Title

Development and Validation of a N.O.V. (New, Original and Valid) Tool for Assessing the Quality of Individual Patient Data Meta-Analysis

Author

Arora, Rashi

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Development and Validation of a N.O.V. (New, Original and Valid) Tool for Assessing the Quality of Individual Patient Data Meta-Analysis

A thesis submitted in partial satisfaction

for the degree Master of Science

in Oral Biology

by

Rashi Arora

2013

ABSTRACT OF THE THESIS

Background.

Evidence based Dentistry and Patient Centered Care are the two paradigms which influence the process of clinical decision making in modern dental care. Although they are complementary means to improve quality, they may seem at odds with each other. In the last decade, a need has emerged to examine the intersection of these two paradigms. Meta-analyses are a hallmark of evidence based dentistry as they succeed in showing statistically significant results by combining the results from individual studies. However, they are limited when the individual studies are heterogeneous. Meta-regression may help investigate this heterogeneity, but its limited ability to identify which patient features are related to the size of treatment effect is answered by using an individual patient data approach. Poor reporting of individual patient data meta-analysis diminishes its value to clinicians, policy makers, and other users. Our goal is to develop and validate a tool to assess the quality of an individual patient data meta-analysis.

Methods.

We develop a tool based on the literature available in the field of Individual patient data meta-analysis. A modification of the PRISMA is done without altering its intent of use. In total, twelve Individual patient data meta-analyses in oral and maxillofacial medicine are identified to validate the tool and to establish its reliability. This is carried out by two independent readers who are standardized and trained to avoid

misinterpretation of the Tool items. A Pearson's r coefficient is calculated to test inter-rater; intra-rater reliability and criterion validity.

Results

A 9 item tool with a minimum of 3 criteria per item is constructed. The lowest score possible per question is 1 and the highest score possible per question is 4. The minimum total score possible is 9 and the maximum total score possible is 36. The new instrument has high content validity by virtue of its construction process. Testing the tool for criterion validity yields a Pearson's r of 0.957. The tool has construct validity based on the evidence of Test content, Internal structure, Response process, Relations to other variables, and Consequences of testing. It is also reliable with a Pearson's r of 0.927 for inter-rater reliability and a Pearson's r of 0.972 and 0.904 for intra-rater reliability.

Discussion

Our study led to the development of a novel instrument specifically designed for assessing the quality of individual patient data meta-analysis. The tool scores for the 12 papers quantified the degree to which each paper satisfied certain criteria that are established in the literature as determinant factors of highest quality meta-analysis in general, and individual patient data meta-analysis in particular. The tool has significant implications for clinicians and researchers. We recognize the need for further testing of the Tool to increase its validity. New evidence in areas of methodology of IPDMA may call for a need to update the Tool.

The thesis of Rashi Arora is approved.

Carl Maida

Neal Garrett

Francesco Chiappelli, Committee Chair

University of California, Los Angeles

2013

DEDICATION

I dedicate this work to Dr.Chiappelli for fostering me in this foreign land and without whose guidance none of this would be possible, my mother Saroj Arora for making me the person I am today and my husband, Ashvin Domadia for supporting my decisions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

ABBREVIATIONS

PCC	Patient Centered Care
EBD	Evidence Based Dentistry
ADA	American Dental Association
RCT	Randomized Controlled Trial
IPD	Individual Patient Data
IPDMA	Individual Patient Data Meta Analysis
AHRQ	Agency for Healthcare Research and Quality

ACKNOWLEDGEMENTS

I thank Dr. Garrett and Dr. Maida for their critical addition and feedback and mentorship on the thesis, Megan Scott for her encouragement from day one, Paulina Nguyen and Molly Uyeda for being the independent readers. Last but certainly not the least, I am grateful to my family.

CHAPTER ONE: INTRODUCTION/BACKGROUND

A. Patient Centered Care: What is it? Why is it important?

Modern dental care is influenced by two paradigms: 'evidence-based dentistry' and 'patient-centered care'. In the last decade, both paradigms have rapidly gained in popularity and are now both supposed to affect the process of clinical decision making during the daily practice of dentists.

Patient-Centered Care (PCC), although not a new phenomenon, has recently attracted renewed attention. It challenges the traditional-model of the doctor-patient relationship which is generally asymmetric in terms of power - the doctor asks the patient questions and the doctor makes most of the treatment decisions [Freeman et al., 1999]. Health care professionals believe this traditional model lacks efficacy as patients do not necessarily follow doctor's orders, especially when it comes to the prevention and treatment of chronic diseases. Patients are now far more active players in their treatment decisions and are more likely to express their needs and opinions to their health care providers. The increased role of the patient has resulted in a more balanced relationship between doctor and patient. PCC basically has a humanistic, bio-psychosocial perspective, combining ethical values on 'the ideal physician', with psychotherapeutic theories on facilitating patients' disclosure of real worries, and negotiation theories on decision making. PCC has been described by Stewart et al as the interweaving of six components: exploration of "*both the disease and the illness experience, understanding the whole person, finding common ground, incorporating prevention and health*

promotion, enhancing the patient-doctor relationship, and being realistic." [Stewart et al., 1995]. It puts a strong focus on patient participation in clinical decision making by taking into account the patients' perspective, and tuning dental care to the patients' needs and preferences. Research has shown that patient-centered clinical methods benefits factors such as patient satisfaction and adherence to treatment, doctor satisfaction and health outcomes [Stewart et al., 2000]. Patient-centered clinical approaches may help dentists interact with their patients, especially those with different social or cultural backgrounds. This approach can also improve patients' adherence to treatments and help to improve health-related behaviors such as oral hygiene and nutrition.

B. Evidence Based Dentistry: What is it?

Evidence-based dentistry is a rather young concept that entered the scientific literature in the early 1990s. It has basically a positivistic, biomedical perspective. Its focus is on offering clinicians the best available evidence about the most adequate treatment for their patients. The American Dental Association (ADA) defines Evidence-Based Dentistry as follows:

"Evidence-based dentistry (EBD) is an approach to oral healthcare that requires the judicious integration of systematic assessments of clinically relevant scientific evidence, relating to the patient's oral and medical condition and history, with the dentist's clinical expertise and the patient's treatment needs and preferences." [American Dental Association's Center for Evidence-Based Dentistry. "http://www.ada.org/1754.aspx" Accessed February 20, 2013.]

EBD consists of two principal and interdependent elements[Chiappelli et al., 2010]:

- Identification of the best available research evidence
- Integration of the best available evidence into treatment intervention

Identification of the best available evidence involves a step-by-step process consisting of:

- Formulating the patient-centered questions – i.e., the PICO question (patient-intervention-comparison-outcome)
- Searching for the appropriate evidence – i.e., the initial step of Research synthesis
- Critically appraising the evidence to yield Systematic reviews and Meta-Analysis

C. Relationship between Patient Centered Care and Evidence Based Dentistry.

Healthcare is a complex enterprise in practice and in theory. Some view it as a scientific endeavor while others see it as a service to individuals bounded in unique context [Sackett et al., 1998]. As it is a necessary system within our society, changes have been difficult to implement because of the interests of powerful organizations. However, for the past several decades, patient satisfaction has increasingly been recognized as an issue. Patient satisfaction has become a key quality indicator used by many hospitals [Walker et al., 2006] in an attempt to appear more patient-centered. In some ways, EBD and PCC are complementary means to improve quality; but it can also appear that, by virtue of their methods of changing dental practice, they are fundamentally at odds. Patient-centered outcomes are the key in this relationship and they demand more focus than current, if one desires to defy these odds. The current problem lies in the paucity of research on Patient-centered outcomes to support an EBD approach.

In general, the goal of EBD has been to improve quality through the standardization of medical care. EBD has typically been implemented through clinical guidelines, protocols, or best practices, all which are used to standardize (depending on the specificity of the research), not individualize, patient care. PCC on the other hand uses the foundations of narrative medicine to better understand the patient's story and integrate what is important to them into decisions about their oral health care. In essence, PCC aims to improve quality by individualizing, rather than standardizing, health care interactions. EBD guidelines are derived from population-based studies, while early teaching modules on PCC are based on general, or average, health beliefs among

subpopulations. Ultimately, for EBD and PCC to work together, we will need evidence that EBD can be patient-centered and that PCC can demonstrably improve health outcomes. For example, do patients feel that implementation of EBD, in practice, makes them feel that their perspective of the best treatment for them has been understood by the clinician? And does PCC lead to fewer errors and better health outcomes?

All of which is to say that we are in desperate need of a research agenda to examine the intersection between EBD and PCC. Fundamental questions at this intersection include, what evidence is needed to show that respect for individual preferences is worthwhile? And, how can the tools of EBD, such as guidelines, be adapted to foster the patient's sense of being respected and participating in their own health care decisions?

D. Meta-Analysis: A hallmark of Evidence Based Dentistry.

In many medical and dental specialties it is common to find that several trials have attempted to answer similar questions about clinical effectiveness. Often many of the individual trials will fail to show a statistically significant difference between the two treatments. However, when the results from the individual studies are combined using appropriate techniques (meta-analysis), significant benefits of treatment may be shown. The ADA defines it Meta-analysis as follows

"*Meta-analysis is a review that uses quantitative methods to combine the statistical measures from two or more studies and generates a weighted average of the effect of an intervention, degree of association between a risk factor and a disease, or accuracy of a diagnostic test.* " [American Dental Association Center for Evidence-Based Dentistry. "http://www.ada.org/1754.aspx" Accessed February 20, 2013].

Systematic review methodology is therefore at the heart of meta-analysis. The objective of systematic reviews is to present a balanced and impartial summary of the existing research, enabling decisions on effectiveness to be based on all relevant studies of adequate quality. Frequently, such systematic reviews provide a quantitative (statistical) estimate of net benefit aggregated over all the included studies. This approach offers a rational and helpful way of dealing with a number of practical difficulties that beset anyone trying to make sense of effectiveness research.

E. Disadvantages of Meta-Analysis.

A major concern about meta-analyses is the extent to which they mix studies that are different in kind (heterogeneity). One widely quoted definition of meta-analysis is: ‘a statistical analysis which combines or integrates the results of several independent clinical trials considered by the analyst to be “combinable”'[Huque et al., 1988]. The key difficulty lies in deciding which sets of studies are ‘combinable’. Clearly, to get a precise answer to a specific question, only studies that exactly match the question should be included. Most meta-analyses rely on randomized clinical trials (RCT) as the ‘gold standard’ of finding evidence for the most adequate treatments in oral health care [Ebrahim et al., 1997]. Without wanting to undermine the enormous relevance RCTs had and will have for the scientific development of medicine and dentistry, one major drawback must be considered: Results from RCTs are *not* generalized. Patients enroll in RCTs because they fulfill a very clear set of inclusion criteria, which are only based on the strictly defined diagnostic criteria of the disease under study. However, the majority of the patients have symptoms that do not fit exactly in the diagnostic criteria formulated by the researchers[Glasziou et al., 1995].

Randomized clinical trials are performed on homogeneous patient groups, that are constructed by banning many patients, while the consultation room is filled with patients that show a wide diversity in related symptom patterns and an even wider diversity in the way they evaluate and cope with these symptoms, sometimes for the better, sometimes for the worse. Studies can differ on the types of patient studied (disease severity or co-morbidity), the nature of local healthcare facilities, the intervention given and the primary

endpoint (death, disease, disability{all of which are not typical Patient-centered outcomes}). These systematic differences between studies can influence the amount of treatment benefit (the effect size) in a meta-analysis, leading to heterogeneity between studies.

Groups of patients which may seem homogeneous in public health terms can be very heterogeneous in their individual characteristics, necessitating the use of different interventions in different people [James et al., 1998].There is no such thing as an average patient who represents all others. Practice guidelines and recommendations often are created from research conducted with specific patient groups. The uniqueness of patients, their individual needs and preferences, and their emotional status are easily neglected as relevant factors in decision-making. Before applying these guidelines to the care of a particular patient, clinicians should ask how well the study sample represents that patient.

The presence or absence of heterogeneity influences the methods of the meta-analysis. If heterogeneity is within acceptable limits, then the analysis employs what is termed fixed-effects modeling. This assumes the size of treatment effect is the same (fixed) across all studies and the variation seen between studies is due only to chance. Random-effects models assume that the treatment effect really does vary between studies. Such models tend to increase the variance of the summary measure, making it more difficult to obtain significant results. When the amount of heterogeneity is large, it may even be inappropriate to calculate an overall summary measure of effect size.

When heterogeneity is detected, it is important to investigate what may have caused it. Meta-regression is a technique which allows researchers to explore which types

of patient-specific factors or study design factors contribute to the heterogeneity. The simplest type of meta-regression uses summary data from each trial, such as the average effect size, average disease severity at baseline, and average length of follow-up. This approach is valuable, but it has only limited ability to identify important factors. In particular, it struggles to identify which patient features are related to the size of treatment effect [Schmid et al., 2004].

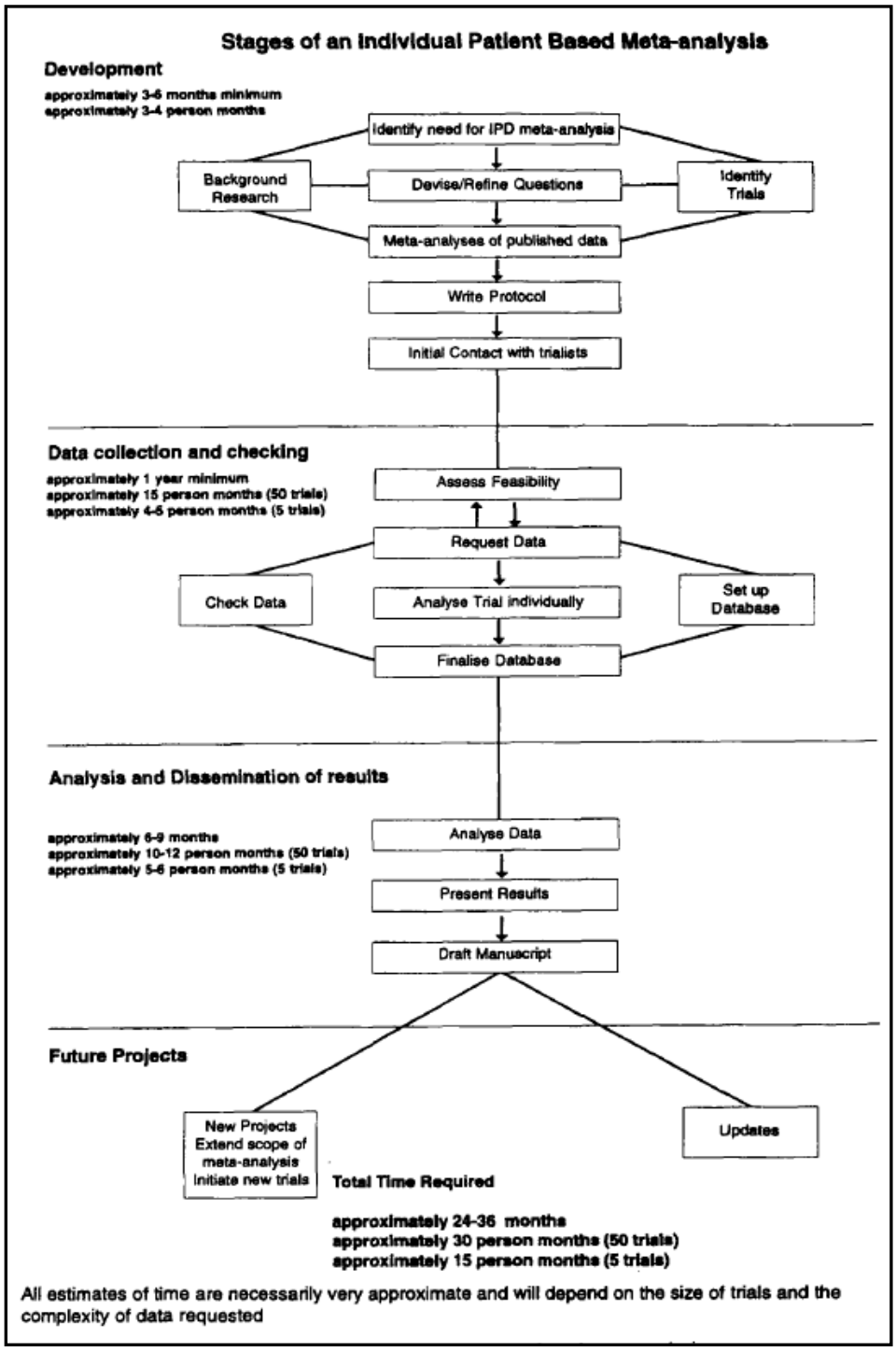
Another approach, using individual patient data, will give answers to the important question: what types of patients are most likely to benefit from this treatment?

F. Individual Patient Data Meta-Analysis.

Traditional methods for meta-analysis synthesize aggregate study level data obtained from study publications or study authors, such as a treatment effect estimate (for example, an odds ratio) and its associated uncertainty (for example, a standard error or confidence interval). An alternative but increasingly popular approach is to do a meta-analysis of individual patient data, in which the raw individual level data for each study are obtained and used for synthesis [Stewart et al., 1993].

Individual Patient Data Meta-Analysis(IPDMA) involve the central collection, validation, and reanalysis of "raw" data from all clinical trials worldwide that have addressed a common research question with data obtained from those responsible for the original trials [Stewart et al., 1995]. The overall philosophy is the same as for other types of well-designed and well-conducted systematic reviews. The methodology should differ only in terms of organizational structure, data collection, analysis, and the same basic methods should apply (Figure 1) [Stewart et al., 1995].

Figure 1 - Stages Of An Individual Patient Data Meta-Analysis.



Legend to Figure 1: Figure 1 shows the stages executing an Individual Patient Data Meta-Analysis.

The statistical implementation of an individual participant data meta-analysis crucially must preserve the clustering of patients within studies; it is inappropriate to simply analyze individual participant data as if they all came from a single study. Clusters can be retained during analysis by using a two step or a one step approach [Simmonds et al., 2005]. In the two step approach, the individual participant data are first analyzed in clusters in each separate study independently by using a statistical method appropriate for the type of data being analyzed. This step produces aggregate data for each study, such as a mean treatment effect estimate and its standard error. These data are then synthesized in the second step using a suitable model for meta-analysis of aggregate data. In the one step approach, the individual participant data from all studies are modeled simultaneously while accounting for the clustering of participants within studies. This approach again requires a model specific to the type of data being synthesized, alongside appropriate specification of the assumptions of the meta-analysis [Riley et al., 2010]. Detailed statistical articles regarding the implementation and merits of one step and two step individual participant data meta-analysis methods are available [Turner et al., 2000; Higgins et al., 2001; Whitehead et al., 2001; Tudur-Smith et al., 2005; Jones et al., 2009; Riley et al., 2008; Riley et al., 2008(1)].The two approaches have been shown to give very similar results, particularly when the meta-analysis aims to estimate a single treatment effect of interest[Mathew et al., 1999;Olkin et al.,1998;Riley et al., 2008].

G. Advantages of Individual Patient Data Meta-Analysis .

When carrying out an IPD meta-analysis, there are advantages to be gained both from the nature of the data itself and from the processes involved in reviewing evidence, as an international multidisciplinary team. There are undoubtedly limits of relying only on data presented in published reports. By definition, unpublished trials are not included, data may be inconsistent or incompatible across trials, and papers frequently present inadequate information. Aggregate data is more likely to be reported (and in greater detail) when statistically or clinically significant, amplifying the threat of publication bias and within study selective reporting. On the contrary, having individual participant data facilitates standardization of analyses across studies and direct derivation of the information desired, independent of significance or how it was reported. IPDMA also have a longer follow-up time, more participants, and more outcomes than were considered in the original study publication. This means that individual participant data meta-analyses are potentially more reliable than aggregate data meta-analyses, and the two approaches may lead to different conclusions. Therefore systematic reviews should look beyond the data presented in publications which could be accomplished by collecting IPD.

