Research article

WHO systematic review of maternal mortality and morbidity: methodological issues and challenges

A Metin Gülmezoglu*, Lale Say, Ana P Betrán, Jose Villar and Gilda Piaggio

Address: UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, Department of Reproductive Health and Research, World Health Organization, Geneva, (1211) Switzerland

Email: A Metin Gülmezoglu* - gulmezoglum@who.int; Lale Say - sayl@who.int; Ana P Betrán - betrana@who.int; Jose Villar - villarj@who.int; Gilda Piaggio - piaggiog@who.int

* Corresponding author

Abstract

Background: Reducing maternal mortality and morbidity are among the key international development goals. A prerequisite for monitoring the progress towards attainment of these goals is accurate assessment of the levels of mortality and morbidity. In order to contribute to mapping the global burden of reproductive ill-health, we are conducting a systematic review of incidence and prevalence of maternal mortality and morbidity.

Methods: We followed the standard methodology for systematic reviews. We prepared a protocol and a form for data extraction that identify key characteristics on study and reporting quality. An extensive search was conducted for the years 1997–2002 including electronic and hand searching.

Results: We screened the titles and abstracts of about 65,000 citations identified through 11 electronic databases as well as various other sources. Four thousand six hundred and twenty-six full-text reports were critically appraised and 2443 are included in the review so far. Approximately one third of the studies were conducted in Asia and Africa. The reporting quality was generally low with definitions for conditions and the diagnostic methods often not reported.

Conclusions: There are unique challenges and issues regarding the search, critical appraisal and summarizing epidemiological data in this systematic review of prevalence/incidence studies. More methodological studies and discussion to advance the field will be useful. Considerable efforts including leadership, consensus building and resources are required to improve the standards of monitoring burden of disease.

Background

Levels of maternal mortality and morbidity tell us about the risk attributable to pregnancy and childbirth as well as the performance of health systems in terms of access to health care and the quality of care provided. However, accurate assessment of these indicators has been problematic. The World Health Organization (WHO) has developed estimates of maternal mortality [1,2], anaemia during pregnancy, low birth weight and unsafe abortion at national, regional and global levels using modelling techniques. The lack of good quality data for many countries and different methodologies used to estimate levels of mortality complicate monitoring of the trends and comparisons between countries [3,4].

Although considerable amounts of facility-based data on maternal morbidity are generated, these may not reflect the actual health status of women in the whole community or area. Population-based data on the status of women's health are more useful and needed, yet scarce. Even when available, the challenge remains as to how to compile and summarize the data and thus map the burden of reproductive ill-health. A logical approach is to extend the concept of systematic reviews from randomised controlled trials to observational data [5-7].

For more than a decade, systematic reviews of randomised controlled trials (RCTs) have been used increasingly to evaluate the effectiveness of various health care interventions. The Cochrane Library as of 2004 includes more than 3000 systematic reviews [8]. Considerable experience of methodological issues such as literature search, critical appraisal of identified studies and methods for summarising data has been gained and tools have been developed for the reviews and meta-analysis of RCTs [8].

However, systematic reviews of observational studies are rather rare and the relevant experience is limited [6]. Most of the work in this area relates to questions for which RCTs are difficult, impossible or unethical to conduct (e.g. testing aetiological hypothesis, less common adverse effects in drugs) [9]. Methodological issues with regard to inclusion of studies with different designs, population and setting characteristics, and statistical methods to combine the data are evolving and need to be improved [6,10].

With these considerations, we are conducting a systematic review of prevalence/incidence of maternal mortality and morbidities from 1997 to 2002. The primary objective is to contribute to mapping the global burden of reproductive ill-health. The review will provide a comprehensive, standardized and reliable tabulation of available data on the incidence/prevalence of maternal morbidity and mortality, and case-fatality rates for maternal morbid conditions. The review will also assist us in identifying the most commonly used set of definitions for some pregnancy related conditions, in testing a set of critical appraisal and data-extraction instruments that can be used in future reviews of observational studies, and in guiding future research in this field.

