country first author**		**Affiliation university**	Yes (1)			
Administration (1)	Diagnostics (7)	Geriatrics (13)				No (0)			
Airway & Respiratory (2)	Education/Teaching/EBM (8)	Infectious Diseases (14)	**QUALITY ASSESSMENT – Jadad scale**						
Cardiovascular (3)	EMS/Pre Hospital Care (9)	Informatics (15)	Randomization?		Double blind bonus?				
Injury/Trauma (4)	Neurosciences (10)	Pediatrics (16)	Randomization bonus?		Double blind deduction?				
Psychiatry (5)	Toxicology (11)	Other no classified (17)	Randomization deduction?		Withdrawals described?				
Type funding	**Declared**	Yes (1)	No (0)	Double blind study?		Jadad score			
Government (Yes 1, No 0)	Associations (Yes 1, No 0)	**CONCEALMENT OF TREATMENT ALLOCATION**							
Industry (Yes 1, No 0)	Other (s)	None	Adequate (1)	Inadequate (0)	Unclear (2)				
RESULTS									
Study design	Parallel (0)	Factorial (1)	Crossover (2)	Other (S) (3)	**Multicenter**	Yes (1)	No (0)		
Pilot study	Yes (1)	No (0)	**Total sample size (N)**		**# Withdrawals**				
Population disease		**Primary outcome**	Selected (0)	Inferred (1)	Stated (2)				
Author's conclusions			**Significance**	Yes (1)	No (0)	Not reported (2)			
Positive (1)	Negative (0)	Neutral/unclear (2)	**If significant**	Favour control (0)	Favour Tx (1)	Unclear (2)	N/A (3)		
Type of primary outcome	Dichotomous (0)	Continuous (1)							