H. Reporting and Quality of Meta-Analysis.

It is now evident that IPDMA are essential to summarize evidence relating to efficacy and safety of health care interventions accurately and reliably. The clarity and transparency of these reports, however, is not optimal. Poor reporting of IPDMA diminishes its value to clinicians, policy makers, and other users. Currently the PRISMA statement is the standard for investigators when reporting their findings and also provides a benchmark by which meta-analyses may be appraised. Reporting guidelines such as the PRISMA (See Appendix A- The PRISMA Checklist) when considered by the researchers at the outset of their work, are likely to yield better designed studies that will be easier to understand when the work is published. It ensures the clarity and transparency of reporting of systematic reviews and meta analysis in general but it is not specific to the individual participant data approach. A review of 33 applied individual participant data meta-analyses from between 1999 and 2001 noted that "clear reporting of the statistical methods used was rare" and that only a few studies actually referred to a protocol for their individual participant data project [Simmonds et al., 2005]. Clearly these shortcomings must be addressed.

I. Addressing the Research Question.

How does one assess the reporting and in turn the quality of an Individual Patient data meta-analysis?

Aim#1: Developing a tool to assess the quality of an IPDMA.

Aim#2: Validation of the developed tool.

Aim #3: Establishing the reliability of the developed tool.

CHAPTER TWO: METHODS

A. Development of the Tool

The tool was built upon a framework of "domains" and "elements" defined by the AHRQ as deemed appropriate for systematic reviews and meta analysis[West et al., 2002]. A “domain” of study methodology reflects factors to be considered in assessing the extent to which the study’s results are reliable or valid (i.e., study quality). Each domain has specific “elements” that one might use in determining whether a particular instrument assessed that domain; in some cases, only one element defines a domain. The domains are study question, search strategy, inclusion and exclusion criteria, interventions, outcomes, data extraction, study quality and validity, data synthesis and analysis, results, discussion, and funding or sponsorship. Table 1 shows the domains and elements for systematic reviews and meta analysis defined by the AHRQ [West et al., 2002].

A checklist was then formed from the literature available in the field of IPDMA to make the tool specific to the field. The checklist's content was generated from a review of the literature[Stewart et al., 2008; Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration, 2008]. The inclusion criteria used to select the checklist's content was that they should measure “scientific quality” and they should be applicable to IPDMA. The initial tool was developed by designing items that corresponded to the checklist. The content of the individual items were guided by this checklist. At least three criteria were constructed per item and it was graded on a point-based system. The items were refined

through an iterative process of discussions and revision. Once the tool was revised, the reviewer team, consisting of two independent readers reviewed the resulting draft to determine whether the remaining items were ambiguous or vague, or had awkward wording or biased language. Items were discussed until consensus was reached on additional revision of items.

TABLE 1 - DOMAINS AND ELEMENTS FOR SYSTEMATIC REVIEWS AND META-ANALYSIS BY AHRQ

Domain	Elements
Study Question	• **Question clearly specified and appropriate**
Search Strategy	• ***Sufficiently comprehensive and rigorous with attention to possible publication biases***
	• *Search restrictions justified (e.g., language or country of origin)*
	• Documentation of search terms and databases used
	• Sufficiently detailed to reproduce study
Inclusion and Exclusion Criteria	• **Selection methods specified and appropriate, with *a priori* criteria specified if possible**
Interventions	• **Intervention(s) clearly detailed for all study groups**
Outcomes	• **All potentially important harms and benefits considered**
Data Extraction†	• Rigor and consistency of process
	• Number and types of reviewers
	• Blinding of reviewers
	• Measure of agreement or reproducibility
	• Extraction of clearly defined interventions/exposures and outcomes for all relevant subjects and subgroups
Study Quality and Validity	• ***Assessment method specified and appropriate***
	• Method of incorporation specified and appropriate
Data Synthesis and Analysis	• ***Appropriate use of qualitative and/or quantitative synthesis, with consideration of the robustness of results and heterogeneity issues***
	• Presentation of key primary study elements sufficient for critical appraisal and replication
Results	• **Narrative summary and/or quantitative summary statistic and measure of precision, as appropriate**
Discussion	• **Conclusions supported by results with possible biases and limitations taken into consideration**
Funding or Sponsorship	• ***Type and sources of support for study***

B. Modification of the PRISMA

There have been parallel developments in meta-analysis. As major undertakings of work, their results may be influential to health care providers, researchers, and decision makers. Thus, the need for a consistent framework of reporting was recognized. This led to the compilation of the Quality of Reporting of Meta-analyses (QUOROM) statement [Moher et al., 1999], which was aimed at improving the quality of published meta-analyses of randomized controlled trials. Recently, the QUOROM statement [Moher et al., 1999] has been superseded by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [Liberati et al., 2009]. This was in response to developments in systematic review methodology and to widen the scope beyond randomized controlled trials. Also, the PRISMA statement is built upon the framework of "domains" and "elements defined by the AHRQ as deemed appropriate for systematic reviews and meta-analyses (Refer to Table 1, page 17 for the table).Currently this statement is the standard for investigators when reporting their findings and also provides a benchmark by which meta-analyses may be appraised. The PRISMA checklist was thus selected to be the most- up-to-date instrument for rating the quality of IPDMA.

The PRISMA Statement consists of a 27-item checklist and a four-phase flow diagram. Each checklist item is attached with a brief description and a box for the Page # it is reported on. Previous research has outlined that critical appraisal may have aspects of objective and subjective assessment that cannot be reduced to a simple check list [Crowe and Sheppard, 2011], therefore, we modified the PRISMA in the following manner: If the

item was found in the literature and reported, it would yield two points, and if the item was not found and not reported it would yield one point. It is a quantitative nominal scale of "1" and "2". This format was true for all of the 27 items. When the point value for each item had been scored, they can be summarized. The score ranges from 27-54. This does not change the intent of its use. It simply deemphasizes the importance of any one item even though there is recognition fit, and it helps to lay out the criteria to establish criterion validity.

C. Methodology

We appropriately formed a P.I.C.O. question , which suggested subject heading keywords for inclusion and exclusion purposes. The inclusion criteria were:

- IPDMA
- Interventions/ treatments addressing the diseases in the oral and maxillofacial region
- Any population

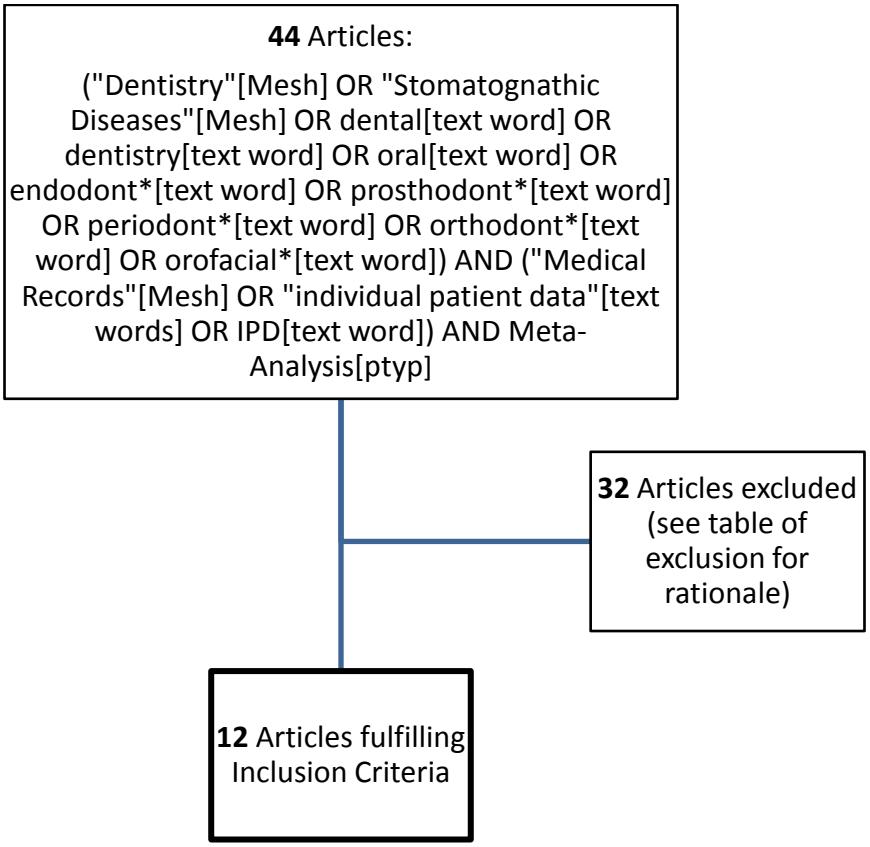
The exclusion criteria were:

- Aggregate data meta-analysis
- Narrative reviews or editorial letters to editors
- In vitro or animal studies
- Non- English articles

We searched the National Library or Medicine (Pubmed) with the help of a professional biomedical librarian, Rikke Ogawa. The search string used was: " ("Dentistry"[Mesh] OR "Stomatognathic Diseases"[Mesh] OR dental[text word] OR dentistry[text word] OR oral[text word] OR endodont*[text word] OR prosthodont*[text word] OR periodont*[text word] OR orthodont*[text word] OR orofacial*[text word]) AND ("Medical Records"[Mesh] OR "individual patient data"[text words] OR IPD[text word]) AND Meta-Analysis[ptyp]" This yielded total of 44 articles. Upon examination of the 44 articles, 32 articles were excluded (See Table 2 - Excluded Studies and Reason for Exclusion) and the remaining 12 articles were identified to meet the inclusion criteria. (See Figure 2 - Search Strategy). They are :

Roberts et al., 1991;Edwards et al., 2002;Clauser et al.,2003; Baujat et al., 2006a;Baujat et al., 2006b;McDaid et al., 2009;Steiner et al.,2009;Brin et al.,2009;Tandon et al., 2010;Moore et al.,2011;Blanchard et al.,2011;Chambrone et al.,2012 [Refer to Appendix D: Included Studies with their Abstracts].

FIGURE 2 - SEARCH STRATEGY



Legend to Figure 2.Figure 2 shows the search strategy used to find articles. The initial search yielded 44 articles. Out of these 12 articles fulfilled the Inclusion criteria.

TABLE 2 - EXCLUDED STUDIES AND REASON FOR EXCLUSION (Sorted by Date Published)		
	Study	Reason for exclusion
1.	Henegan et al., 2012	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of self-monitoring of oral coagulants.
2.	Mason et al., 2012	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of Acamprosate for alcohol dependence.
3.	Kovalchik et al., 2012	Is not based on intervention/treatment for a condition of the maxillofacial region, deals with the relation between mother's CD4+ counts and late postnatal HIV-free survival of breastfed children.
4.	Lehert et al., 2011	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of Racecadotril for childhood gastroenteritis.
5.	Hurwitz et al., 2011	Not an IPDMA
6.	Cassidy et al., 2011	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of capecitabine versus 5-fluorouracil in colorectal and gastric cancers
7.	Home et al., 2010	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the risk of

		hypoglycemia in people with type 2 diabetes using NPH insulin
8.	Fonseca et al., 2010	Not an IPDMA
9.	Sieber et al., 2010	Is not based on intervention/treatment for a condition of the maxillofacial region, compares the effectiveness of various interventions for pollen allergens.
10.	Lee et al., 2010	Is not based on intervention/treatment for a condition of the maxillofacial region, investigates the validity of cross-trial comparisons for competing treatments in advanced breast cancer
11.	Okines et al., 2009	Is not based on intervention/treatment for a condition of the maxillofacial region, compares interventions for advanced oesophago-gastric cancer
12.	De Backer et al.,2012	Update of the previous article [De Backer et al.,2008]
13.	Perera et al., 2008	Not an IPDMA, only an IPDMA protocol
14.	De Backer et al.,2008	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of Naftidrofuryl for intermittent claudication
15.	Harris et al.,2008	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the risk of Ibandronate and clinical fractures in women with postmenopausal osteoporosis

16.	Gerlinger et al., 2007	Is not based on intervention/treatment for a condition of the maxillofacial region, Evaluates menstrual bleeding patterns
17.	Mahr et al., 2007	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of methotrexate for treatment of giant cell Arteritis
18.	Lampl et al., 2007	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy and s9afety of effervescent As.10.pirin in migraine headaches.
19.	Sakamoto et al.,2007	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of an adjuvant therapy with uracil-tegafur for curatively resected rectal cancer
20.	Sakamoto et al.,2005	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of an adjuvant therapy with carmofur for curatively resected rectal cancer
21.	Pignon et al., 2005	Article in French
22.	Hashiguchi et al.,2004	Is not based on intervention/treatment for a condition of the maxillofacial region, Compares interventions for one-month effectiveness and safety after elective coronary stenting

23.	Dávalos et al., 2002	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of citicoline in acute ischemic stroke
24.	Sakamoto et al.,2001	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of adjuvant therapy with carmofur for curatively resected colorectal cancer
25.	Collins et al., 2001	Not an IPDMA
26.	Pan et al., 2000	Not an IPDMA
27.	Sakamoto et al.,1999	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of fluoropyrimidines for curatively resected colorectal cancer
28.	Groves et al., 1998	Is not based on intervention/treatment for a condition of the maxillofacial region, Compares drugs for treatment of spasticity
29.	The Atrial Fibrillation Investigators., 1997	Is not based on intervention/treatment for a condition of the maxillofacial region, Assesses the efficacy of aspirin in patients with atrial fibrillation
30.	Steinberg et al., 1997	Compares a meta-analysis of summary data vs. a meta-analysis of IPD for Ovarian cancer studies
31.	Moore et al., 1997	Compares the efficacy of various analgesics
32.	Pawinski et al., 1996	Not an IPDMA

Two independent readers(Rater 2 and Rater 3) were trained and standardized by Rater 1 as described in our previous paper [Chiappelli et al.,2012; Supplementary material: Table 1: Protocol to establish the reliability of research quality assessment tools]. This was done to eliminate any inconsistencies among the readers and to prevent any misinterpretations of the Tool items. Rater 1 was also responsible for the verification of the precision of reading for Rater 2 and Rater 3. Rater 2 and Rater 3 made notes of the scores on the article for a particular item and the location of the criteria being met. They later transferred theses scores to an Excel worksheet. Rater 1 ensured that the scores were transferred correctly without errors. A preliminary trial of the Tool scoring was performed to ensure that each of the two readers read the literature critically and consistently. The trial was run with the modified PRISMA and for the entire body of literature. Another test trial was conducted for the "Introduction" portion(Item 1) of the Tool for the entire body of literature. Any discrepancies in the scores were discussed by a third reader until a consensus was attained for the manner in which they followed the Tool's scoring criteria. After the two readers had been trained and their judgments were standardized, the reading and scoring of the articles were done independently among all of the readers. All readers were blind from one another’s scoring. After an interval of two weeks, the two readers assessed six arbitrarily assigned articles using the tool in order to establish intra-rater reliability.

D. Data Analysis

After collecting the scores, the two readers met to discuss disagreements. Consensus was achieved by discussion. An additional member of the research team (Rater 1) compiled the data, averaged the scores (to obtain the means of scores), and analyzed the scores of the readers so the analyses and interpretations of the data are unbiased. The following statistical tests were performed:

- Pearson r
- Mean
- SD

Pearson product-moment correlation coefficient r was performed . It is appropriate because it is a common measure of the correlation linear dependence between two variables X and Y, in this case between the two raters(for inter-rater reliability), or between the PRISMA and Tool scores(for criterion validity) giving a value between +1 and -1 inclusive. Prior to collecting data, an alpha level is predetermined. It is the probability of making a Type I error (false positives, occurs when you see things that are not there) is denoted by the Greek letter alpha (α). It is the likelihood of being incorrect when we say the relationship we found in our sample reflects a relationship in the population. Alpha can range from 0 to 1 where 0 means there is no chance of making a Type I error and 1 means it is unavoidable. Following Fisher, the critical level of alpha for determining whether a result can be judged statistically significant is conventionally set at .05. In order to determine if the r value we found with our sample meets that requirement, we will use a critical value table for Pearson's Correlation

Coefficient (See Appendix C: The critical value table for Pearson's Correlation Coefficient). The degree of freedom is N-2 where N is the total number of articles being assessed. The degree of freedom is 10 in this case; and the value a the intersection of alpha .05 and 10 degrees of freedom is 0.576. If the absolute value of the correlation coefficient is above .576(See Appendix C: The critical value table for Pearson's Correlation Coefficient), the null hypothesis (there is no relationship)is rejected and the alternative hypothesis is accepted: There is a statistically significant relationship between the two variables.

The mean describes the central location of the data set; the mean is the sum of the observations divided by the number of observations.

The standard deviation is a measure of variability or dispersion of a population, a data set, or a probability distribution. A low standard deviation indicates that the data points tend to be very close to the same value (the mean); while a high standard deviation indicates that the data are "spread out" over a large range of values.

RESULTS

A. The Developed Tool.

This is a 9-item tool developed to assess the quality of IPDMA. The lowest score possible per question is 1 and the highest score possible per question is 4. One may argue that by giving 1 point to any article even when it satisfies none of the criteria is an incorrect process, however, the absolute score of each article is unimportant, instead, it is the relative score of each article that matters. The box is provided at end of the criteria for the reader to check if the criteria is met. With a total of 9 questions, the minimum total score possible a IPDMA will receive is 9 and the maximum total score possible is 36. For qualitative purposes, one may make the arbitrary assignments to the IPDMA with an overall score of:

- 9-18 = poor
- 19-27 = average
- 28-36 = good

The Tool

1. Was a structured summary(abstract) provided for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The background & objectives is present in the abstract.	☐
(B) The search methods and data analysis done is present in the abstract.	☐
(C) The main results and author's conclusions are outlined in the abstract.	☐

Explanation: Abstracts provide key information that helps readers to understand the scope, processes, and findings of a IPDMA and to decide whether or not to read the full report. Sometimes, the abstract may be all that is readily available to a reader, for example, in a bibliographic database. Therefore it should present a balanced assessment of the IPDMA's findings that mirrors, briefly, the main text of the report. Structured abstracts provide readers with a series of headings pertaining to the purpose, methods, results, and conclusions being reported. They give readers more complete information and facilitate finding information more easily than unstructured abstracts.

Criteria (A) The background & objectives is present in the abstract. This sets the context for readers and explains the importance of the review question along with using elements of PICO to state the primary objective of the review.

Criteria (B) The search methods and data analysis done is present in the abstract. This summarizes the sources that were searched, the inclusion criteria, and the appraisal methods used to integrate or summarize the data.

Criteria (C) The main results and author's conclusions are outlined in the abstract. This provides numerical results with confidence intervals for the most

important outcomes. Ideally, the IPDMA should specify the amount of evidence in these analyses (numbers of studies and numbers of participants). The authors should provide clear and balanced conclusions that are closely linked to the objective and findings of the review.

2. Was a Background (Introduction) provided for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Background gives a context of what is already known from existing literature. ☐
(B) The Background gives a Rationale and reason for why the IPDMA approach was sought. ☐
(C) The Background outlines the aims and outcomes to be achieved with the IPDMA approach. ☐

Explanation: The Background/Introduction helps the reader to understand the rationale behind the study and what the IPDMA may add to what is already known.

Criteria (A) The Background gives a context of what is already known from existing literature. This helps define the importance of the review question from

different perspectives (e.g., public health, individual patient, or health policy) and gives an idea of the current knowledge and its limitations.