We prepared a protocol [11] and a form for data extraction (See Additional file 1) which were both peer-reviewed. In this manuscript, we present our experience with the methodological, technical and practical challenges encountered in conducting the review.

Methods
Criteria for considering studies
Types of studies
For maternal morbidity, any study type providing prevalence, incidence or case-fatality rates for specified maternal morbid conditions is considered. These include mainly cross-sectional and cohort studies, clinical trials, and incidence/prevalence surveys. Case-control studies are included if the cases selected correspond to all cases in a given population where the denominator is also known. Intervention and control arms of controlled trials are treated separately.

For maternal mortality, studies providing estimates of maternal mortality levels derived from direct counting, or from special surveys are considered for inclusion. Estimates derived from modelling of other variables or extrapolations from other populations are excluded.

Types of participants
Women either pregnant or within one year of termination of pregnancy.

Types of outcomes
Maternal mortality and morbid conditions defined according to the International Classification of Diseases, 10th revision (ICD-10) [12].

Studies are ineligible if any of the following apply: (i) data collection dates are not reported, (ii) data are collected only before 1990, (iii) part of the data is collected before 1980 and disaggregation by year is not possible (in order to exclude data before 1990), (iv) number of study participants is less than 200 (this criterion imposed arbitrarily), (v) the study design is case-control and incidence/prevalence estimates from the defined population cannot be calculated, (vi) the methodology is not described.

Search strategy for identification of studies
We searched for published and unpublished studies reported between 1997 and 2002 in any language. The decision to start from 1997 was arbitrary and based on the concept of reviewing recent data.

The sources searched to identify studies included electronic databases (Medline, Popline, EMBASE, LILACS, CAB Abstracts, SocioFile, CINAHL, Econlit, BIOSIS, PAIS International, Index Medicus for the Eastern Mediterranean Region (EMRO) – on-line database of WHO/EMRO); web pages from Ministries of Health for official information and other potentially relevant internet sources (e.g. reproductive health gateway, development gateway, dissertation abstracts, Google). Additional file 2 includes detailed strategy for electronic search.

In addition, we checked reference lists of retrieved articles, proceedings and abstract books of related congresses. We hand searched journals at WHO headquarters' library that are not indexed in electronic databases and countries' statistical reports held at the WHO library. We contacted country focal experts such as WHO representatives and staff from collaborating centres, non-governmental organizations (NGOs), and other organizations known to be active in the field.

A WHO specialised librarian and the trial search coordinator of the Cochrane Collaboration Pregnancy and Childbirth Group developed the search strategy for each of the electronic databases according to their specific subject headings or searching structure in collaboration with the reviewers. We tested the search strategy for citations from 1997, modified the strategy and ran it for the whole period. We used Reference Manager® software [13] to keep track of the citations identified. We downloaded the citations identified in electronic searches into Reference Manager® and entered those retrieved from other sources manually (e.g. hand searching, reference lists). We deleted duplicates and assigned a unique identification number for each citation.

Screening and data-extraction form

Initially, we evaluated all identified citations on the basis of titles and/or abstracts against the eligibility criteria. Those deemed to be irrelevant were excluded and reasons for exclusion noted. A list of excluded reports and the reasons for exclusion are available from the authors upon request. When the information provided by titles/abstracts was insufficient to decide on inclusion/exclusion, or the titles/abstracts were relevant to the project, we retrieved and evaluated the full-text. As of January 2004, we screened titles/abstracts of a total of 64,586 citations from years 1997 to 2002. Among these, 59,960 were excluded and we retrieved full-text reports of the remaining 4626 (Figure 1).