Criteria (B) The Background gives a Rationale and reason for why the IPDMA approach was sought. This discusses the extent to which the limitations of the existing evidence base may be overcome by the IPDMA approach.

Criteria (C) The Background outlines the aims and outcomes to be achieved with the IPDMA approach. The questions being addressed, and the rationale for them, are one of the most critical parts of a IPDMA and stating them precisely helps the reader to understand the scope and applicability of the study to their interest.

3. Did the Methods section provide a Protocol for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Methods section specifies the information regarding how and where the protocol can be accessed. ☐
(B) The Methods section specifies that the Protocol contains the methods and analyses to be used. ☐

(C) The Methods section specifies that the Protocol contains the outcomes & patient characteristics to be analyzed. ☐

Explanation: As with any formal research, some form of written plan or protocol should be produced for the meta-analysis. The highly collaborative nature of IPDMA, the time consumed in establishing this collaboration and the huge costs involved, further warrant a protocol for the process. Developing a written plan or protocol makes setting up a meta-analysis more rigorous by helping to identify problems and clarify issues early in the project.

Criteria (A) The Methods section specifies the information regarding how and where the protocol can be accessed. The ease of access to the protocol increases transparency in the study.

Criteria(B) The Methods section specifies that the Protocol contains the methods and analyses to be used. The methods and analyses provide a guidance to how reviewers will extract information , and methods that reviewers might use to quantitatively summarize the outcome data.

Criteria (C) The Methods section specifies that the Protocol contains the outcomes & patient characteristics to be analyzed. Having a protocol can help restrict the likelihood of biased post hoc decisions in review methods, such as selective outcome reporting.

4. Did the Methods section provide a Eligibility Criteria for considering studies for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Inclusion and Exclusion criteria specifies the types of studies (RCTs, language restrictions, publication status etc). ☐
(B) The Inclusion and Exclusion criteria specifies the types of participants in the studies (age group, gender etc). ☐
(C) The Inclusion and Exclusion criteria specifies the types of interventions being compared. ☐
(D) The Inclusion and Exclusion criteria defines the outcome measures for studies to be considered. ☐

Explanation: Knowledge of the eligibility criteria is essential in appraising the validity, applicability, and comprehensiveness of a IPDMA.

Criteria(A) The Inclusion and Exclusion criteria specifies the types of studies (RCTs, language restrictions, publication status, etc). The types of studies helps evaluate level of the evidence used in the meta-analysis.

Criteria(B) The Inclusion and Exclusion criteria specifies the types of participants in the studies (age group, gender etc). Providing information about the type of participants, their defining characteristics of interest (often disease), and possibly the setting of care considered helps the reader to evaluate the applicability of the IPDMA.

Criteria(C) The Inclusion and Exclusion criteria specifies the types of interventions being compared. The interventions (exposures) under consideration in the IPDMA need to be transparently reported.

Criteria(D) The Inclusion and Exclusion criteria defines the outcome measures for studies to be considered. The outcomes of the intervention being assessed, such as overall survival, disease free survival, or quality of life improvements, should be clearly defined as they are required to interpret the validity and generalizability of the systematic review's results.

5. Did the Methods section provide a Search Strategy for identification of studies for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The different databases searched and the search strategy for each database are specified.	☐
(B) The methods used to avoid publication bias are specified.	☐
(C) It is specified if the searches are regularly updated and the last date this was done.	☐
(D) The studies obtained are assessed independently by two or more authors to reach a consensus on their eligibility for inclusion in the IPDMA.	☐

Explanation: To ensure that all the relevant trials are included, a robust search strategy is important.

Criteria(A) The different databases searched and the search strategy for each database are specified. Retrieval from any single database, even by an experienced searcher, may be imperfect, which is why more than one database should be searched. A detailed search strategy allows interested readers to assess the comprehensiveness and completeness of the search, and to replicate it. Authors should also report who developed and conducted the search [Zhang et al., 2006].

Criteria(B) The methods used to avoid publication bias are specified. Most reviews are limited to randomized controlled trials (RCTs), although not all studies reported as such are in fact randomized. On further inquiry, it can transpire that “randomization” has been done by birth date, date of clinic visit, or by alternate allocation, all of which could be biased. It is widely appreciated that

trials with positive results are more likely to be published than those with negative results [Dickersin, 1990; Dickersin et a.,1992; Easterbrook et al.,1991; Stern et al., 1997]. Clearly, any meta-analysis that uses only data from published reports will then be at risk of publication bias. Therefore the methods used to avoid publication bias should be specified.

Criteria(C) It is specified if the searches are regularly updated and the last date this was done. The regular updating of the search ensures that the results are current.

Criteria(D) The studies obtained are assessed independently by two or more authors to reach a consensus on their eligibility for inclusion in the IPDMA. Efforts to enhance objectivity and avoid mistakes in study selection are important. Thus authors should report whether the studies obtained were assessed by two or more authors, who these people were, and, what was the process was for resolving disagreements. The use of at least two investigators may reduce the possibility of rejecting relevant reports [Edwards et al., 2002]. The benefit may be greatest for topics where selection or rejection of an article requires difficult judgments [Cooper et al.,1989]. For these topics, authors should ideally tell readers the level of inter-rater agreement.

6. Did the Methods section specify the Data Collection Strategy for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The number of authors approached and the proportion that provide the data should be mentioned along with the number of authors who did not provide the data, the reasons why, and the number of patients in the respective study is specified. ☐
(B) The authors who provide the data, if they give all of their data or only a proportion and in case of the latter, the reasons why information was omitted. ☐
(C) The reasons for excluding(or including) any patients who were originally excluded(or included) by the source authors. ☐
(D) The data collected is checked for validity, consistency & integrity of randomization and follow-up. ☐

Explanation: A well organized Data Collection Strategy is an important step to ensure a well executed IPDMA.

Criteria (A) The number of authors approached and the proportion that provide the data should be mentioned along with the number of authors who did not provide the data, the reasons why, and the number of patients in the respective study is specified. The number of authors approached and the proportion that

provide the data is important as the unavailable data may impact the results of the analysis. The proportion of unavailable data helps gauge the significance of this impact.

Criteria (B) The authors who provide the data, if they give all of their data or only a proportion and in case of the latter, the reasons why information was omitted. Partial data reporting by source authors, if related to treatment, could seriously bias the results, for example, if patients are excluded because they are unable to tolerate the allocated therapy or follow the treatment schedule. Thus it is important that reasons for providing a proportion of data are succinct.

Criteria (C) The reasons for excluding(or including) any patients who were originally excluded(or included) by the source authors. There can be good clinical reasons for excluding(or including) certain types of individuals. However, to be unbiased, any exclusions(or inclusions) should be prespecified and applied objectively and uniformly across trials. Ideally, their effect should be assessed by sensitivity analyses (including and excluding patients to determine whether it influences the estimated treatment effect).

Criteria (D) The data collected is checked for validity, consistency & integrity of randomization and follow-up. The reason for checking data is primarily to make sure that we represent data accurately, insure follow-up, and carry out an intention-to-treat analysis of all randomized patients. This ensures that we

have the most unbiased, up-to-date estimate of the effect of the intervention. It also gives a greater insight into the design and characteristics of a trial that can assist in interpretation of both the individual trial and meta-analysis results. Any missing data, obvious errors, inconsistencies between variables, or extreme values can be discussed with the trialist and corrected where necessary. Collecting IPD that include the time interval between the randomization and the event of interest enables time-to-event analyses to be conducted. These include, for example, time to recovery, time free of seizures, time to conception and time to death. Indeed, one of the main reasons that IPD meta-analyses have been so important is that time-to-event analysis of survival is vital in evaluating therapies.

7. Did the Results section provide the necessary category of results ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The description of characteristics of the included & excluded studies including the number of patients in each study is specified. ☐
(B) The results of the main analysis along with statistical details such as how the clustering of patients within studies was accounted for. ☐

(C) The assessment of risk of bias in the included studies is done by two or more authors based on the risk of bias tool. ☐

Explanation: Authors should in general report syntheses for all the outcome measures they set out to investigate (i.e., those described in the protocol; see Item 4) to allow readers to draw their own conclusions about the implications of the results. Readers should be made aware of any deviations from the planned analysis.

Criteria (A) The description of characteristics of the included & excluded studies including the number of patients in each study is specified. Initial descriptions of the characteristics of the evidence covered in the review may tell readers important things about the study populations and the design and conduct of studies. These descriptions can help in the examination of patterns across studies. They may also provide important information about applicability of evidence, suggest the likely effects of any major biases, and allow consideration, in a systematic manner, of multiple explanations for possible differences of findings across studies.

Criteria (B) The results of the main analysis along with statistical details such as how the clustering of patients within studies was accounted for. Most IPD meta-analyses to date have used a two-stage approach to analysis. In the first stage, each individual study is analysed in the same way, as set out in the meta-analysis protocol or analysis plan. In the second step, the results, or summary statistics, of each of these individual study analyses are combined to provide a pooled estimate

of effect in the same way as for a conventional systematic review [Simmonds et al., 2005]. More complex approaches using multilevel modelling have been described depending on the type of data. The IPD approach is the most practical way to carry out Time-To-Event analysis and Subgroup Analyses**.** If a Time-To-Event analysis is done, a Hazard Ratio must be calculated. A hazard ratio makes use of the time (from randomization) at which each individual event takes place and also uses information from patients who have not yet experienced the event (censored patients). Subgroup Analyses refer to analyses that investigate whether any observed effect of an intervention is consistent across well-defined groups of patients, for example do women gain a smaller or larger benefit from treatment than men. However, subgroup analyses based on IPD remain stratified by trial. Thus, for example, the effect of treatment compared to control is calculated for men, and the effect of treatment compared to control is calculated for women within each trial. The individual trial results for a particular subgroup can then be combined to give a pooled estimate of treatment effect. Such subgroup analyses can give valuable clinical insights. Sensitivity analyses may be done to explore the degree to which the main findings are affected by changes in its methods or in the data used from individual studies (e.g., study inclusion criteria, results of risk of bias assessment).

Criteria (C) The assessment of risk of bias in the included studies is done by two or more authors based on the risk of bias tool. Authors should present the results of any assessments of risk of bias across studies the risk of bias tool addresses the

domains of sequence generation, allocation concealment, blinding, incomplete outcome data addressed, free of selective reporting[Higgins et al.,2011].

8. Did the Discussion summarize the main findings ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The summary of main findings is presented in a brief and balanced manner.	☐
(B) The results are compared with the published results and if they are not comparable, the reasons why should be mentioned.	☐
(C) The limitations of the IPDMA due to unavailable data are specified and if it might impact the results obtained.	☐

Explanation: The Discussion should summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).

Criteria (A) The summary of main findings is presented in a brief and balanced manner. Authors should give a brief and balanced summary of the nature and findings of the review. Sometimes, outcomes for which little or no data were

found should be noted due to potential relevance for policy decisions and future research. Applicability of the review's findings, to different patients, settings, or target audiences, for example, should be mentioned.

Criteria (B) The results are compared with the published results and if they are not comparable, the reasons why should be mentioned. It is important to evaluate the findings of the IPDMA in context to the results of previous literature which may be based on aggregate data analysis. This confirms if the IPD approach was successful or not.

Criteria (C) The limitations of the IPDMA due to unavailable data are specified and if it might impact the results obtained. As with any other formal research, it is important for the researchers to identify the limitations of their study. In case of IPDMA's the main source of limitation is unavailability of data. If unavailability is related to the study results, for example if investigators are keen to supply data from studies with promising results but reluctant to provide data from those that were less encouraging, then ignoring the unavailable studies could bias the results of the IPD review. If only a limited number of studies are able to provide IPD for analysis, then the value of using the IPD approach is questionable.

9. Did the Conclusion provide a general interpretation of the results in context of other evidence?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The implications for clinical practice are specified.	☐
(B) The implications for future research are specified.	☐
(C) The sources of funding for the IPDMA and the role of funders is specified.	☐

Explanation: Authors should try to relate the results of the review to other evidence, as this helps readers to better interpret the results.

Criteria (A) The implications for clinical practice are specified. The relevance of the results of the IPDMA for clinical decision makers should be mentioned. Authors may discuss the results of their review in the context of existing evidence regarding other interventions.

Criteria (B) The implications for future research are specified. Gaps in the field of research being investigated are evident at the end of the results of an evidence generation process such as a IPDMA. These gaps must be identified as potential research recommendations.

Criteria (C) The sources of funding for the IPDMA and the role of funders is specified. Given the potential role of IPDMA's in decision making, we believe authors should be transparent about the funding and the role of funders, if any. Any level of funding or services provided to the team should be reported. Authors should also report whether the funder had any role in the conduct or reporting of the analysis. Beyond funding issues, authors should report any real or perceived conflicts of interest related to their role or the role of the funder in the reporting of the data.

B. Validity

Validity is often defined as the extent to which an instrument measures what it purports to measure[Streiner et al., 2008]. It is a matter of degree and is not a reflection of the instrument itself·, but rather a reflection of the instrument relative to its specific use [Nunnally, 1978]. This means that while we speak of the validity of a test or instrument, validity is not a property of the test itself. Instead, validity is the extent to which the interpretations of the results of a test are warranted, which depend on the test's intended use (i.e., measurement of the underlying construct). It is ‘‘the most fundamental consideration in developing and evaluating’’ a tool [American Educational Research Association et al., 1999, p. 9].

Validity requires that an instrument is reliable, but an instrument can be reliable without being valid. For example, a scale that is incorrectly calibrated may yield exactly the same, albeit inaccurate, weight values. Therefore without proper and thorough validity testing, it is irrelevant whether a tool has reliability [American Educational Research Association et al., 1999, pp. 9–11]. In other words, if all raters independently agree on a score a paper should receive (reliability) this is immaterial if the score does not accurately reflect what is being measured (validity). Therefore, validation of the proposed tool was required before reliability[Crowe et al., 2012] could be examined., There are three types of validity: content, criterion, and construct [Carmines & Zeller, 1979].

B.1.Content Validity

Face and content validity are closely linked concepts that describe whether a measure is assessing the relevant aspects for the purpose, and whether the domains covered are appropriate, important and sufficient[Higginson et al., 2007].This type of validity addresses how well the items developed to operationalize a construct provide an adequate and representative sample of all the items that might measure the construct of interest. Because there is no standardization and no statistical test to determine whether a measure adequately covers a content area or adequately represents a construct, content validity usually depends on the judgment of experts in the field. The new instrument already has high content validity by virtue of its construction process. This is illustrated by the Tool being based on the upon a framework of "domains" and "elements" defined by the AHRQ and simultaneously catering to the specifications of an IPDMA by designing items that are specific to it.

B.2.Criterion Validity

Criterion validity refers to whether the measure correlates with another instrument that measures similar aspects. Preferably, the other instrument is the ‘gold standard’, meaning it has been validated, and is widely used and accepted in the field. The reasons for using the PRISMA as the gold standard have already been discussed. A modification of the PRISMA was done in the following manner: If the item was found in the literature and reported, it would yield two points, and if the item was not found and not reported it would yield one point. It merely translated the classic response of a "Yes/No" to a binary scale of "1" and "2". When the point value for each item had been scored, they were summarized. This modification did not change the intent of its use. It simply deemphasized the importance of any one item even though there is recognition of it, and it helps to lay out the criteria to establish criterion validity.

Criterion validity provides evidence about how well scores on the new measure correlate with other measures of the same construct or very similar underlying constructs that theoretically should be related. For a new measure, the correlation with the gold standard is expected to be between 0.4–0.8 for it to have an acceptable criterion validity[Streiner et al., 2008].

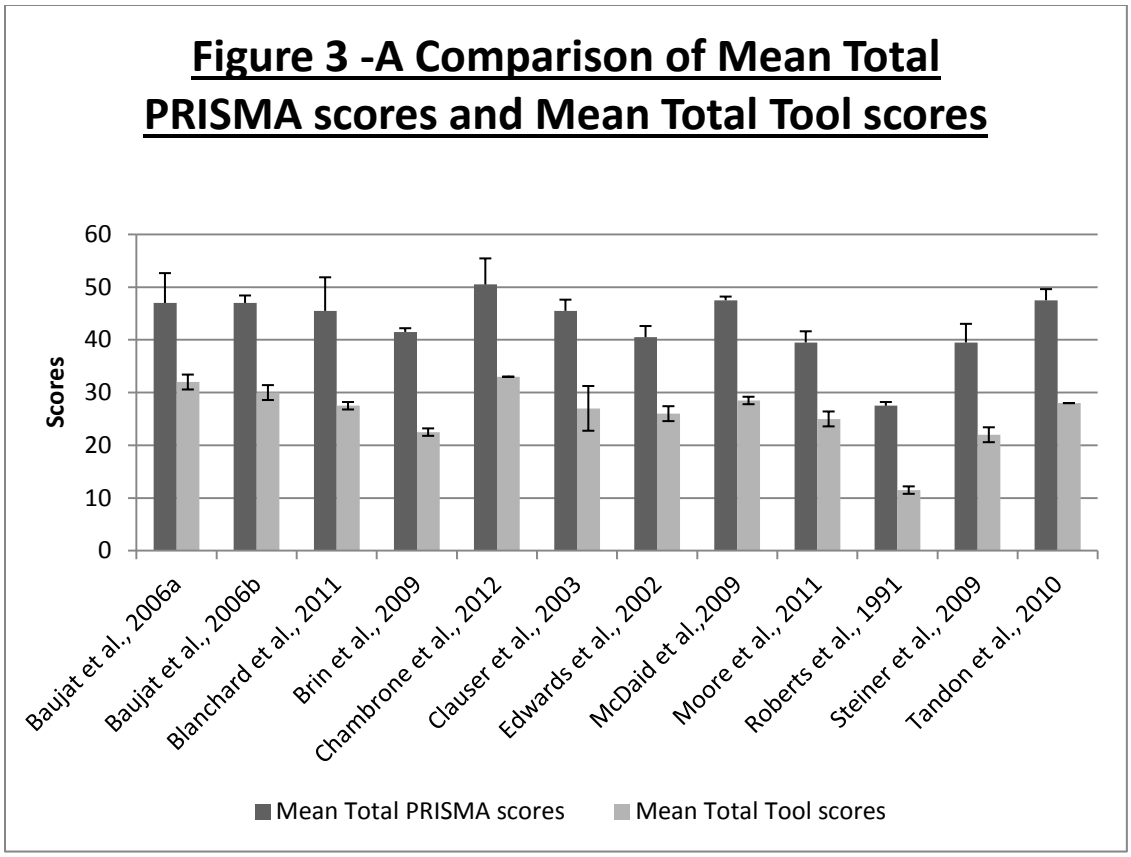
Table 3 shows the Total PRISMA scores for all the 12 articles by the two raters (2,3).

TABLE 3 -TOTAL PRISMA SCORES OF ALL 12 ARTICLES BY THE TWO RATERS		
Paper	**Rater 2**	**Rater 3**
Baujat et al., 2006a	51	43
Baujat et al., 2006b	48	46
Blanchard et al., 2011	50	41
Brin et al., 2009	42	41
Chambrone et al., 2012	54	47
Clauser et al., 2003	47	44
Edwards et al., 2002	42	39
McDaid et al.,2009	48	47
Moore et al., 2011	41	38
Roberts et al., 1991	28	27
Steiner et al., 2009	42	37
Tandon et al., 2010	49	46

Table 4 shows the mean scores for Total PRISMA and the mean scores for the Total Tool by the two raters, 2 and 3. At the bottom of the table is the Pearson product moment correlation coefficient r, 0.957. Since this value is above 0.576, the null hypothesis (there is no relationship)is rejected. Therefore there is a statistically significant relationship between the Mean Total PRISMA scores and the Mean Total Tool scores. Figure 3 compares the Mean Total PRISMA scores to the Mean Total Tool scores of the 12 articles. Figure 4 shows the criterion validity of the Tool as a function of the proportion of shared variance(r^2) of the mean total PRISMA scores and mean total Tool scores represented as a percentage.

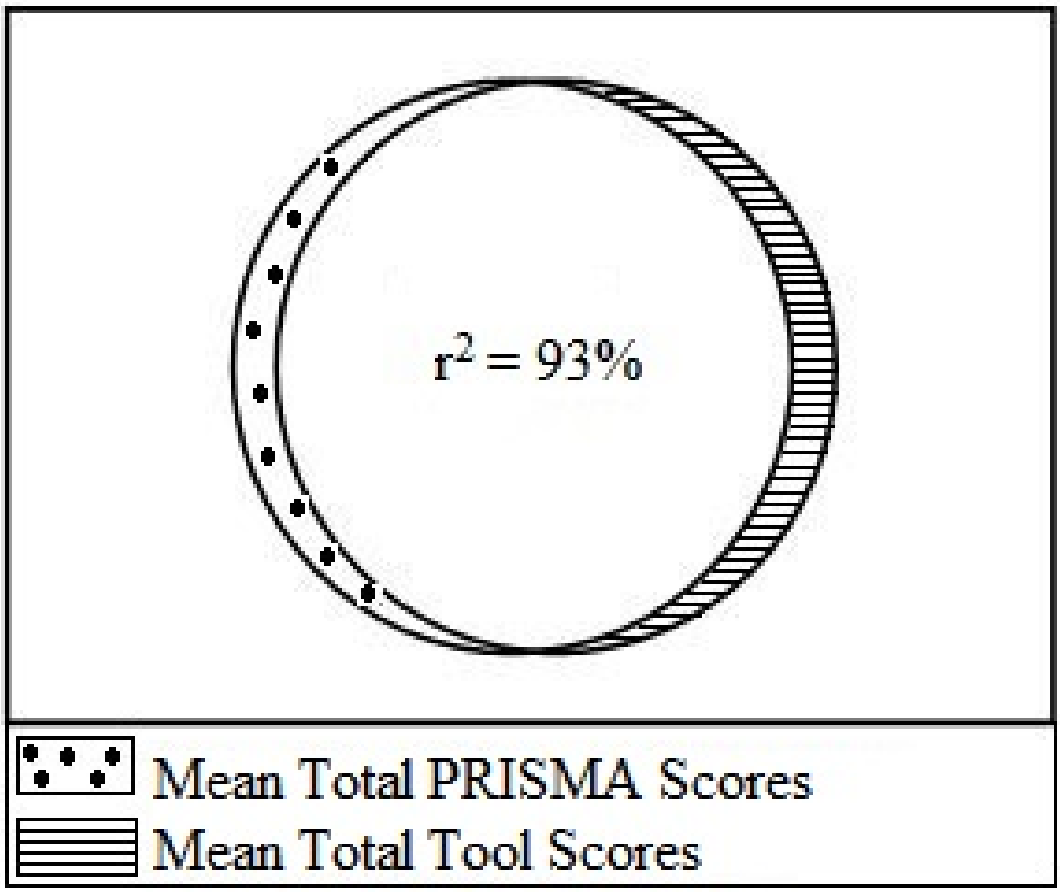
TABLE 4 - MEAN SCORES OF TOTAL PRISMA AND TOTAL TOOL

Paper	Mean Total PRISMA ($\bar{x}$ PRISMA)	Mean Total Tool ($\bar{x}$ TOOL)
Baujat et al., 2006a	47	32
Baujat et al., 2006b	47	30
Blanchard et al., 2011	45.5	27.5
Brin et al., 2009	41.5	22.5
Chambrone et al., 2012	50.5	33
Clauser et al., 2003	45.5	27
Edwards et al., 2002	40.5	26
McDaid et al.,2009	47.5	28.5
Moore et al., 2011	39.5	23.5
Roberts et al., 1991	27.5	11.5
Steiner et al., 2009	39.5	22
Tandon et al., 2010	47.5	28
Pearson's r	**0.965**	
r^2	0.930	
r2 (in %)	93	



Legend to Figure 3: Figure 3 compares the Mean Total PRISMA scores to the Mean Total Tool scores of the 12 articles. It also shows the Standard deviation for each score for each Article. Chambrone et al.,2012 scores high on the PRISMA and the Tool. Roberts et al.,1991 scores low on PRISMA and the Tool.

Figure 4- The Criterion Validity Of The Tool.



Legend to Figure 4:Figure 4 shows the Criterion Validity of the Tool. Mean Total PRISMA scores are represented by the dotted area and Mean Total Tool scores are represented by the lined area. The clear area is the proportion of shared variance(r^2) of the mean total PRISMA scores and mean total Tool scores represented as a percentage.

B.3.Construct Validity

Construct validity is the degree to which an instrument measures the construct for which it was intended . It is a judgment based on the accumulation of evidence from numerous studies using a specific measuring instrument. All evidence of validity, including content- and criterion-related validity, contributes to the evidence of construct validity. Evaluating construct validity is considered a mixture of reasoned argument, theoretical foundations, and empirical evidence which together support the credibility of score interpretation [Strauss et al, 2009]. To evaluate construct validity, five types of evidence are gathered: Test content, Internal structure, Response process, Relations to other variables, and Consequences of testing [American Educational Research Association et al., 1999, pp. 11–17].

Test content explores the specification of the construct, analysis of test content against the construct (e.g. themes, words, formats, questions, procedures, guidelines), The major threats to construct validity are construct underrepresentation and construct-irrelevant variance, either or both of which may be present within a test [American Educational Research Association et al., 1999, p. 10; Messick,1995]. Construct underrepresentation is when a test is too narrowly focused and fails to include important aspects of a construct. Construct-irrelevant variance is when a test is too broad and includes items that are not relevant to the construct being measured [American Educational Research Association et al., 1999, p. 10; Messick, 1995]. These threats were negated by building the tool on a framework of "domains" and "elements" defined by the

AHRQ as deemed appropriate for systematic reviews and meta analysis, and at the same time catering to the special needs of an IPDMA. Preliminary trial of the tool aided further refinement via consensus to reduce subjectivity. This helped ensure that application of the tool was consistent.

The internal structure of the tool was designed so that each item could be considered a one-dimensional construct and that the items did not overlap [Crowe and Sheppard, 2011]. Furthermore, each item is scored separately based on 2 principles. First, scoring was not simply a check list but allowed for a combined objective (tick boxes) and subjective scoring of each category based on the user guide. A scoring system with an objective and subjective component was chosen because previous research has outlined that critical appraisal may have aspects of objective and subjective assessment that cannot be reduced to a simple check list [Crowe and Sheppard, 2011]. Second, only items which are applicable to a research design are included in the appraiser's score. In other words, only items that are present and should be present, and items that are absent but should be present contribute to a category score. The recommendation for using the scores is to rank papers based on the total score The inferences which can be made are that the higher the total score, augmented by and including the category scores, implies the higher the credibility of the paper being appraised and the results obtained by that research.

Response process ensures that there is a fit between the processes used by a test taker to deliver a response and the construct being tested. It means giving a reader the ability to include where they found evidence for different aspects of the research and why

they thought this constituted evidence for or against giving a particular score given to the research. The tool developed aides the response process by the inclusion of tick boxes for each of the criteria for an item so that the reviewer can account for the elements of the construct that were present. The tick boxes also help a reviewer to keep track of their appraisal .

Relations to other variables means that scores should be tested against existing instruments, where the existing instrument have validity and reliability data available, and are reported to test similar or the same constructs as the new instrument. This encompasses evidence of criterion validity and has already been discussed.

Consequences of testing refers to the tool not over-stating its usefulness in appraising a paper and not being used outside the research methods it was designed to appraise. This has been achieved as the scope of this tool is limited to IPDMA and it would be inappropriate to use the scores obtained from it another types of articles (for example systematic review, RCT) because the contexts would be different. The user guide should be followed to ensure that the proposed tool is applied correctly.

C. Reliability of the Tool

Reliability is concerned with the extent to which a measurement is repeatable [Nunnally, 1978]. Reliability estimates are used to evaluate (1) the equivalence of sets of items from the same test (internal consistency) or of different observers scoring a behavior or event using the same instrument (inter-rater reliability).
(2) the stability of measures administered at different times to the same individuals or using the same standard (test–retest reliability or intra-rater reliability).

C.1. Inter-Rater Reliability

Inter-rater reliability (also called inter observer agreement) establishes the equivalence of ratings obtained with an instrument when used by different observers. Since the tool yields a continuous measurement in the form of a score, inter-rater reliability is determined by the correlation of the scores from two independent raters.

Table 5 shows the Scores for the Individual Items on the Tool by rater 2. Table 6 shows the scores for the individual Items in the Tool by rater 3. The horizontal rows indicate the score for each article. The vertical column indicates the Item on the tool.

TABLE 5 - SCORES FOR INDIVIDUAL ITEMS ON THE TOOL BY RATER 2										
Article	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**TOTAL**
Baujat et al., 2006a	4	4	2	4	4	1	4	4	4	31
Baujat et al., 2006b	2	4	4	3	4	3	3	4	2	29
Blanchard et al., 2011	3	4	4	1	2	3	3	4	4	28
Brin et al., 2009	3	4	1	4	2	2	3	3	1	23
Chambrone et al., 2012	4	4	4	4	4	2	4	4	3	33
Clauser et al., 2003	4	4	1	4	4	3	3	4	3	30
Edwards et al., 2002	4	4	1	4	3	3	2	3	3	27
McDaid et al.,2009	4	4	1	4	4	2	4	3	3	29
Moore et al., 2011	3	4	1	4	1	4	3	3	1	24
Roberts et al., 1991	3	1	1	1	1	1	1	1	1	11
Steiner et al., 2009	3	4	1	4	1	1	3	3	1	21
Tandon et al., 2010	3	4	1	4	4	2	4	4	2	28

Row 1 shows Baujat et al.,2006a scored a 4 on Item 1 of the Tool by the rater 2, 4 on Item 2 of the tool, and so forth. The score is totaled to the right and Baujat et al.,2006a scored 31 by the rater 2.

Row 2 shows Baujat et al.,2006b scored a 2 on Item 1 of the Tool by the rater 2, 4 on Item 2 of the tool, and so forth. The score is totaled to the right and Baujat et al.,2006b scored 29 by the rater 2.

The highest scoring IPDMA by rater 2 is Baujat et al.,2006a with 31 points, while, the lowest scoring article is Roberts et al.,1991 with 11 points.

When examining the columns, the first column, corresponds to question item 1 of the Tool - Abstract. Examining the scores in the column shows a range from 2 to 4 by rater 2.

The second column corresponds to question item 2 of the Tool - Background. Examining the scores in the column shows a range from 1 to 4 by rater 2.

TABLE 6 - SCORES FOR INDIVIDUAL ITEMS ON THE TOOL BY RATER 3

Article	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Baujat et al., 2006a	4	4	4	4	4	2	4	4	3	33
Baujat et al., 2006b	2	4	4	4	4	4	3	4	2	31
Blanchard et al., 2011	4	4	4	3	2	2	3	2	3	27
Brin et al., 2009	3	4	1	4	1	2	2	4	1	22
Chambrone et al., 2012	4	4	4	4	4	2	4	4	3	33
Clauser et al., 2003	4	4	1	4	1	2	1	4	3	24
Edwards et al., 2002	3	4	1	4	2	2	4	3	2	25
McDaid et al.,2009	4	2	1	4	3	4	4	3	3	28
Moore et al., 2011	3	3	1	4	2	2	2	3	3	23
Roberts et al., 1991	3	1	1	1	1	1	1	1	2	12
Steiner et al., 2009	3	4	4	3	1	1	3	3	1	23
Tandon et al., 2010	4	4	3	4	4	2	2	3	2	28

Row 1 shows Baujat et al.,2006a scored a 4 on Item 1 of the Tool by the rater 3, 4 on Item 2 of the tool, and so forth. The score is totaled to the right and Baujat et al.,2006a scored 33 by the rater 3.

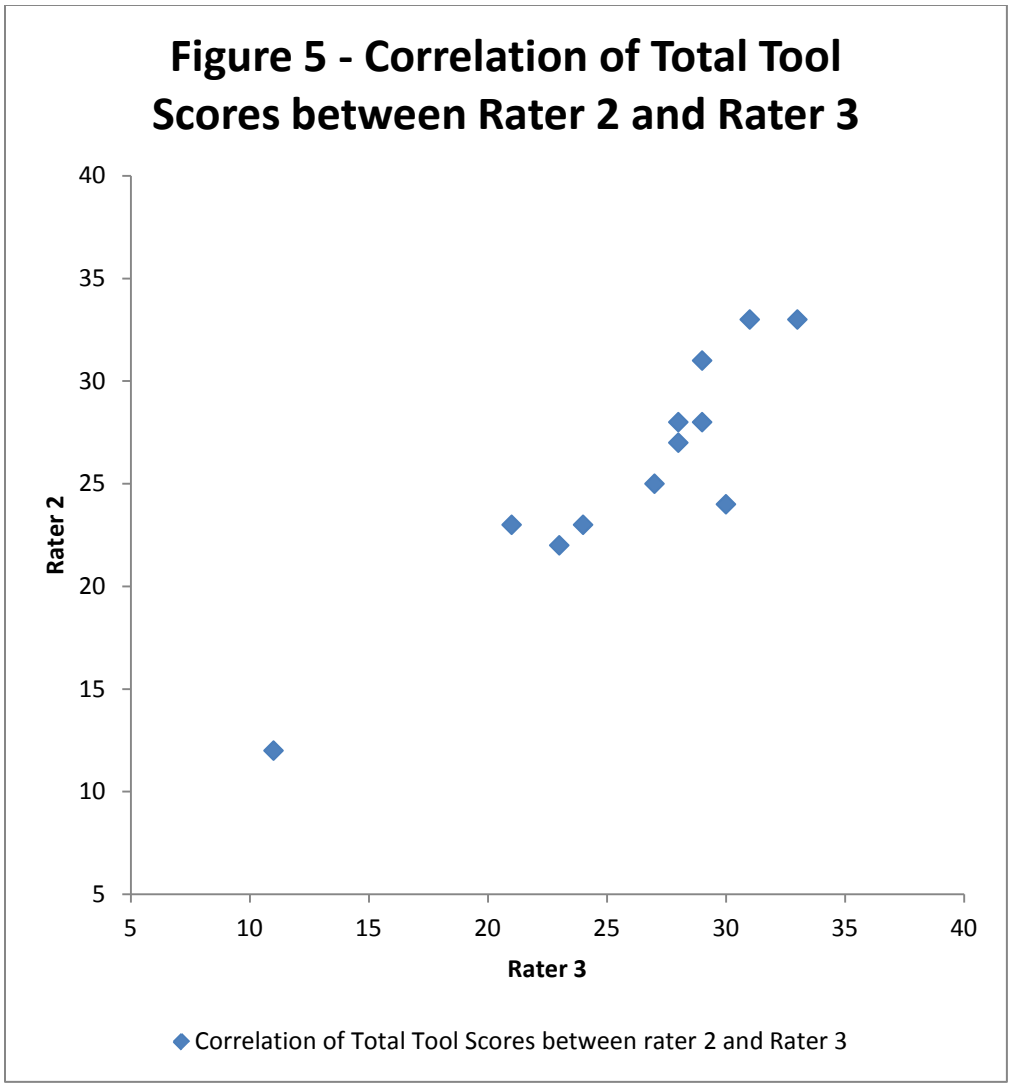
Row 2 shows Baujat et al.,2006b scored a 2 on Item 1 of the Tool by the rater 3, 4 on Item 2 of the tool, and so forth. The score is totaled to the right and Baujat et al.,2006b scored 31 by the rater 3.

The highest scoring IPDMA by rater 3 are Baujat et al.,2006a and Chambrone et al., 2012 with 33 points, while, the lowest scoring article is Roberts et al.,1991 with 12 points.

When examining the columns, the first column, corresponds to question item 1 of the Tool - Abstract. Examining the scores in the column shows a range from 2 to 4 by rater 3.

The second column corresponds to question item 2 of the Tool - Background. Examining the scores in the column shows a range from 1 to 4 by rater 3.

Figure 5 shows a correlation of the Total Tool scores by the 2 raters for all the twelve articles.



Legend to Figure 5: Figure 5 shows a correlation of the Total Tool scores by the 2 raters for all the twelve articles. The x-axis denotes the Total Tool scores by Rater 3. The y-axis denotes the Total Tool scores by Rater 2. The diamond represents the correlation of Total Tool score by the two raters for each paper.

Table 7 shows the overall total TOOL score of all the 12 articles by the two raters (2,3). At the bottom of the table is the Pearson product moment correlation coefficient r, 0.907.

Row 1 shows Blanchard et al.,2011 scored a 31 on the Tool by the rater 2 and a 33 by the rater 3. The mean score is 32.0 with a standard deviation of 1.4.

Row 2 shows Blanchard et al.,2011(1) scored a 29 on the Tool by the rater 2 and a 31 by the rater 3. The mean score is 30.0 with a standard deviation of 1.4.

The highest scoring IPDMA is Chambrone et al., 2012 with a mean total score of 33 points and a standard deviation of 1.4, while, the lowest scoring article is Roberts et al.,1991 with a mean total score of 11.5 points and a standard deviation of 0.7.

At the bottom of the table is the Pearson product moment correlation coefficient r, 0.920. Since this value is above 0.576, the null hypothesis (there is no relationship)is rejected. Therefore there is a statistically significant relationship between the Total Tool scores by Rater 2 and Rater 3.

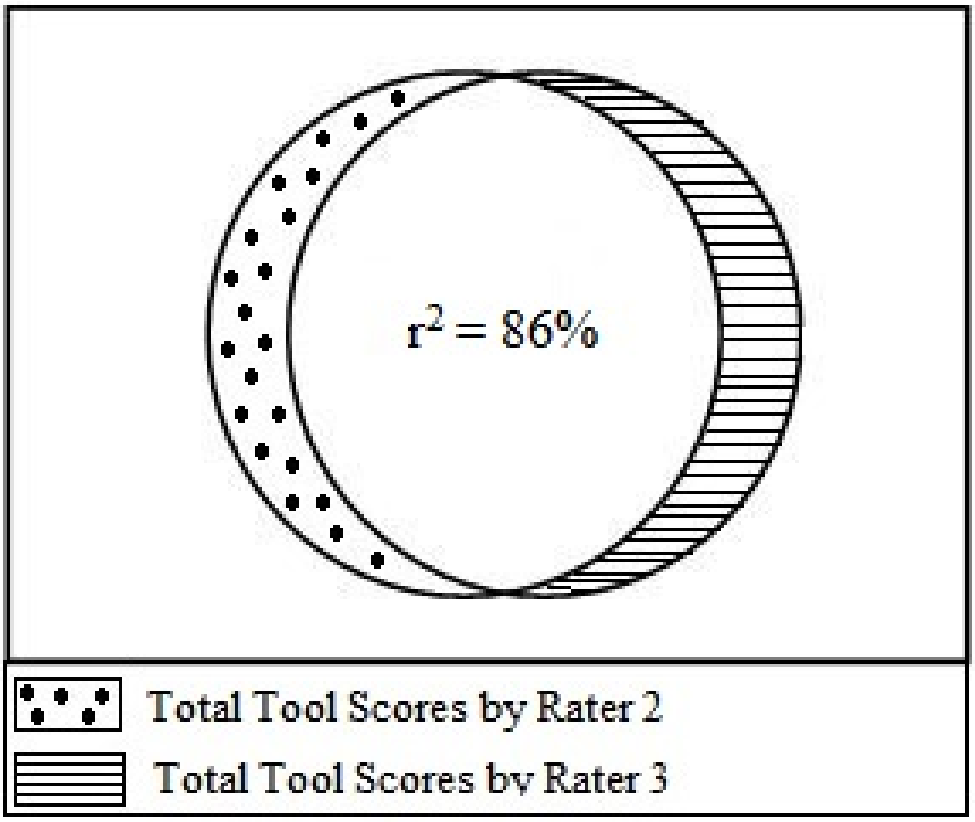
Figure 6 shows the inter-rater reliability as a function of the proportion of shared variance(r^2) of the Total Tool scores by Rater 2 and the Total Tool scores by Rater 3 represented as a percentage.