We completed a specially designed screening form for each full-text evaluated report. This form was used to collect information on whether the report was included or not and if excluded, the reason for this. For reports meeting more than one exclusion criterion, only one reason – following the order on the screening form – was reported as the reason for exclusion. We extracted data from the included studies using a specifically designed data extraction form (See Additional file 1). This form includes 48 questions distributed in five modules. Modules were designed to collect information on (i) the general characteristics of the study such as design, population, setting, (ii) prevalence/incidence of maternal morbid conditions, (iii) maternal mortality, (iv) quality assessment of morbidity reports and (v) quality assessment of studies report-

ing maternal mortality. We also developed a manual for providing definitions and instructions on how to extract the data (available upon request from the authors). We tested both screening and data-extraction forms for a group of studies of different designs and revised prior to use.

Agreement between the reviewers in screening and data extraction

Two reviewers independently screened titles/abstracts from a sample of citations identified through the electronic search. In order to estimate the level of disagreement between two reviewers when including studies in the systematic review within 2.5% of the true value, a total of 560 studies needed to be classified. This sample size assumes a 95% confidence interval and that the level of disagreement between the two reviewers will not exceed 10% [14]. The percentage of agreement was 88.9% (95% CI 86.0% to 91.4%). The inter-observer agreement beyond chance was calculated using the Kappa statistics and found to be 0.60 (95% CI 0.52 to 0.69). This value corresponds to moderate to substantial agreement between the reviewers [15].

Two reviewers independently assessed full-texts of 50 articles for inclusion in the review following the initial screening process and completed the data extraction forms for those that were eligible for inclusion. The reviewers then compared results and disagreements were resolved following discussion. The decision on the number of this sample was agreed on arbitrarily at the beginning and deemed to be sufficient following the discussions on the completed forms.

Data processing

We categorised variables of interest and developed codes for responses to open-ended questions to facilitate data entry and statistical analysis. Initially, we classified morbidities according to the ICD-10 [12], using the classifications described mainly in chapter XV, which addresses pregnancy, childbirth and postpartum conditions. We assigned unique codes to some conditions (e.g. obstructed labour) that are classified with different codes in ICD-10 according to etiological distinctions. (See Additional file 3). These changes were made to facilitate the coding of the conditions during the data extraction process since definitions do not generally include etiological distinctions in incidence/prevalence studies of maternal conditions.

We preferred to extract raw data but where only percentages or rates were available, we also included these. Data presented in graphs and figures were used only if numbers (or percentages) were described in the text or labelled in the graph. Such data were not used if extrapolation was
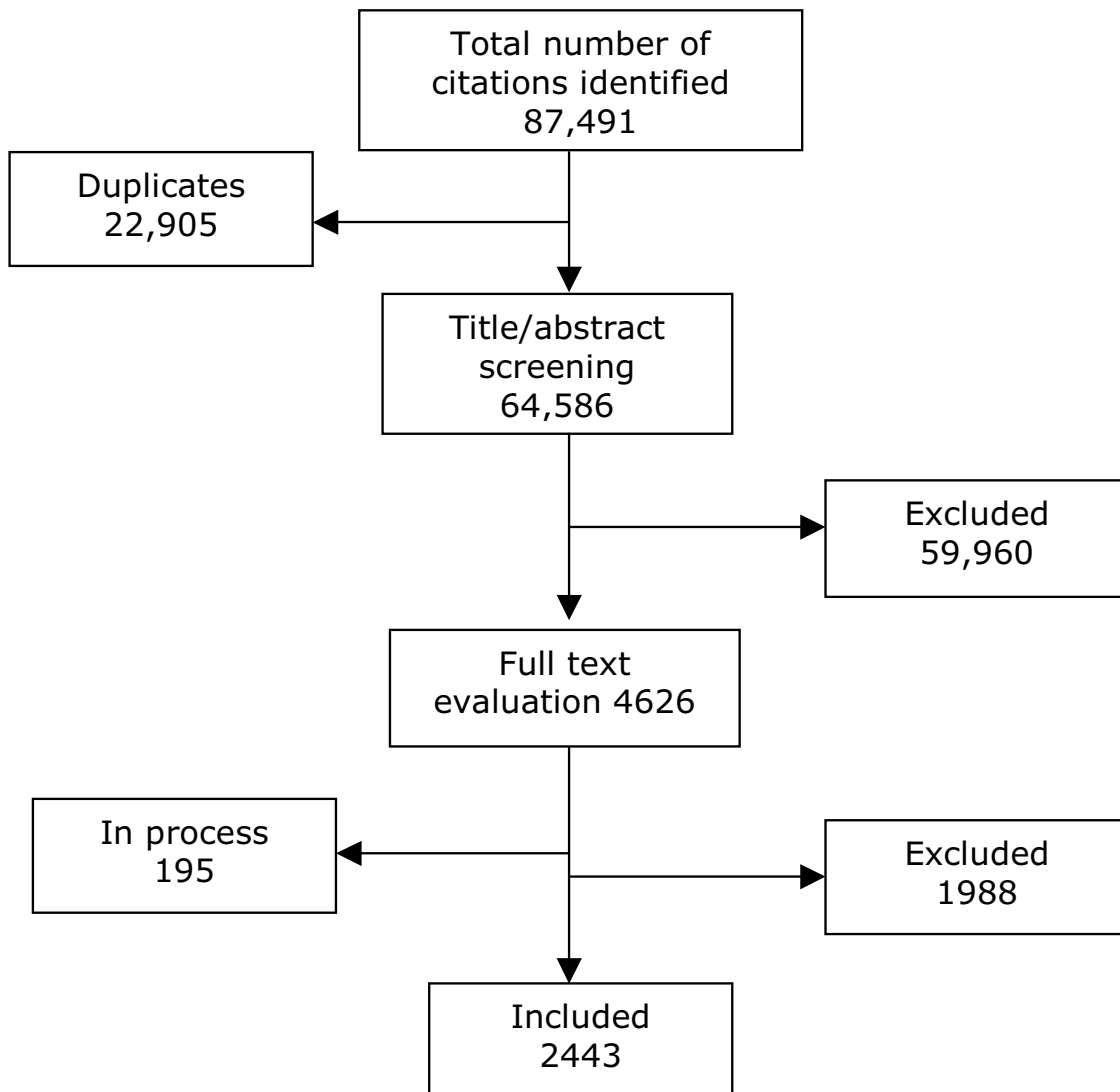
Figure 1
Flow diagram of the process of identifying and including references for the systematic review.

required. Once data extraction was completed, data were reviewed to identify duplicate data, for example the same results published in more than one journal or published papers whose unpublished drafts had been identified previously. Data were manually double entered in a specific database and processed with SAS software®.

Appraising methodological quality of primary studies
We excluded studies that did not state the methodology used to obtain data. For quality appraisal, we extracted information on (i) study design, sampling method,

sources of data and completeness of follow-up or records and, (ii) reported definitions and diagnostic procedures regarding outcome measures.

The evaluation of methodological and reporting quality was used to assess the reliability and accuracy of the data as objectively as possible. For example, the selection criteria and certain characteristics of participants such as economic status, ethnicity, age group or health status allow assessment of the external validity or generalisability of results in addition to presentation of stratified analysis for

different categories. Likewise, information on the proportions and characteristics of losses to follow-up, non-responders or those not included in the final analysis after having been initially selected for the study was used to assess the internal validity of a study.

Furthermore, we assumed that the presence of definitions of conditions and description of diagnostic methods or procedures could be regarded as an indication of higher quality. For studies which reported maternal mortality, in addition to categorising definitions of maternal mortality, we recorded information about special efforts to capture all maternal deaths and the method to confirm deaths as maternal (e.g. confidential enquiry, verbal autopsy) as indication of higher quality.

Results

We identified and screened about 65,000 different citations for the period 1997–2002. As of January 2004, 4626 citations were identified as potentially eligible for full-text evaluation, 2443 of which were included and 1988 excluded. The remaining 195 are in the process of retrieval and evaluation (Figure 1). Citations were mostly excluded because of no relevant data (57%), sample size less than 200 (16%) and no dates of data collection period (11%). The number of included studies for which data extraction and entry is complete is 2204. The distribution of these studies according to their designs, selected characteristics of the population and settings are summarised in Table 1. Most studies use a cross sectional design (69.5%). The study population is urban in 17% of the studies, rural in 6.7%, mixed in 43.6% and unknown in 32.7%. Nearly two-thirds of the data is facility based while most of the rest is either nationally or regionally representative.