TABLE 7 - TOTAL TOOL SCORES OF ALL 12 ARTICLES BY THE TWO RATERS

Paper	Total Score by Rater2	Total Score by Rater3	Mean Total Score ($\overline{X}$ TOOL)	Standard Deviation
Baujat et al., 2006a	31	33	32.0	1.4
Baujat et al., 2006b	29	31	30.0	1.4
Blanchard et al., 2011	28	27	27.5	0.7
Brin et al., 2009	23	22	22.5	0.7
Chambrone et al., 2012	33	33	33.0	0.0
Clauser et al., 2003	30	24	27.0	4.2
Edwards et al., 2002	27	25	26.0	1.4
McDaid et al.,2009	25	28	26.5	2.1
Moore et al., 2011	24	23	23.5	0.7
Roberts et al., 1991	11	12	11.5	0.7
Steiner et al., 2009	21	23	22.0	1.4
Tandon et al., 2010	28	28	28.0	0.0

Pearson's r	**0.927**
r^2	0.858
r^2 (in %)	86

Figure 6 - The Inter-Rater Reliability Of The Tool



Legend to Figure 6: Figure 6 shows the Inter-rater reliability of the Tool. Total Tool scores by rater 2 are represented by the dotted area and the Total Tool scores by Rater 3 are represented by the lined area. The clear area is the proportion of shared variance(r^2) of the Total Tool scores by Rater 2 and the Total Tool scores by Rater 3 represented as a percentage.

C.2. Intra-Rater Reliability

Intra-rater reliability or stability of measurement, or test–retest reliability, is determined by administering a test at two different points in time to the same individuals and determining the correlation or strength of association of the two sets of scores.

Table 8 shows the Scores for the Individual Items on the Tool by rater 2 after an interval of 2 weeks(T_1) from the initial assessment (T_0). Table 9 shows the scores for the individual Items in the Tool by rater 3 after an interval of 2 weeks(T_1) from the initial assessment (T_0). The horizontal rows indicate the score for each article. The vertical column indicates the Item on the tool.

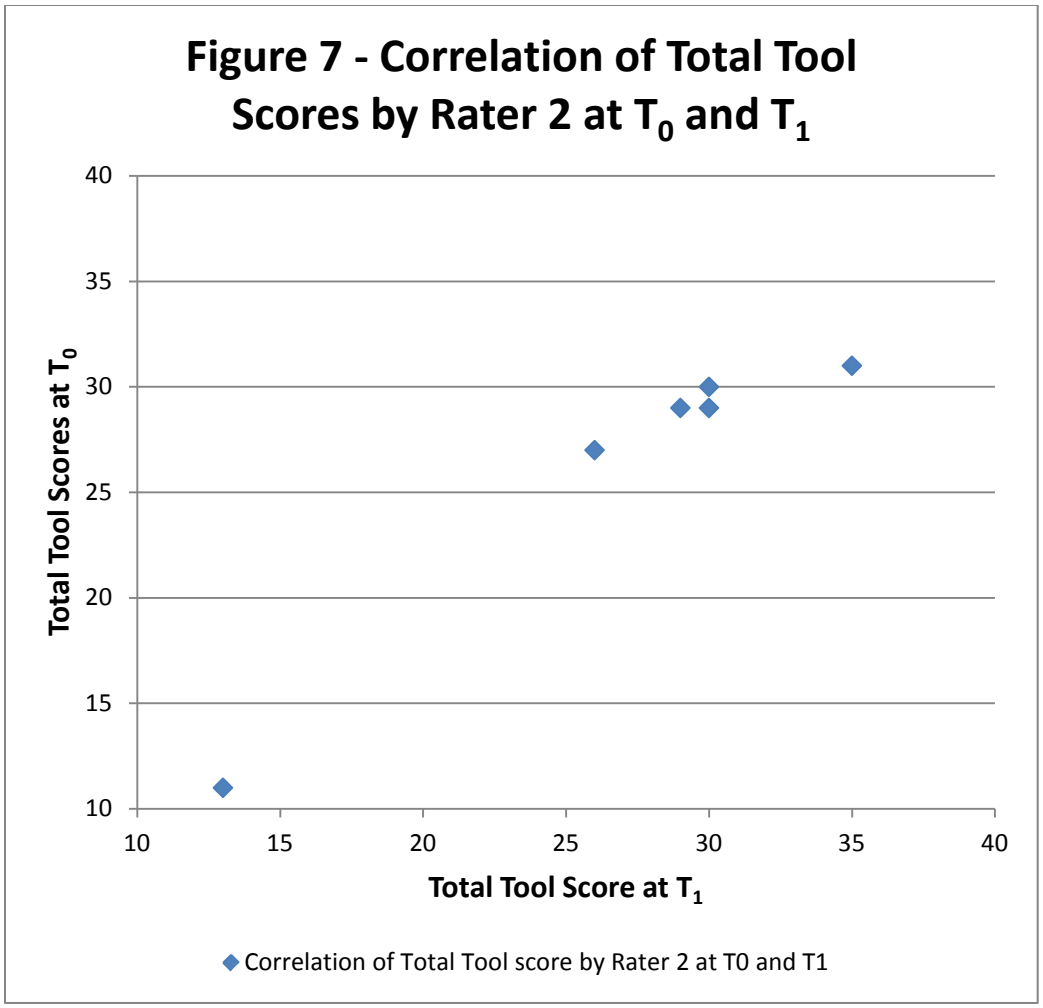
TABLE 8 - SCORES FOR INDIVIDUAL ITEMS ON THE TOOL BY RATER 2 AFTER AN INTERVAL OF 2 WEEKS(T_1)										
Article	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**TOTAL**
Baujat et al., 2006a	4	4	3	4	4	4	4	4	4	35
Baujat et al., 2006b	4	4	3	3	4	2	3	4	2	29
Clauser et al., 2003	4	4	1	4	4	3	3	4	3	30
Edwards et al., 2002	4	4	1	4	3	2	3	3	2	26
McDaid et al.,2009	4	4	1	4	4	2	4	4	3	30
Roberts et al., 1991	3	1	1	1	1	1	1	1	3	13

TABLE 9 - SCORES FOR INDIVIDUAL ITEMS ON THE TOOL BY RATER 3 AFTER AN INTERVAL OF 2 WEEKS(T1)										
Article	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Blanchard et al., 2011	3	4	4	4	4	2	2	2	4	29
Brin et al., 2009	4	4	1	4	1	3	3	3	2	25
Chambrone et al., 2012	4	4	4	4	4	3	4	4	3	34
Moore et al., 2011	1	4	1	4	2	2	2	3	3	22
Steiner et al., 2009	4	4	1	4	1	3	2	3	2	24
Tandon et al., 2010	4	2	1	4	4	3	3	3	2	26

Table 10 shows the Overall Total Scores for the Tool by the rater 2 at T_0 and T_1. At the bottom of the table is the Pearson product moment correlation coefficient r. Pearson's coefficient r for Rater 2 is 0.972. . The degree of freedom changes to 4 in this case, and the value at the intersection of alpha .05 and 4 degrees of freedom is 0.811. Since the value of the correlation coefficient is above 0.811, the null hypothesis (there is no relationship)is rejected. There is a statistical significance between the scores of Rater 2 at the two intervals. Figure 7 depicts the correlation of the Total Tool scores by Rater 2 at T_0 and T_1.

TABLE 10 - TOTAL SCORES FOR THE TOOL BY RATER 2 AT T_0 AND T_1

Article	**T_0**	**T_1**
Baujat et al., 2006a	31	35
Baujat et al., 2006b	29	29
Clauser et al., 2003	30	30
Edwards et al., 2002	27	26
McDaid et al.,2009	29	30
Roberts et al., 1991	11	13
Pearson's r	**0.972**	

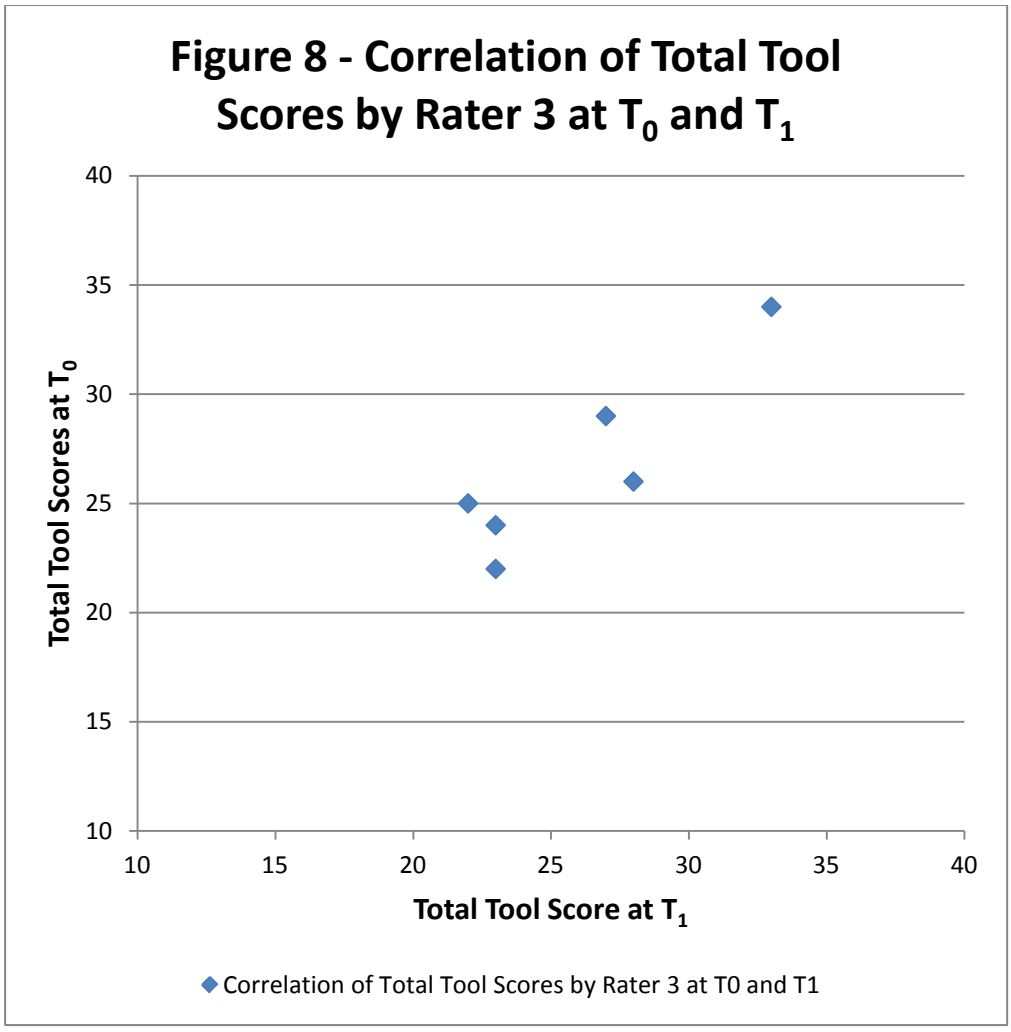


Legend to Figure 7: Figure 7 shows a correlation of the Total Tool scores by Rater 2 at T_0(initial) and T_1(after 2 weeks). The x-axis denotes the Total Tool scores by Rater 2 at T_1. The y-axis denotes the Total Tool scores by Rater 2 at T_0. The diamond represents the correlation of the Total Tool scores by Rater 2 at T_0 and T_1.

Table 11 shows the Total Scores for the Tool by the rater 3 at T_0 and T_1. At the bottom of the table is the Pearson product moment correlation coefficient r. Pearson's coefficient r for 3 is 0.904. The degree of freedom changes to 4 in this case, and the value at the intersection of alpha .05 and 4 degrees of freedom is 0.811. Since the value of the correlation coefficient is above 0.811, the null hypothesis (there is no relationship)is rejected. There is a statistical significance between the scores of Rater 2 at the two intervals. Figure 8 depicts the correlation of the Total Tool scores by Rater 3 at T_0 and T_1.

TABLE 11 - TOTAL SCORES FOR THE TOOL BY RATER 3 AT T_0 AND T_1

Article	**T_0**	**T_1**
Blanchard et al., 2011	27	29
Brin et al., 2009	22	25
Chambrone et al., 2012	33	34
Moore et al., 2011	23	22
Steiner et al., 2009	23	24
Tandon et al., 2010	28	26
Pearson's r	**0.904**	



Legend to Figure 8: Figure 8 shows the correlation of the Total Tool scores by Rater 3 at T_0(initial) and T_1(after 2 weeks). The x-axis denotes the Total Tool scores by Rater 3 at T_1. The y-axis denotes the Total Tool scores by Rater 3 at T_0. The diamond represents the correlation of the Total Tool scores by Rater 3 at T_0 and T_1.

DISCUSSION

Our purpose was to help users of IPDMA to critically appraise IPDMA. Therefore, we set out with the goal of developing a new instrument for assessing the methodological quality and reporting of IPDMA. In this study, we used the PRISMA as the criterion of reference to assess the quality of meta-analysis as currently this is the standard for investigators when reporting their findings and it provides a benchmark by which meta-analyses may be appraised. Previous research has outlined that critical appraisal may have aspects of objective and subjective assessment that cannot be reduced to a simple check list, therefore, we adapted the scoring of the PRISMA from a binomial nominal scale (yes/no) to a quantitative binomial scale (yes=2,No=1). This did not change the intent of its use. It simply deemphasized the importance of any one item even though there is recognition fit, and it helped to lay out the criteria to establish criterion validity.

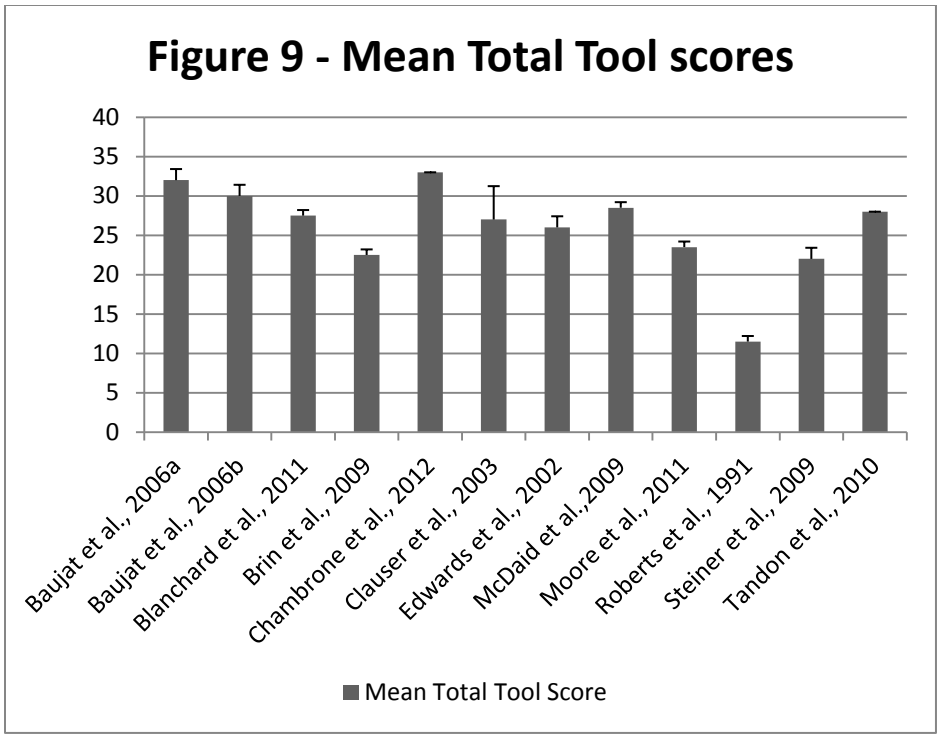
We constructed our Tool from a review of the literature by the Cochrane and keeping the appropriate framework for meta analysis as defined by the AHRQ. Following content refining, the Tool was established to consist of 9 Items. The content validity of the finalized Tool was established on the grounds of its construction process and the reasoning for each criteria in each of the items. We established construct validity of the Tool on the basis of evidence of test content, internal structure, response process, relation to other variables and consequences of testing.

We designed the scoring of the Tool as follows - each of the 9 items has at minimum 3 criteria. If a article satisfied 3 of the criteria, then 4 points were allocated. If it satisfied 2 of the criteria, then 3 points were allocated. If it satisfied 1 of the criteria then 2 points were allocated. Even if none of the criteria were satisfied, 1 point was allocated. One may argue that by giving 1 point to any article although it satisfies none of the criteria is a false process. However, the absolute score of each article is unimportant, instead, it is the relative score of each article that matters. In summary, the Tool we developed was established to have strong construct validity and strong content validity, and to yield semi-continuous scoring.

The Tool is relatively simple to implement and the scores obtained can then be directly compared. We standardized two independent readers(Rater 2 and Rater 3) in its use. This was accomplished by having them score for the "Introduction" portion(Item 1) of the Tool for the entire body of literature.Meta-analyses unrelated to interventions/ treatments addressing the diseases in the oral and maxillofacial region were excluded. Any divergence was resolved by a third reader(Rater 1). Once standardized, the readers scored the 12 articles that met our inclusion criteria with the modified PRISMA, as criterion, and the Tool. The resulting correlation demonstrated congruent criterion validity assessments across both readers. Having established the validity of our instrument, we established its inter-rater reliability (r=0.927). Intra-rater reliability is a Pearson correlation coefficient between two readers. The resulting ,statistically significant , correlation aided in establishing the reliability of the Tool. We did not see any specific Item being targeted while testing the Intra-rater reliability. The differences

caused between the readings at T_0 and T_1 (Refer to Table 10 and Table 11) are present all over the Tool in a random fashion, therefore pose no threat to the consistency of the Tool.

In brief, our study led to the development of a novel instrument specifically designed for assessing the quality of individual patient data meta-analysis. We established its construct, content and criterion validity; and we established its reliability. With this Tool, we obtained independent scores for each paper reviewed, which quantified the degree to which each paper satisfied certain criteria that are established in the literature as determinant factors of highest quality meta-analysis in general, and individual patient data meta-analysis in particular. Thus, we obtained total scores, which reflected the overall quality of each paper. Figure 9 shows the Mean Total Tool scores across both readers for each papers.



Legend to Figure 9: Figure 9 shows the Mean of Total Tool scores across both readers for all 12 articles. It also shows the standard deviation for each mean. Chambrone et al.,2012 has the highest mean total Tool score and Roberts et al.,1991 has the lowest mean total Tool score.

Roberts et al.,1991 has the lowest mean score value. This indicates that few criteria of meta-analysis quality were met in this paper. It may therefore be questionable as to whether the information provided in this paper may not be more harmful than beneficial to patients. By contrast, the highest mean score recorded for Chambrone et al.,2012 (Figure 9) reveals in an unambiguous quantified manner of the high quality of this paper. We infer that clinicians can trust more the information from Chambrone et al.,2012 than from Roberts et al.,1991.