Almost half of the studies are from Europe and North America and one-third are from Asia and Africa (Table 2). Similarly, half of the studies are conducted in 43 industrialised countries while the remainder are from 95 less developed and 46 least developed countries (Table 3).

The data were disaggregated by study periods, age groups, ethnic groups, settings and interventions used (i.e. different arms of RCTs) where possible and entered in the database as separate data sets. From 2204 included studies, we obtained 3805 data sets most of which include prevalence/incidence data on more than one morbidity. Morbidities of interest in our review were reported 5933 times in these data sets. Table 4 presents the distribution of reported morbidities and shows that hypertensive disorders of pregnancy and stillbirth were most frequently reported (14.9 % and 13.9%, respectively).

A preliminary assessment of the reporting quality of studies shows that the quality is generally low. For example,

Table 1: Selected characteristics of included studies.

	Number (%)
Study design	
Cross-sectional	1532 (69.5)
Cohort	274 (12.4)
Incidence/prevalence survey	220 (10.0)
Controlled trial	52 (2.4)
Census	22 (1.0)
Others	104 (4.7)
Population characteristics	
Urban	375 (17.0)
Rural	147 (6.7)
Mixed	962 (43.6)
Unknown	720 (32.7)
Characteristics of the setting	
National	414 (18.8)
Provincial	288 (13.1)
City-based	93 (4.2)
Other	39 (1.8)
Health facility-based	1370 (62.1)
Overall	**2204**

for hypertensive disorders, about 50% included definitions and only 10% described the diagnostic procedure. More than half of the studies of maternal mortality did not report the definition used for maternal death, and two-thirds did not use any method to confirm the death as maternal. Similarly, less than 20% attempted to capture all maternal deaths among the population studied.

Discussion

In this paper, we present our initial experience with conducting this large-scale systematic review of observational studies. We discuss methodological challenges as well as barriers encountered at both technical and logistic levels. We present preliminary results on the descriptive characteristics of the data set and expect to generate more discussion and empirical research in this area.

One of the main strengths of this systematic review is the comprehensive search strategy including multiple electronic databases. This search strategy yielded a large number of disparate records. This is partly due to the fact that searching according to study type is possible only for controlled trials. Initial screening by titles and abstracts to select relevant studies reduced the number of potentially relevant reports to a reasonably manageable level. However, it was not always straightforward to judge relevance from abstracts and this has been a tedious and time-consuming process.

Identification of duplicate records has been another time-consuming activity. Different databases use different formats for indexing titles and/or authors. For example, authors of articles written in Spanish tend to present two

Table 2: Regional distribution of included studies and reports of maternal morbidity and mortality [24].

Regions	N reporting only morbidity	N reporting only mortality	N reporting morbidity and mortality	Total
Europe	482	39	34	555
Asia	374	68	30	472
North America	405	28	4	437
Africa	201	102	70	373
Latin America and the Caribbean	156	85	23	264
Australia/Oceania	60	5	1	66
Other (e.g. multicountry)	26	6	5	37
Overall	**1704**	**333**	**167**	**2204**

Table 3: Distribution of included studies according to the development status of the country where the study was conducted [24].

Country type	N reporting only morbidity	N reporting only mortality	N reporting morbidity and mortality	Total
Industrialised (N = 43)	971	75	39	1085
Less developed (N = 95)	584	183	87	854
Least developed (N = 46)	123	69	36	228
Other (e.g. multicountry)	26	6	5	37
Overall	**1704**	**333**	**167**	**2204**

surnames. The first surname may not be recognised correctly or surname and first name are not always differentiated [16,17]. This lack of uniformity in formatting of citations across databases causes several different recordings of the same article if indexed in more than one database. We experienced an exhaustive process of manual searching and deleting of duplicate records to address this issue.

Although efforts to develop methodologies for searching for and summarising data from observational studies exist, these focus largely on effects of health care interventions that are difficult or impossible to evaluate through RCTs [18]. The associations between risk factors and disease occurrence, evaluation of diagnostic and screening tests or prognostic variables are other areas that require reviews of observational studies and pose particular methodological challenges [9].

The specific issues relating to analysis of data from systematic reviews of incidence/prevalence studies need to be systematically explored in order to guide developments in this field. Evaluating the comparability in terms of design, population and setting, and summarizing results pose specific methodological challenges.

A great deal of variation in the incidence/prevalence of maternal conditions between studies has been shown to be related to variations in definitions [19-21]. In addition, for many conditions studied, we identified that a variety

of diagnostic tests with different levels of precision were used. We extracted detailed information on such characteristics of the studies in order to explore the contribution of these factors to the heterogeneity of the results. We envisage providing a tabulation of the most commonly used definitions and diagnostic procedures for each condition. In a second step, we will examine why the 'most recommended' or 'official' definition or diagnostic procedure is or is not used and how this affects the outcomes. This could provide a background for initiatives to standardize the definitions and improve the accuracy of measurements.

Another issue of concern is the general poor reporting quality of the studies. Characteristics of the populations and/or settings, definitions of outcomes and diagnostic procedures are not systematically described in the reports. This could limit the comparability of the studies and utility of some data extracted.

An important challenge for systematic reviews on burden of disease is to identify all available data from less developed countries [22]. Studies from these countries are likely to be published in non-indexed and non-English journals. The amount of research conducted in those countries may also be less considering the difficulties of securing funds for research. Nevertheless, we think that the final data set will include substantial amount of data from developing countries.

Table 4: Maternal morbidities most frequently reported in included studies

Morbidity	N (%)
Hypertensive disorders of pregnancy	885 (14.9)
Stillbirth	828 (13.9)
Preterm delivery	489 (8.2)
Induced abortion	400 (6.7)
Haemorrhage (antepartum, intrapartum, postpartum, unspecified	365 (6.2)
Anaemia	267 (4.5)
Placenta anomalies (praevia, abruptio, etc.)	245 (4.1)
Spontaneous abortion	235 (4.0)
Gestational diabetes	224 (3.8)
Ectopic pregnancy	146 (2.5)
Premature rupture of membranes	140 (2.4)
Perineal laceration	139 (2.3)
Uterine rupture	116 (2.0)
Obstructed labour	102 (1.7)
Depression (postpartum, during pregnancy)	96 (1.6)
Puerperal infection	86 (1.5)
Violence during pregnancy	77 (1.3)
Urinary tract infections	66 (1.1)
Malaria	54 (0.9)
Other conditions	973 (16.4)
Overall	**5933**

The first decade of this new millennium will be a test of our capability and ability to cope with the ever-increasing amounts of information produced. To analyse all available information in a reliable way, we need systematic reviews that include comprehensive searches, critical evaluation of studies and advances in statistical and other (searching, appraisal) methodologies. Considering the fact that it took almost 20 years to achieve satisfactory standards of synthesising the research into the effects of health care interventions, it is clear from our initial work that similar efforts, including leadership, consensus building and resources are required to improve the standards of monitoring the burden of disease.

The conceptualisation of this review, completion of a protocol and actual conduct has taken three years with several staff members involved. Full results will be available at the end of 2004. Although complex and time-consuming this systematic review and others on this issue will contribute to the Millennium Project [23] in several ways. Firstly, quantifying the burden not only through modelling approaches but also through in depth analyses of empirical studies will improve our understanding of the magnitude of the problem. Secondly, by identifying the gaps in the methodology and reporting, future research could be designed more rigorously.

Competing interest
None declared.

Authors' contributions
AMG and JV initiated the project. AMG, LS, APB and JV coordinated the review. APB, LS, AMG and JV prepared the manuscript. LS, APB and AMG developed the search strategy. LS and APB performed the title/abstract screening of the records. LS and APB, extracted data. GP provided statistical expertise.