Baujat et al.,2006a and Baujat et al.,2006b have quite different scores although they are from the same author (Refer Table 3 for PRISMA scores and Table 7 for Tool scores). This may be attributed to a different reporting style of the data. Despite the variability between the scores by the same authors, there is a correlation between the scores by the two readers. Both readers have scored Baujat et al.,2006ahigher than Baujat et al.,2006b. Considering our data of intra-rater reliability that emphasize how strongly standardized our readers were in this study, we infer that the scoring discrepancies for the paper are not due to systematic errors (e.g., unreliable scoring), but due to random errors.

We also note that the standard deviations are generally less than 5% of the means, which confirms a sound degree of replicablility between the standardized readers. However, Figure 9 also evinces that the standard deviation of the score for Clauser et al.,2003 is rather large (15% of the mean), which indicates variability between the readers' scores. Considering our data of intra-rater reliability that emphasize how strongly standardized our readers were in this study, we infer that the scoring discrepancies for the paper are not due to systematic errors (e.g., unreliable scoring), but to random errors, such as deriving from ambiguities in the paper itself. Therefore, we conclude that Clauser et al.,2003, while it may be of high or satisfactory quality for clinical relevance, may be unclear and ambiguous to the point of being unhelpful in generating the consensus statement.

In brief, the Tool we have developed and validated based on stringent psychometric criteria yields data for each paper that permit a systematic analysis aimed at determining whether or not a paper should be accepted, or not, in the process of crafting

the consensus statement. This acceptable sampling analysis is an extension of a similar analysis previously described (Kung et al., 2010; Phi et al.,2012).

A. Limitations

The tool has its limitations.

Any quantitative rating reduces and simplifies the complexity of the original article itself. For example, when the tool was utilized to rate Chambrone et al., 2012 the complexity of the article is narrowed to a single number, 33. This is the inherent limitation of any rating instrument.

There are limitations to our evaluation of the Tool as only two readers were used to collect data for this study. If more readers were used, preferably from different educational backgrounds and fields, the study may have had greater strength. However, since the purpose was to explore the construct and criterion validity of the tool, and two observations were made for each paper, which from a testing theory view is adequate, this was a good start given that no other tool exists to assess the quality of IPDMA. One of the two reviewers was involved in derivation of the instrument. A separate evaluation by independent users would help support the generalizability of the scoring tool.

Another possible limitation of this study is its sample size(12 papers). A search strategy carried out by a professional biomedical librarian yielded only 12 papers that were IPDMA in the field of oral and maxillo-facial medicine. If more papers were used, the study may have had greater strength. Due to the lack of number of IPDMA in the field of oral and maxillofacial medicine, further evaluation can be done as more data is gathered in future research.

B. Research Implications

There are several potential future applications of the Tool. Our instrument is an attempt to achieve consensus amongst current mainstream opinions. Inevitably, new evidence will modify current thinking in some areas of methodology of IPDMA and at that point the Tool will be updated. We recognize the need for further testing of the Tool. Additional studies are necessary with a focus on the reproducibility and construct validity of the Tool before strong recommendations can be made on its use. Validation of the tool by independent users will help support the generalizability of the tool. Validation of the Tool in fields other than oral and maxillofacial medicine will also support its use to assess the quality of IPDMA.

C. Dental Implications

The implications of the tool for a dental practice are particularly important for dentists.

The tool sets new standard for IPDMA to meet. Overall, it pushes the field of evidence based dentistry towards responsiveness to the needs, values and expressed preferences of the individual patient, leading to an improvement in evidence based dentistry. IPDMA ensures that when treatment is provided to any patient sitting in the waiting office of a dental clinic, the uniqueness of a patient is maintained. Instead of implementing evidence that has been obtained based on group data meta-analyses which may not always be generalized, the IPDMA approach maintains the patient's individuality rather than treating him/her as an "average patient". It essentially improves quality of care by individualizing the patient's treatment plan , thereby enabling dentists to increase patient satisfaction. Our Tool makes it simpler to evaluate and appraise these IPDMA studies, thereby easing their way inside a dental practice. The Tool will thus make the dentist want to approach evidence-based dentistry and apply it to their daily practice.

The tool in itself is a learning tool to teach methodology of IPDMA. The items have been dissected and the reasoning for each criteria has been explained. This helps increase one's understanding of IPDMA.

D. Future Directions

A significant need of a research is to connect the patient centered research with evidence based dentistry. With more patient-centered endpoints, such as toxicity and quality of life, becoming increasingly common, the next generation of systematic reviews will need to adapt accordingly. This may mean new methods of analyses will need to be developed and that methodological research will break new ground. The tool will therefore remain a living document and advances in empirical methodological research will be reflected in further improvements to the instrument. IPDMA is a way of evidence based dentistry adapting to foster the patient's sense of being respected and participating in their own health care decisions. It helps preserve the uniqueness of patients, their individual needs and preferences in the decision-making process. Our tool is a good start given that no other tool exists to assess the quality of IPDMA.

Knowledge is essentially resources misspent, if it is not disseminated and utilized appropriately by its stakeholders. Besides clinicians and patients, other potential stakeholders in this case are policymakers, scientists and lay-oriented organizations such as civic networks and mutual help organizations. IPDMA that are not well executed and clearly reported are of diminished value to the clinicians as well other stakeholders. Our tool facilitates the dissemination and utilization of IPDMA in daily clinical practices. However, it also faces an impending challenge of its own dissemination and utilization, inherent to any new critical appraisal tool. Via further development and testing, the Tool's potential can be fully realized, and an interest needs to be stimulated amongst the stakeholders, encouraging them to adopt and embrace it and thereby diffusing it's innovation. Ultimately, the Tool will help promote better decision-making by health care

providers and their "active" patients, so that both may meaningfully engage in informed colloquy about the nature and quality of care.

APPENDIX A- THE PRISMA CHECKLIST

Checklist of items to include when reporting a systematic review or meta-analysis

Section/topic	Item No	Checklist item	Reported on page No
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both	
Abstract			
Structured summary	2	Provide a structured summary including, as applicable, background, objectives, data sources, study eligibility criteria, participants, interventions, study appraisal and synthesis methods, results, limitations, conclusions and implications of key findings, systematic review registration number	
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	
Methods			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (such as web address), and, if available, provide registration information including registration number	
Eligibility criteria	6	Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale	
Information	7	Describe all information sources (such as	

Section/topic	Item No	Checklist item	Reported on page No
sources		databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	
Study selection	9	State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	
Data collection process	10	Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	
Data items	11	List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	
Summary measures	13	State the principal summary measures (such as risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I^2 statistic) for each meta-analysis	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)	

Section/topic	Item No	Checklist item	Reported on page No
Additional analyses	16	Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram	
Study characteristics	18	For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15)	
Additional analysis	23	Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)	
Discussion			
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as	

Section/topic	Item No	Checklist item	Reported on page No
		health care providers, users, and policy makers)	
Limitations	25	Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research	
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (such as supply of data) and role of funders for the systematic review	

APPENDIX B - THE TOOL

1. Was a structured summary(abstract) provided for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The background & objectives is present in the abstract.	☐
(B) The search methods and data analysis done is present in the abstract.	☐
(C) The main results and author's conclusions are outlined in the abstract.	☐

2. Was a Background (Introduction) provided for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Background gives a context of what is already known from existing literature.	☐
(B) The Background gives a Rationale and reason for why the IPDMA approach was sought.	☐
(C) The Background outlines the aims and outcomes hopes to be achieved with the IPDMA approach.	☐

3. Did the Methods section provide a Protocol for the IPDMA ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Methods section specifies the information regarding how and where the protocol can be accessed. ☐
(B) The Methods section specifies that the Protocol contains the methods and analyses to be used. ☐
(C) The Methods section specifies that the Protocol contains the outcomes & patient characteristics to be analyzed. ☐

4. Did the Methods section provide a Eligibility Criteria for considering studies for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The Inclusion and Exclusion criteria specifies the types of studies (RCTs, language restrictions, publication status etc).	☐
(B) The Inclusion and Exclusion criteria specifies the types of participants in the studies (age group, gender etc).	☐
(C) The Inclusion and Exclusion criteria specifies the types of interventions being compared.	☐
(D) The Inclusion and Exclusion criteria defines the outcome measures for studies to be considered.	☐

5. Did the Methods section provide a Search Strategy for identification of studies for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The different databases searched and the search strategy for each database are specified.	☐
(B) The methods used to avoid publication bias are specified.	☐
(C) It is specified if the searches are regularly updated and the last date this was done.	☐

(D) The studies obtained are assessed independently by two or more authors to reach a consensus on their eligibility for inclusion in the IPDMA. ☐

6. Did the Methods section specify the Data Collection Strategy for the IPDMA?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The number of authors approached and the proportion that provide the data should be mentioned along with the number of authors who did not provide the data, the reasons why, and the number of patients in the respective study is specified. ☐
(B) The authors who provide the data, if they give all of their data or only a proportion and in case of the latter, the reasons why information was omitted. ☐
(C) The reasons for excluding(or including) any patients who were originally excluded(or included) by the source authors. ☐
(D) The data collected is checked for validity, consistency & integrity of randomization and follow-up. ☐

7. Did the Results section provide the main results in an orderly manner?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The description of characteristics of the included & excluded studies including the number of patients in each study is specified. ☐
(B) The results of the main analysis along with statistical details such as how the clustering of patients within studies was accounted for. ☐
(C) The assessment of risk of bias in the included studies is done by two or more authors based on the risk of bias tool. ☐

8. Did the Discussion summarize the main findings ?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The summary of main findings is presented in a brief and balanced manner. ☐
(B) The results are compared with the published results and if they are not comparable, the reasons why should be mentioned. ☐

(C) The limitations of the IPDMA due to unavailable data are specified and if it might impact the results obtained. ☐

9. Did the Conclusion provide a general interpretation of the results in context of other evidence?

If it satisfies 3 of the criteria --> 4

If it satisfies 2 of the criteria --> 3

If it satisfies 1 of the criteria --> 2

If it satisfies 0 of the criteria --> 1

(that is even though none of the criteria were satisfied, the article will receive one point)

Criteria are:

(A) The implications for clinical practice are specified. ☐
(B) The implications for future research are specified. ☐
(C) The sources of funding for the IPDMA and the role of funders is specified. ☐

APPENDIX C: THE CRITICAL VALUE TABLE FOR PEARSON'S CORRELATION COEFFICIENT

	Level of Significance of a One-Tailed or Directional Test					
	H_0: $\rho \leq 0$ or H_0: $\rho \geq 0$					
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0005$
	Level of Significance of a Two-Tailed or Nondirectional Test					
	H_0: $\rho = 0$					
df	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	$\alpha = 0.001$
1	0.9511	0.9877	0.9969	0.9995	0.9999	0.9999
2	0.8000	0.9000	0.9500	0.9800	0.9900	0.9990
3	0.6870	0.8054	0.8783	0.9343	0.9587	0.9911
4	0.6084	0.7293	0.8114	0.8822	0.9172	0.9741
5	0.5509	0.6694	0.7545	0.8329	0.8745	0.9509
6	0.5067	0.6215	0.7067	0.7887	0.8343	0.9249
7	0.4716	0.5822	0.6664	0.7498	0.7977	0.8983
8	0.4428	0.5494	0.6319	0.7155	0.7646	0.8721
9	0.4187	0.5214	0.6021	0.6851	0.7348	0.8470
10	0.3981	0.4973	0.5760	0.6581	0.7079	0.8233
11	0.3802	0.4762	0.5529	0.6339	0.6835	0.8010
12	0.3646	0.4575	0.5324	0.6120	0.6614	0.7800
13	0.3507	0.4409	0.5140	0.5923	0.6411	0.7604
14	0.3383	0.4259	0.4973	0.5742	0.6226	0.7419
15	0.3271	0.4124	0.4821	0.5577	0.6055	0.7247
16	0.3170	0.4000	0.4683	0.5425	0.5897	0.7084
17	0.3077	0.3887	0.4555	0.5285	0.5751	0.6932
18	0.2992	0.3783	0.4438	0.5155	0.5614	0.6788
19	0.2914	0.3687	0.4329	0.5034	0.5487	0.6652
20	0.2841	0.3598	0.4227	0.4921	0.5368	0.6524
21	0.2774	0.3515	0.4132	0.4815	0.5256	0.6402
22	0.2711	0.3438	0.4044	0.4716	0.5151	0.6287
23	0.2653	0.3365	0.3961	0.4622	0.5052	0.6178
24	0.2598	0.3297	0.3882	0.4534	0.4958	0.6074
25	0.2546	0.3233	0.3809	0.4451	0.4869	0.5974
30	0.2327	0.2960	0.3494	0.4093	0.4487	0.5541
35	0.2156	0.2746	0.3246	0.3810	0.4182	0.5189
40	0.2018	0.2573	0.3044	0.3578	0.3932	0.4896
50	0.1806	0.2306	0.2732	0.3218	0.3542	0.4432
60	0.1650	0.2108	0.2500	0.2948	0.3248	0.4079
70	0.1528	0.1954	0.2319	0.2737	0.3017	0.3798
80	0.1430	0.1829	0.2172	0.2565	0.2830	0.3568
90	0.1348	0.1726	0.2050	0.2422	0.2673	0.3375
100	0.1279	0.1638	0.1946	0.2301	0.2540	0.3211
150	0.1045	0.1339	0.1593	0.1886	0.2084	0.2643
300	0.0740	0.0948	0.1129	0.1338	0.1480	0.1884
500	0.0573	0.0735	0.0875	0.1038	0.1149	0.1464
1000	0.0405	0.0520	0.0619	0.0735	0.0813	0.1038

APPENDIX D: INCLUDED STUDIES WITH THEIR ABSTRACTS.

(Sorted by date Published)

Study	Abstract
Roberts et al., 1991	Strategies for the advancement of surgical methods in cleft lip and palate. This paper examines the clinical research methodologies used for the evaluation of cleft lip and palate therapies. A survey of clinical reports in the Cleft Palate Journal between 1964 and 1988 revealed that almost all used retrospective methods (96%). The authors examine the merits and biases associated with retrospective evaluation of therapies and compared these to prospective randomized clinical trials. The strengths and weaknesses of clinical trials are discussed in relation to the long-term evaluation of primary surgery in cleft patients. For these to be successful, further work is needed to investigate questions such as sample size, possible predictors of long-term outcome, and improved methods of presurgical assessment. The authors conclude that if the uncertainties associated with the choice of primary cleft surgery are to be resolved, the challenge of multicenter prospective clinical trials must be faced by the various disciplines involved in cleft palate clinical research.
Edwards et al., 2002	Combination analgesic efficacy: individual patient data meta-analysis of single-dose oral tramadol plus acetaminophen in acute postoperative pain. The primary aims of this study were to assess the analgesic efficacy and adverse effects of single-dose oral tramadol plus acetaminophen in acute postoperative pain and to use meta-analysis to demonstrate the efficacy of

	the combination drug compared with its components. Individual patient data from seven randomized, double blind, placebo controlled trials of tramadol plus acetaminophen were supplied for analysis by the R.W. Johnson Pharmaceutical Research Institute, Raritan, New Jersey, USA. All trials used identical methods and assessed single-dose oral tramadol (75 mg or 112.5 mg) plus acetaminophen (650 mg or 975 mg) in adult patients with moderate or severe postoperative pain. Summed pain intensity and pain relief data over six and eight hours and global evaluations of treatment effect after eight hours were extracted. Number-needed-to-treat (NNT) for one patient to obtain at least 50% pain relief was calculated. NNTs derived from pain relief data were compared with those derived from pain intensity data and global evaluations. Information on adverse effects was collected. Combination analgesics (tramadol plus acetaminophen) had significantly lower (better) NNTs than the components alone, and comparable efficacy to ibuprofen 400 mg. This could be shown for dental but not postsurgical pain, because more patients were available for the former. Adverse effects were similar for the combination drugs and the opioid component alone. Common adverse effects were dizziness, drowsiness, nausea, vomiting, and headache. In sum, this meta-analysis demonstrated analgesic superiority of the combination drug over its components, without additional toxicity.
Clauser et al.,2003	Evidence-based mucogingival therapy. Part 2: Ordinary and individual patient data meta-analyses of surgical

	treatment of recession using complete root coverage as the outcome variable. BACKGROUND: The literature (1970-2000) on the outcome of surgical root coverage has been revised and summarized in a companion paper. The overall conclusion was that the various procedures are effective, but it was not possible to determine which procedure was best indicated in different clinical conditions. In this study, meta-analysis techniques were used to seek evidence for guiding clinical decisions when planning root coverage surgery. The aim of this study was to illustrate the differences between meta-analyses applied to summarized and individual patient data (IPD) and to present suggestions for reducing the costs of IPD meta-analysis.METHODS: Only clinical trials and case series that included data on the number of teeth treated, baseline recession depth (BRD) and the proportions of postoperative complete root coverage (CRC) were considered. The first group of meta-analyses (the outcome of each procedure based on summarized data) covered 65 studies dealing with coronally advanced flap (CAF), epithelial free gingival graft (EFGG), connective tissue graft (CTG), and guided tissue regeneration (GTR) procedures. The second group of meta-analyses was done to determine the outcome of each procedure on the basis of 26 studies that reported IPD for at least baseline recession depth (BRD) and final CRC for each site. The third group of meta-analysis compared the outcomes of CTG and GTR in 5 randomized studies, 4 of which reported only summarized data.RESULTS: The first

	analysis showed that CRC was achieved more often in non-randomized than in randomized studies. The heterogeneity tests revealed great variability of results in both the randomized and non-randomized studies, which makes it difficult to draw any definite conclusions. In the second analysis all the tested techniques revealed similar trends: greater baseline recession depths were always associated with a decreased CRC. The third analysis showed that CRC was achieved more frequently in the sites treated with CTG as opposed to GTR. The small sample size and the lack of IPD rendered the analyses inconclusive despite the randomized design. CONCLUSIONS: Few studies reported individual patient data; they are a valuable contribution to clinical decision making, but IPD published in the literature are still insufficient to provide a reliable guide for clinical decision making. Therefore, decisive steps should be taken to facilitate the publication of IPD, in electronic format, whenever a clinical study is published in a leading journal.
Baujat et al., 2006a	Chemotherapy as an adjunct to radiotherapy in locally advanced nasopharyngeal carcinoma. BACKGROUND: A previous meta-analysis investigated the role of chemotherapy in head and neck locally advanced carcinoma. This work had not been performed on nasopharyngeal carcinoma. OBJECTIVES: The aim of the project was to study the effect of adding chemotherapy to radiotherapy on overall survival (OS) and event-free survival (EFS) in patients with nasopharyngeal carcinoma. SEARCH STRATEGY: We

	searched MEDLINE (1966 to October 2003), EMBASE (1980 to October 2003) and the Cochrane Central Register of Controlled Trials (CENTRAL) (The Cochrane Library, Issue 3, 2003) and trial registers. Handsearches of meeting abstracts, references in review articles and of the Chinese medical literature were carried out. Experts and pharmaceutical companies were asked to identify trials. SELECTION CRITERIA: Randomised trials comparing chemotherapy plus radiotherapy to radiotherapy alone in locally advanced nasopharyngeal carcinoma were included. DATA COLLECTION AND ANALYSIS: The meta-analysis was based on updated individual patient data. The log rank test, stratified by trial, was used for comparisons and the hazard ratios (HR) of death and failure (loco-regional/distant failure or death) were calculated. MAIN RESULTS: Eight trials with 1753 patients were included. One trial with a 2 x 2 design was counted twice in the analysis. The analysis was performed including 11 comparisons based on 1975 patients. The median follow up was six years. The pooled hazard ratio of death was 0.82 (95% confidence interval (CI) 0.71 to 0.95; P = 0.006) corresponding to an absolute survival benefit of 6% at five years from chemotherapy (from 56% to 62%). The pooled hazard ratio of tumour failure or death was 0.76 (95% CI 0.67 to 0.86; P < 0.00001) corresponding to an absolute event-free survival benefit of 10% at five years from chemotherapy (from 42% to 52%). A significant interaction was observed between chemotherapy timings and overall survival (P = 0.005), explaining the

	heterogeneity observed in the treatment effect (P = 0.03) with the highest benefit from concomitant chemotherapy. AUTHORS' CONCLUSIONS: Chemotherapy led to a small but significant benefit for overall survival and event-free survival. This benefit was essentially observed when chemotherapy was administered concomitantly with radiotherapy.
Baujat et al., 2006b	Chemotherapy in locally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis of eight randomized trials and 1753 patients. OBJECTIVES: To study the effect of adding chemotherapy to radiotherapy (RT) on overall survival and event-free survival for patients with nasopharyngeal carcinoma.METHODS AND MATERIALS: This meta-analysis used updated individual patient data from randomized trials comparing chemotherapy plus RT with RT alone in locally advanced nasopharyngeal carcinoma. The log-rank test, stratified by trial, was used for comparisons, and the hazard ratios of death and failure were calculated.RESULTS: Eight trials with 1753 patients were included. One trial with a 2 x 2 design was counted twice in the analysis. The analysis included 11 comparisons using the data from 1975 patients. The median follow-up was 6 years. The pooled hazard ratio of death was 0.82 (95% confidence interval, 0.71-0.94; p = 0.006), corresponding to an absolute survival benefit of 6% at 5 years from the addition of chemotherapy (from 56% to 62%). The pooled hazard ratio of tumor failure or death was 0.76 (95% confidence interval, 0.67-0.86; p < 0.0001), corresponding to an absolute event-free survival

	benefit of 10% at 5 years from the addition of chemotherapy (from 42% to 52%). A significant interaction was observed between the timing of chemotherapy and overall survival (p = 0.005), explaining the heterogeneity observed in the treatment effect (p = 0.03), with the highest benefit resulting from concomitant chemotherapy. CONCLUSION: Chemotherapy led to a small, but significant, benefit for overall survival and event-free survival. This benefit was essentially observed when chemotherapy was administered concomitantly with RT.
McDaid et al., 2009	Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. OBJECTIVES: To determine the clinical effectiveness, safety and cost-effectiveness of continuous positive airway pressure (CPAP) devices for the treatment of obstructive apnoea-hypopnoea syndrome (OSAHS), compared with the best supportive care, placebo and dental devices. DATA SOURCES: The main search was of fifteen electronic databases, including MEDLINE, EMBASE and the Cochrane Library, up to November 2006.REVIEW METHODS: Randomised controlled trials (RCTs) comparing CPAP with best supportive/usual care, placebo, and dental devices in adults with a diagnosis of OSAHS were included. The primary outcomes of interest were subjective daytime sleepiness assessed by the Epworth Sleepiness Scale (ESS) and objective sleepiness assessed by the Maintenance of Wakefulness Test (MWT) and the

	Multiple Sleep Latency Test (MSLT). A new economic model was developed to assess incremental cost per quality-adjusted life-year (QALY). The cost-effectiveness of CPAP was compared with that of the use of dental devices and conservative management. The costs and QALYs were compared over a lifetime time horizon. Effectiveness was based on the RCT evidence on sleepiness symptoms (ESS), which was 'mapped' to utilities using individual patient data from a subset of studies. Utilities were expressed on the basis of generic HRQoL instruments [the EQ-5D (EuroQoL-5 Dimensions) in the base-case analysis]. The base-case analysis focused on a male aged 50. A series of subgroup and scenario analyses were also undertaken. RESULTS: The searches yielded 6325 citations, from which 48 relevant clinical effectiveness studies were identified, 29 of these providing data on daytime sleepiness. The majority of the included RCTs did not report using an adequate method of allocation concealment or use an intention-to-treat analysis. Only the studies using a sham CPAP comparator were double blinded. There was a statistically significant benefit with CPAP compared with control (placebo and conservative treatment/usual care) on the ESS [mean difference (MD) -2.7 points, 95% CI -3.45 to -1.96]. However, there was statistical heterogeneity, which was reduced when trials were subgrouped by severity of disease. There was also a significant benefit with CPAP compared with usual care on the MWT. There was a non-statistically significant difference between CPAP and dental devices (six trials)

	in the impact on daytime sleepiness (ESS) among a population with moderate symptom severity at baseline (MD -0.9, 95% CI -2.1 to 0.4). A review of five studies evaluating the cost-effectiveness of CPAP was undertaken. All existing cost-effectiveness studies had limitations; therefore a new economic model was developed, based on which it was found that, on average, CPAP was associated with higher costs and benefits than dental devices or conservative management. The incremental cost per QALY gained of CPAP was below 20,000 pounds in the base-case analysis and most alternative scenarios. There was a high probability of CPAP being more cost-effective than dental devices and conservative management for a cost-effectiveness threshold of 20,000 pounds per QALY gained. CONCLUSIONS: CPAP is an effective and cost-effective treatment for OSAHS compared with conservative/usual care and placebo in populations with moderate to severe daytime sleepiness, and there may be benefits when the disease is mild. Dental devices may be a treatment option in moderate disease but some uncertainty remains. Further research would be potentially valuable, particularly investigation of the effectiveness of CPAP for populations with mild sleepiness and further trials comparing CPAP with dental devices.
Steiner et al.,2009	Gastrointestinal tolerability of aspirin and the choice of over-the-counter analgesia for short-lasting acute pain. RATIONALE: For the management of common disorders producing short-lasting pain, there is very good

	evidence of the efficacy of aspirin. Yet paracetamol is often preferred, despite that evidence of its efficacy is much less sound. The reason for this appears to be a concern over gastrointestinal (GI) toxicity. If this concern is misplaced, so may be the preference for paracetamol, with the consequence of widespread sub-optimal treatment. Our purpose in this analysis of pooled individual patient data from clinical studies of aspirin is to adduce the evidence that will show whether or not this is so, for the benefit of consumers and health-care professionals who advise them. METHODS: The frequencies of all and GI adverse events (AEs) and adverse drug reactions (ADRs) were calculated from the pooled individual patient data of nine similar randomized, double-blind, placebo controlled clinical trials of single-doses of aspirin 1000 mg in the treatment of acute migraine attacks, episodic tension-type headache and dental pain. Absolute differences between active and placebo AE and ADR rates, and numbers-needed-to-harm (NNH), were calculated. RESULTS: Of 2852 patients included in the analysis, 1581 were treated with aspirin and 1271 with placebo. Reported AE rates were 14.9% and 11.1% amongst patients allocated to aspirin and placebo respectively (NNH: 26), with the GI system most frequently affected (aspirin: 5.9%; placebo: 3.5%; NNH: 42). Reported ADR rates were much lower (aspirin: 6.3%; placebo: 3.9%; NNH: 42), especially for the GI system (aspirin: 3.1%; placebo: 2.0%; NNH: 91). Most of the AEs and ADRs were mild or moderate, and none was serious. CONCLUSIONS: The GI ADR

	differences between aspirin and placebo are not great enough to support decision choices for short-lasting acute pain based on tolerability: these are better based on efficacy.
Brin et al.,2009	Safety and tolerability of onabotulinumtoxinA in the treatment of facial lines: a meta-analysis of individual patient data from global clinical registration studies in 1678 participants. BACKGROUND: OnabotulinumtoxinA for the treatment of facial lines is a widely used cosmetic medical procedure and, as such, the safety and tolerability profile is of interest to health care providers and patients. Based on data from individual studies that were conducted according to regulatory guidelines to provide adequate safety and efficacy data to support product licensure (registration studies), the overall benefit:risk profile of onabotulinumtoxinA for facial lines has been favorable. OBJECTIVE: Our objective was to increase statistical power through meta-analysis to detect treatment group differences in adverse event (AE) incidence that may not have been evident in individual registration studies. METHODS: Individual participant data (n = 1678) were from 6 randomized, double-blind, placebo-controlled and 3 open-label studies. Two double-blind, placebo-controlled studies were for lateral canthal lines (3-18 U/side) and all others were for glabellar lines (10 or 20 U). Doses used reflect global product labeling in countries where licensed. RESULTS: Participant population was non-Hispanic white (43%) or Asian (52%) and predominantly female (88%). In double-blind,

	placebo-controlled studies, overall AE incidence did not significantly differ by treatment group (onabotulinumtoxinA vs placebo). The only individual AEs with significantly greater incidence in the onabotulinumtoxinA group were eyelid sensory disorder (2.5% vs 0.3%, P = .004; verbatim phrases "tight," "pressured," "heavy," "drooping feeling," "feeling of droopiness") and eyelid ptosis (1.8% vs 0%, P = .02), both present only in glabellar studies. Overall treatment-related (per investigator) AE incidence was greater in the onabotulinumtoxinA group versus placebo (24% vs 16%, P = .005), and treatment-related eyelid edema was an additional AE with significantly higher incidence in the onabotulinumtoxinA group versus placebo (P = .04). Incidence of all 3 of these AEs significantly decreased as number of treatment cycles increased. Eyelid sensory disorder and eyelid edema were more common in Asian participants. Acne, injection site pruritus, oral herpes, rash, lower respiratory tract infection, dental caries, and eye pain were significantly more common in placebo-treated compared with onabotulinumtoxinA-treated participants. Serious AE incidence did not significantly differ by treatment (onabotulinumtoxinA vs placebo) and no serious AEs were treatment related. There were no symptoms of weakness remote to the injection site or related to the central nervous system. LIMITATIONS: Limitations included: (1) highly visible efficacy of onabotulinumtoxinA may have resulted in reporting bias; (2) reliance on participant intervisit recall; (3) a relatively short follow-up period (1 year); (4) conclusions are based

	solely on the doses analyzed (ie, those used in the respective trials); and (5) exclusion of patients with severe medical disease in registration studies. CONCLUSION: This meta-analysis confirms the safety and tolerability of onabotulinumtoxinA for glabellar and lateral canthal lines, at the doses studied, based on the most comprehensive controlled safety analysis of onabotulinumtoxinA performed to date. The AEs observed were generally mild to moderate; most treatment-related AEs were related either to physical injection of product or local pharmacologic effects. Even with the increased statistical power of a large sample size, no new onabotulinumtoxinA-associated AEs emerged.
Tandon et al., 2010	A systematic review of p53 as a prognostic factor of survival in squamous cell carcinoma of the four main anatomical subsites of the head and neck. OBJECTIVES: To summarize existing evidence about whether the presence of mutant or upregulated p53 is a prognostic factor for patients presenting with squamous cell carcinoma arising from the larynx, oropharynx, hypopharynx, or oral cavity. METHOD: Relevant articles were identified using strict criteria for systematic searches. Associations between mutant or upregulated p53 versus wild-type or low/undetectable p53 in relation to overall survival and DFS were summarized by extracting or deriving hazard ratio (HR) estimates. Random-effects meta-analyses were used to account for between-study heterogeneity and to summarize the effect of p53 across studies. RESULTS: The meta-analyses

	gave a statistically significant pooled HR for overall survival in oral cavity [pooled HR, 1.48; 95% confidence interval, (95% CI), 1.03-2.11], and for disease-free survival in oral cavity (pooled HR, 1.47; 95% CI, 1.12-1.93) and in oropharynx (pooled HR, 0.45; 95% CI, 0.27-0.73). Despite attempts to limit it, between-study heterogeneity was large in the majority of meta-analyses and the prognostic value of p53 was generally inconsistent and inconclusive across studies. CONCLUSION: The meta-analysis results highlight that current evidence about the prognostic value of p53 in patients with squamous cell carcinoma of the head and neck is inconclusive. Large heterogeneity exists across studies in study-level and patient-level characteristics, making it difficult to ascertain a clear picture. Future studies are required in which p53 expression is investigated in a more standardized and biologically informative manner. In particular, prospectively planned individual patient data meta-analyses are needed to establish the prognostic importance of p53 for specific subgroups of patients undergoing specific treatments.
Moore et al.,2011	Minimum efficacy criteria for comparisons between treatments using individual patient meta-analysis of acute pain trials: examples of etoricoxib, paracetamol, ibuprofen, and ibuprofen/paracetamol combinations after third molar extraction. We defined response in acute pain trials according to percentage of maximum possible efficacy. Minimum efficacy criteria (MEC) of 0%, or at least 15%, 30%, 50%, and 70% pain relief were used to examine stability

	over time using total pain relief and summed pain intensity difference (SPID), sex differences, and sensitivity. We used individual patient data from placebo-controlled third molar extraction trials: 4 with single-dose oral etoricoxib 120 mg, and 2 with paracetamol, ibuprofen, and ibuprofen plus paracetamol combinations. With etoricoxib, numbers needed to treat (NNTs) were stable between response levels of at least 15% (MEC15) and 50% pain relief (MEC50), and similar for total pain relief and SPID. NNTs were higher (worse) at extremes of MEC, especially with SPID. Results for women and men were similar. NNTs of lower efficacy treatments (paracetamol 500 and 1000 mg) rose rapidly at higher MEC. NNTs of high efficacy treatments (ibuprofen plus paracetamol combinations) showed greater separation at higher MEC. The highest degree of discrimination between treatments was with MEC50 and MEC70. Etoricoxib 120 mg (NNT for ≥50% maximum 6-hour pain relief 1.7) and ibuprofen 200/400 mg plus paracetamol 500/1000 mg (NNTs 1.5 and 1.6, respectively) produced the lowest (best) NNTs in the dental pain model. Timing of patient request for additional analgesia is an alternative analgesic efficacy outcome measure.
Blanchard et al.,2011	Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. INTRODUCTION: The recently updated meta-analysis of chemotherapy in head and neck cancer (MACH-NC) demonstrated the benefit of the addition of chemotherapy in terms of overall survival in head and neck squamous

	cell carcinoma (HNSCC). The magnitude of the benefit according to tumour site is unknown as well as their potential interactions with patient or trial characteristics. METHODS: Eighty seven randomized trials performed between 1965 and 2000 were included in the present analysis. Patients were divided into four categories according to tumour location: oral cavity, oropharynx, hypopharynx and larynx. Patients with other tumour location were excluded (999, 5.7%). For each tumour location and chemotherapy timing, the logrank-test, stratified by trial, was used to compare treatments. The hazard ratios of death or relapse were calculated. Interactions between patient or trial characteristics and chemotherapy effect were studied. RESULTS: Individual patient data of 16,192 patients were analysed, with a median follow-up of 5.6years. The benefit of the addition is consistent in all tumour locations, with hazard ratios between 0.87 and 0.88 (p-value of interaction=0.99). Chemotherapy benefit was higher for concomitant administration for all tumour locations, but the interaction test between chemotherapy timing and treatment effect was only significant for oropharyngeal ($p<0.0001$) and laryngeal tumours ($p=0.05$), and not for oral cavity ($p=0.15$) and hypopharyngeal tumours ($p=0.30$). The 5-year absolute benefits associated with the concomitant chemotherapy are 8.9%, 8.1%, 5.4% and 4% for oral cavity, oropharynx, larynx and hypopharynx tumours, respectively. CONCLUSION: The benefit of the addition of chemotherapy to locoregional treatment is consistent in all tumour locations of HNSCC. The higher

	benefit of concomitant schedule was demonstrated only for oropharyngeal and laryngeal tumours but this may be only a consequence of a lack of power.
Chambrone et al.,2012	Evidence-based periodontal plastic surgery. II. An individual data meta-analysis for evaluating factors in achieving complete root coverage. BACKGROUND: The aim of this review is to conduct an individual patient data meta-analysis of randomized controlled clinical trials (RCTs) to evaluate whether baseline recession-, patient-, and procedure-related factors can influence the achievement of complete root coverage (CRC). METHODS: A literature search with no restrictions regarding status or the language of publication was performed for MEDLINE (for Medical Literature Analysis and Retrieval System Online), EMBASE (for Excerpta Medica Database), CENTRAL (for Cochrane Central Register of Controlled Trials), and the Cochrane Oral Health Group's Specialized Register databases up to and including March 2011. Only RCTs, with a duration of ≥6 months evaluating recession areas (Miller Class I or II) that were treated by means of root coverage procedures were included. Mixed-effects logistic regression analyses were conducted to evaluate associations between five baseline variables and CRC. RESULTS: Of the 70 potentially eligible trials, 22 were included in the meta-analyses. In total, the data from 320 patients and 16 procedures were evaluated. None of the RCTs were classified as low risk of bias. Of the 602 recessions treated, 310 (51.5%) achieved CRC. Subepithelial connective tissue grafts (SCTGs), matrix

	grafts, and enamel matrix derivative protein (EMD) procedures were superior in achieving CRC when compared to coronally advanced flap (CAF) alone. For the adjusted covariates, the greater the baseline recession depth, the smaller the chance of achieving CRC (individual procedure analysis [odds ratio (OR) = 0.55; 95% confidence interval (CI) = 0.44, 0.70] and grouped procedure analysis [OR = 0.56; 95% CI = 0.45, 0.71]), as well as studies with conflict of interest were more likely to achieve CRC than those without conflict of interest (individual procedure analysis [OR = 6.78; 95% CI = 1.78, 25.86]). CONCLUSIONS: SCTGs, matrix grafts, and EMD were superior to CAF in achieving CRC, but SCTGs showed the best predictability. The impossibility of inclusion of all identified RCTs should be taken into consideration when interpreting the present findings.

REFERENCES

The efficacy of aspirin in patients with atrial fibrillation Analysis of pooled data from 3 randomized trials The Atrial Fibrillation Investigators. Arch Intern Med. 1997 Jun 9;157(11):1237-40. PubMed PMID: 9183235.

American Educational Research Association, American Psychological Association,National Council on Measurement in Education. Standards for Educational and Psychological Testing. 1999.

Baujat B, Audry H, Bourhis J, Chan AT, Onat H, Chua DT, Kwong DL, Al-Sarraf M, Chi KH, Hareyama M, Leung SF, Thephamongkhol K, Pignon JP, MAC-NPC Collaborative Group. Chemotherapy in locally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis of eight randomized trials and 1753 patients. Int J Radiat Oncol Biol Phys. 2006 Jan 1;64(1):47-56. PubMed PMID: 16377415.

Baujat B, Audry H, Bourhis J, Chan AT, Onat H, Chua DT, Kwong DL, Al-Sarraf M, Chi KH, Hareyama M, Leung SF, Thephamongkhol K, Pignon JP, MAC-NPC Collaborative Group. Chemotherapy as an adjunct to radiotherapy in locally advanced nasopharyngeal carcinoma. Cochrane Database Syst Rev. 2006 Oct 18;PubMed PMID: 17054200.

Blanchard P, Baujat B, Holostenco V, Bourredjem A, Baey C, Bourhis J, Pignon JP, MACH-CH Collaborative group. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. Radiother Oncol. 2011 Jul;100(1):33-40. PubMed PMID: 21684027.

Brin MF, Boodhoo TI, Pogoda JM, James LM, Demos G, Terashima Y, Gu J, Eadie N, Bowen BL. Safety and tolerability of onabotulinumtoxinA in the treatment of facial lines: a meta-analysis of individual patient data from global clinical registration studies in 1678 participants. J Am Acad Dermatol. 2009 Dec;61(6):961-70.e1-11. PubMed PMID: 19744746.