Additional material

Additional File 1
Data-extraction forms used in the systematic review.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2288-4-16-S1.pdf]

Additional File 2
Final search strategy for each of the searched electronic databases.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2288-4-16-S2.doc]

Additional File 3
ICD-10 codes and those used in the systematic review for maternal mortality and morbidity conditions.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2288-4-16-S3.doc]

Acknowledgements

References

1. **Revised 1990 estimates of maternal mortality. A new approach by WHO and UNICEF.** *Volume WHO/FRH/MSM/96.11.* Geneva, Switzerland, World Health Organization; 1996.
2. **Maternal mortality in 1995: Estimates developed by WHO, UNICEF and UNFPA.** Geneva, Switzerland, World Health Organization; 2001.
3. Hill K, AbouZahr C, Wardlaw T: **Estimates of maternal mortality for 1995.** *Bulletin of the World Health Organization* 2001, **79:**182-193.
4. Hakkert R: **Country estimates for maternal mortality: an alternative model.** *Statistics in medicine* 2001, **20:**3505-3524.
5. Villar J, Betrán AP, Gülmezoglu AM, Say L: **WHO leads global effort on systematic reviews.** *International Journal of Epidemiology* 2003, **32:**164-165.
6. Dickersin K: **Systematic reviews in epidemiology: why are we so far behind?** *International Journal of Epidemiology* 2002, **31:**6-12.
7. Petticrew M: **Why certain systematic reviews reach uncertain conclusions.** *BMJ* 2003, **326:**756-758.
8. **The Cochrane Database of Systematic Reviews** 2004 [http://www.cochrane.org/reviews/clibintro.htm].
9. Egger M, Smith GD, Altman DG: *Systematic reviews in health care. Meta-analysis in context.* London, BMJ Publishing Group; 2001.
10. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB: **Meta-analysis of observational studies in epidemiology. A proposal for reporting.** *JAMA* 2000, **283:**2008-2012.
11. **WHO systematic review on maternal morbidity and mortality 1997-2002. Study protocol.** *World Health Organization, Geneva, Switzerland* 2002.
12. *International Classification of Diseases and Related Health Problems - 10th revision* Geneva, Switzerland, World Health Organization; 1992.
13. Researchsoft Thomson - ISI: **Reference Manager.** 2004 [http://www.refman.com/].
14. Machin D, Campbell MJ, Fayers P, Pinol A: *Statistical Tables for the Design of Clinical Trials Volume Table 11.1.* 2nd edition edition. Oxford, Blackwell Scientific Publications; 1997.
15. Landis JR, Koch GG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33:**159-174.
16. Fernández E, García AM: **Accuracy of referencing of Spanish names in Medline.** *Lancet* 2003, **361:**351-352.
17. Siebers R, Holt S: **Accuracy of references in five leading medical journals.** *Lancet* 2000, **356:**1445.
18. **The Cochrane Non-Randomised Studies Methods Group (NRSMG)** 2004 [http://www.cochrane.dk/nrsmg/].
19. Higgins JR, de Swiet M: **Blood-pressure measurement and classification in pregnancy.** *Lancet* 2001, **357:**131-135.
20. Ronsmans C, Campbell O, Collumbien M: **Effect of supplementation with vitamin A or beta carotene on mortality related to pregnancy. Slight modifications in definitions could alter interpretation of results.** *BMJ* 1999, **319:**1202-1203.
21. North RA, Taylor RS, Schellenberg JC: **Evaluation of a definition of pre-eclampsia.** *British Journal of Obstetrics & Gynaecology* 1999, **106:**767-773.
22. Smith R: **Publishing research from developing countries.** *Statistics in medicine* 2002, **21:**2869-2877.
23. World Bank: **Millennium Development Goals.** [http://www.developmentgoals.org/].
24. United Nations: **World Population Prospects. The 2000 revision.** New York, United Nations; 2001.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/4/16/prepub