Carmines E, Zeller R. Reliability and Validity Assessment. Newbury Park, CA: Sage Publications; 1979.

Cassidy J, Saltz L, Twelves C, Van Cutsem E, Hoff P, Kang Y, Saini JP, Gilberg F, Cunningham D. Efficacy of capecitabine versus 5-fluorouracil in colorectal and gastric cancers: a meta-analysis of individual data from 6171 patients. Ann Oncol. 2011 Dec;22(12):2604-9. PubMed PMID: 21415237.

Chambrone L, Pannuti CM, Tu YK, Chambrone LA. Evidence-based periodontal plastic surgery II An individual data meta-analysis for evaluating factors in achieving complete root coverage. J Periodontol. 2012 Apr;83(4):477-90. PubMed PMID: 21859324.

Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration. Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: a systematic review and meta-analysis of individual patient data from 18 randomized trials. J Clin Oncol. 2008 Dec 10;26(35):5802-12. PubMed PMID: 19001332; PubMed Central PMCID: PMC2645100.

Chiappelli F, Phil AB, Arora R, Phi L, Giroux A, Uyeda M, Kung J, Ramchandani M. Reliability of Quality Assessments in Research Synthesis: Securing the Highest Quality

Bioinformation for HIT. Bioinformation. 2012;8(14):691-4. PubMed PMID: 23055612; PubMed Central PMCID: PMC3449364.

Chiappelli F, Brant X, Oluwadara O, Neagos N, Ramchandani M. Understanding Evidence-Based Practice: Toward Optimizing Clinical Outcomes. Chiappelli F, Brant X, Oluwadara O, Neagos N, Ramchandani M, editors. Heidelberg: Springer–Verlag; 2010.

Clauser C, Nieri M, Franceschi D, Pagliaro U, Pini-Prato G. Evidence-based mucogingival therapy Part 2: Ordinary and individual patient data meta-analyses of surgical treatment of recession using complete root coverage as the outcome variable. J Periodontol. 2003 May;74(5):741-56. PubMed PMID: 12816306.

Collins SL, Edwards J, Moore RA, Smith LA, McQuay HJ. Seeking a simple measure of analgesia for mega-trials: is a single global assessment good enough?. Pain. 2001 Mar;91(1-2):189-94. PubMed PMID: 11240091.

Cortina JM. What is coefficient alpha: an examination of theory and applications. Journal of Applied Psychology. 1993 ; 78:98-104.

Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychomerika. 1951 ; 16:297-334.

Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol. 2011 Jan;64(1):79-89. PubMed PMID: 21130354.

Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol. 2012 Apr;65(4):375-83. PubMed PMID: 22078576.

Dávalos A, Castillo J, Alvarez-Sabín J, Secades JJ, Mercadal J, López S, Cobo E, Warach S, Sherman D, Clark WM, Lozano R. Oral citicoline in acute ischemic stroke: an individual patient data pooling analysis of clinical trials. Stroke. 2002 Dec;33(12):2850-7. PubMed PMID: 12468781.

De Backer TL, Vander Stichele R, Lehert P, Van Bortel L. Naftidrofuryl for intermittent claudication. Cochrane Database Syst Rev. 2008 Apr 16;PubMed PMID: 18425872.

de Backer TL, Vander Stichele R, Lehert P, Van Bortel L. Naftidrofuryl for intermittent claudication. Cochrane Database Syst Rev. 2012 Dec 12;12:CD001368. PubMed PMID: 23235580.

Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA. 1990 Mar 9;263(10):1385-9. PubMed PMID: 2406472.

Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results Follow-up of applications submitted to two institutional review boards. JAMA. 1992 Jan 15;267(3):374-8. PubMed PMID: 1727960.

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet. 1991 Apr 13;337(8746):867-72. PubMed PMID: 1672966.

Ebrahim S, Smith GD. Systematic review of randomised controlled trials of multiple risk factor interventions for preventing coronary heart disease. BMJ. 1997 Jun 7;314(7095):1666-74. PubMed PMID: 9193292; PubMed Central PMCID: PMC2126863.

Edwards JE, McQuay HJ, Moore RA. Combination analgesic efficacy: individual patient data meta-analysis of single-dose oral tramadol plus acetaminophen in acute postoperative pain. J Pain Symptom Manage. 2002 Feb;23(2):121-30. PubMed PMID: 11844632.

Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. Stat Med. 2002 Jun 15;21(11):1635-40. PubMed PMID: 12111924.

Fonseca V, Davidson J, Home P, Snyder J, Jellinger P, Dyhr Toft A, Barnett A. Starting insulin therapy with basal insulin analog or premix insulin analog in T2DM: a pooled analysis of treat-to-target trials. Curr Med Res Opin. 2010 Jul;26(7):1621-8. PubMed PMID: 20429817.

Freeman R. A psychodynamic understanding of the dentist-patient interaction. Br Dent J. 1999 May 22;186(10):503-6. PubMed PMID: 10379083.

Gerlinger C, Endrikat J, Kallischnigg G, Wessel J. Evaluation of menstrual bleeding patterns: a new proposal for a universal guideline based on the analysis of more than 4500 bleeding diaries. Eur J Contracept Reprod Health Care. 2007 Sep;12(3):203-11. PubMed PMID: 17763258.

Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ. 1995 Nov 18;311(7016):1356-9. PubMed PMID: 7496291; PubMed Central PMCID: PMC2551234.

Groves L, Shellenberger MK, Davis CS. Tizanidine treatment of spasticity: a meta-analysis of controlled, double-blind, comparative studies with baclofen and diazepam. Adv Ther. 1998 Jul-Aug;15(4):241-51. PubMed PMID: 10186943.

Harris ST, Blumentals WA, Miller PD. Ibandronate and the risk of non-vertebral and clinical fractures in women with postmenopausal osteoporosis: results of a meta-analysis of phase III studies. Curr Med Res Opin. 2008 Jan;24(1):237-45. PubMed PMID: 18047776.

Hashiguchi M, Ohno K, Nakazawa R, Kishino S, Mochizuki M, Shiga T. Comparison of cilostazol and ticlopidine for one-month effectiveness and safety after elective coronary stenting. Cardiovasc Drugs Ther. 2004 May;18(3):211-7. PubMed PMID: 15229389.

Heneghan C, Ward A, Perera R, Self-Monitoring Trialist Collaboration, Bankhead C, Fuller A, Stevens R, Bradford K, Tyndel S, Alonso-Coello P, Ansell J, Beyth R, Bernardo A, Christensen TD, Cromheecke ME, Edson RG, Fitzmaurice D, Gadisseur AP, Garcia-Alamino JM, Gardiner C, Hasenkam JM, Jacobson A, Kaatz S, Kamali F, Khan TI, Knight E, Körtke H, Levi M, Matchar D, Menéndez-Jándula B, Rakovac I, Schaefer C, Siebenhofer A, Souto JC, Sunderji R, Gin K, Shalansky K, Völler H, Wagner O, Zittermann A. Self-monitoring of oral anticoagulation: systematic review and meta-analysis of individual patient data. Lancet. 2012 Jan 28;379(9813):322-34. PubMed PMID: 22137798.

Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Stat Med. 2001 Aug 15;20(15):2219-41. PubMed PMID: 11468761.

Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928. PubMed PMID: 22008217; PubMed Central PMCID: PMC3196245.

Higginson IJ, Harding R. Outcome measurement. Addington-Hall JM, Bruera E, Higginson IJ, Payne S, editors. Oxford; New York: Oxford University Press; 2007. 99-110p.

Home PD, Fritsche A, Schinzel S, Massi-Benedetti M. Meta-analysis of individual patient data to assess the risk of hypoglycaemia in people with type 2 diabetes using NPH insulin or insulin glargine. Diabetes Obes Metab. 2010 Sep;12(9):772-9. PubMed PMID: 20649629.

Huque M. Experiences with meta-analysis in NDA submissions. Proc Biopharm Section Am Stat Assoc. 1988 ; 2:28-33.

Hurwitz HI, Saltz LB, Van Cutsem E, Cassidy J, Wiedemann J, Sirzén F, Lyman GH, Rohr UP. Venous thromboembolic events with chemotherapy plus bevacizumab: a pooled analysis of patients in randomized phase II and III studies. J Clin Oncol. 2011 May 1;29(13):1757-64. PubMed PMID: 21422411.

James PA, Cowan TM, Graham RP. Patient-centered clinical decisions and their impact on physician adherence to clinical guidelines. J Fam Pract. 1998 Apr;46(4):311-8. PubMed PMID: 9564373.

Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. Clin Trials. 2009 Feb;6(1):16-27. PubMed PMID: 19254930.

Kovalchik SA. Mother's CD4+ count moderates the risk associated with higher parity for late postnatal HIV-free survival of breastfed children: an individual patient data meta-analysis of randomized controlled trials. AIDS Behav. 2012 Jan;16(1):79-85. PubMed PMID: 21739289.

Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. Open Dent J. 2010 Jul 16;4:84-91. PubMed PMID: 21088686; PubMed Central PMCID: PMC2948145.

Lampl C, Voelker M, Diener HC. Efficacy and safety of 1,000 mg effervescent aspirin: individual patient data meta-analysis of three trials in migraine headache and migraine accompanying symptoms. J Neurol. 2007 Jun;254(6):705-12. PubMed PMID: 17406776.

Lee CK, Lord SJ, Stockler MR, Coates AS, Gebski V, Simes RJ. Historical cross-trial comparisons for competing treatments in advanced breast cancer--an empirical analysis of bias. Eur J Cancer. 2010 Feb;46(3):541-8. PubMed PMID: 20036117.

Lehert P, Chéron G, Calatayud GA, Cézard JP, Castrellón PG, Garcia JM, Santos M, Savitha MR. Racecadotril for childhood gastroenteritis: an individual patient data meta-analysis. Dig Liver Dis. 2011 Sep;43(9):707-13. PubMed PMID: 21514257.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009 Jul 21;6(7):e1000100. PubMed PMID: 19621070; PubMed Central PMCID: PMC2707010.

Mahr AD, Jover JA, Spiera RF, Hernández-García C, Fernández-Gutiérrez B, Lavalley MP, Merkel PA. Adjunctive methotrexate for treatment of giant cell arteritis: an individual patient data meta-analysis. Arthritis Rheum. 2007 Aug;56(8):2789-97. PubMed PMID: 17665429.

Mason BJ, Lehert P. Acamprosate for alcohol dependence: a sex-specific meta-analysis based on individual patient data. Alcohol Clin Exp Res. 2012 Mar;36(3):497-508. PubMed PMID: 21895717; PubMed Central PMCID: PMC3288465.

Mathew T, Nordström K. On the equivalence of meta-analysis using literature and using individual patient data. Biometrics. 1999 Dec;55(4):1221-3. PubMed PMID: 11315071.

Maynard A. Evidence-based medicine: an incomplete method for informing treatment choices. Lancet. 1997 Jan 11;349(9045):126-8. PubMed PMID: 8996435.

McDaid C, Griffin S, Weatherly H, Durée K, van der Burgt M, van Hout S, Akers J, Davies RJ, Sculpher M, Westwood M. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. Health Technol Assess. 2009 Jan;13(4):iii-iv, xi-xiv, 1-119, 143-274. PubMed PMID: 19103134.

Messick S. Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning. American Psychologist. 1995 ; 50:741-9.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement Quality of Reporting of Meta-analyses. Lancet. 1999 Nov 27;354(9193):1896-900. PubMed PMID: 10584742.

Moore RA, Straube S, Paine J, Derry S, McQuay HJ. Minimum efficacy criteria for comparisons between treatments using individual patient meta-analysis of acute pain trials: examples of etoricoxib, paracetamol, ibuprofen, and ibuprofen/paracetamol combinations after third molar extraction. Pain. 2011 May;152(5):982-9. PubMed PMID: 21414722.

Moore RA, McQuay HJ. Single-patient data meta-analysis of 3453 postoperative patients: oral tramadol versus placebo, codeine and combination analgesics. Pain. 1997 Feb;69(3):287-94. PubMed PMID: 9085303.

Nunnally JC. Psychometric Theory. 2nd ed. New York: McGraw-Hill Book Company; 1978.

Okines AF, Norman AR, McCloud P, Kang YK, Cunningham D. Meta-analysis of the REAL-2 and ML17032 trials: evaluating capecitabine-based combination chemotherapy and infused 5-fluorouracil-based combination chemotherapy for the treatment of advanced oesophago-gastric cancer. Ann Oncol. 2009 Sep;20(9):1529-34. PubMed PMID: 19474114.

Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. Biometrics. 1998 Mar;54(1):317-22. PubMed PMID: 9544524.

Pan Q, Ornstein S, Gross AJ, Hueston WJ, Jenkins RG, Mainous AG 3rd, Silverstein MD. Antibiotics and return visits for respiratory illness: a comparison of pooled versus hierarchical statistical methods. Am J Med Sci. 2000 Jun;319(6):360-5. PubMed PMID: 10875290.

Pawinski A, Sylvester R, Kurth KH, Bouffioux C, van der Meijden A, Parmar MK, Bijnens L. A combined analysis of European Organization for Research and Treatment of Cancer, and Medical Research Council randomized clinical trials for the prophylactic treatment of stage TaT1 bladder cancer European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group and the Medical Research Council Working Party on Superficial Bladder Cancer. J Urol. 1996 Dec;156(6):1934-40, discussion 1940-1. PubMed PMID: 8911360.

Perera R, Heneghan C, Fitzmaurice D, Self Monitoring Trialists (SMT) collaboration. Individual patient meta-analysis of self-monitoring of an oral anticoagulation protocol. J Heart Valve Dis. 2008 Mar;17(2):233-8. PubMed PMID: 18512497.

Phi L, Ajaj R, Ramchandani MH, Brant XM, Oluwadara O, Polinovsky O, Moradi D, Barkhordarian A, Sriphanlop P, Ong M, Giroux A, Lee J, Siddiqui M, Ghodousi N, Chiappelli F. Expanding the Grading of Recommendations Assessment, Development, and Evaluation (Ex-GRADE) for Evidence-Based Clinical Recommendations: Validation Study. Open Dent J. 2012;6:31-40. PubMed PMID: 22303416; PubMed Central PMCID: PMC3269009.

Pignon JP, Baujat B, Bourhis J. [Individual patient data meta-analyses in head and neck carcinoma: what have we learnt?]. Cancer Radiother. 2005 Feb;9(1):31-6. PubMed PMID: 15804617.

Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010 Feb 5;340:c221. PubMed PMID: 20139215.

Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutitie F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med. 2008 May 20;27(11):1870-93. PubMed PMID: 18069721.

Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med. 2008 Dec 20;27(29):6111-36. PubMed PMID: 18816508.

Roberts CT, Semb G, Shaw WC. Strategies for the advancement of surgical methods in cleft lip and palate. Cleft Palate Craniofac J. 1991 Apr;28(2):141-9. PubMed PMID: 1829965.

Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the "evidence cart". JAMA. 1998 Oct 21;280(15):1336-8. PubMed PMID: 9794314.

Sakamoto J, Hamada C, Kodaira S, Nakazato H, Ohashi Y. Adjuvant therapy with oral fluoropyrimidines as main chemotherapeutic agents after curative resection for colorectal cancer: individual patient data meta-analysis of randomized trials. Jpn J Clin Oncol. 1999 Feb;29(2):78-86. PubMed PMID: 10089948.

Sakamoto J, Kodaira S, Hamada C, Ito K, Maehara Y, Takagi H, Sugimachi K, Nakazato H, Ohashi Y, Meta-Analysis Group of the Japanese Society of Strategies for Cancer Research and Therapy. An individual patient data meta-analysis of long supported adjuvant chemotherapy with oral carmofur in patients with curatively resected colorectal cancer. Oncol Rep. 2001 May-Jun;8(3):697-703. PubMed PMID: 11295105.

Sakamoto J, Hamada C, Rahman M, Kodaira S, Ito K, Nakazato H, Ohashi Y, Yasutomi M. An individual patient data meta-analysis of adjuvant therapy with carmofur in patients with curatively resected colon cancer. Jpn J Clin Oncol. 2005 Sep;35(9):536-44. PubMed PMID: 16155120.

Sakamoto J, Hamada C, Yoshida S, Kodaira S, Yasutomi M, Kato T, Oba K, Nakazato H, Saji S, Ohashi Y. An individual patient data meta-analysis of adjuvant therapy with uracil-tegafur (UFT) in patients with curatively resected rectal cancer. Br J Cancer. 2007 Apr 23;96(8):1170-7. PubMed PMID: 17375049; PubMed Central PMCID: PMC2360162.

Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol. 2004 Jul;57(7):683-97. PubMed PMID: 15358396.

Sieber J, Köberlein J, Mösges R. Sublingual immunotherapy in daily medical practice: effectiveness of different treatment schedules - IPD meta-analysis. Curr Med Res Opin. 2010 Apr;26(4):925-32. PubMed PMID: 20163297.

Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials. 2005;2(3):209-17. PubMed PMID: 16279144.

Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med. 2005 May 15;24(9):1307-19. PubMed PMID: 15685717.

Steinberg KK, Smith SJ, Stroup DF, Olkin I, Lee NC, Williamson GD, Thacker SB. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. Am J Epidemiol. 1997 May 15;145(10):917-25. PubMed PMID: 9149663.

Steiner TJ, Voelker M. Gastrointestinal tolerability of aspirin and the choice of over-the-counter analgesia for short-lasting acute pain. J Clin Pharm Ther. 2009 Apr;34(2):177-86. PubMed PMID: 19250138.

Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ. 1997 Sep 13;315(7109):640-5. PubMed PMID: 9310565; PubMed Central PMCID: PMC2127436.

Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data Cochrane Working Group. Stat Med. 1995 Oct 15;14(19):2057-79. PubMed PMID: 8552887.

Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference?. Lancet. 1993 Feb 13;341(8842):418-22. PubMed PMID: 8094183.

Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, Jordan J. The impact of patient-centered care on outcomes. J Fam Pract. 2000 Sep;49(9):796-804. PubMed PMID: 11032203.

Stewart LA, Tierney JF, Clarke M. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008] ed. Higgins J, Green S, editors. The Cochrane Collaboration: The Cochrane Collaboration; 2008. Chapter 18, Reviews of individual patient data.

Stewart M, Brown JB, Weston WW, McWhinney I, McWilliam C, Freeman T. Patient-centered medicine: transforming the clinical method. Thousand Oaks, Calif: Sage; 1995.

Strauss ME, Smith GT. Construct validity: advances in theory and methodology. Annu Rev Clin Psychol. 2009;5:1-25. PubMed PMID: 19086835; PubMed Central PMCID: PMC2739261.

Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use.. 4th ed. Oxford University Press: Oxford; 2008.

Tandon S, Tudur-Smith C, Riley RD, Boyd MT, Jones TM. A systematic review of p53 as a prognostic factor of survival in squamous cell carcinoma of the four main anatomical subsites of the head and neck. Cancer Epidemiol Biomarkers Prev. 2010 Feb;19(2):574-87. PubMed PMID: 20142252; PubMed Central PMCID: PMC3536470.

Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Stat Med. 2000 Dec 30;19(24):3417-32. PubMed PMID: 11122505.

Walker JK. Patient satisfaction: the past directs the way to the future. J Perinat Neonatal Nurs. 2006 Jan-Mar;20(1):88-90. PubMed PMID: 16508472.

Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. Stat Med. 2001 Aug 15;20(15):2243-60. PubMed PMID: 11468762.

Zhang L, Sampson M, McGowan J. Reporting of the role of the expert searcher in Cochrane Reviews. Evidence Based Library And Information Practice. 2006 ; 1:4.