
INVESTIGACIÓN ORIGINAL

Development of the Metareview Assessment of Reporting Quality (MARQ) Checklist

Desarrollo de la Lista de verificación de la Evaluación de los informes de calidad de las Metarevisiones (MARQ)

Jay P. Singh

Submitted: 25/11/2012 / **Accepted:** 15/12/2012

| [Summary](#) |

Background. Systematic reviews and meta-analysis often come to conflicting conclusions on key issues and have a number of potentially important methodological limitations. A metareview represents one approach to a descriptive investigation of such issues in review literatures; this involves a systematic review of previously published reviews. Metareviews report on the areas that systematic reviews and meta-analyses have covered, investigating the methodological quality of such reviews, comparing methods for reporting results with recommended standards in the field of systematic reviewing and highlighting areas which could benefit from further research.

Objective. The present report was aimed at critically examining the reporting quality of available medical metareviews and encouraging the use of such innovative approach to develop an instrument for assessing metareviews' methodological quality.

Materials and methods. PsycINFO, MEDLINE, EMBASE, and CINAHL were searched for previous medical metareviews as of February 11th, 2012. References regarding identified reports and annotated bibliographies were used to supplement the search.

Results. Four metareviews meeting the inclusion criteria were identified and descriptively analysed. The first set of standardised metareview reporting guidelines' checklist (metareview assessment of reporting quality - MARQ), using

quality checklists developed for primary studies and reviews as models, was introduced to enable transparent and consistent reporting of metareview methodology. An average of 15 (SD = 3) MARQ criteria were met when applied to the four metareviews identified during the systematic search. This indicated a moderate level of reporting quality which should be improved in subsequent applications of the methodology by using the standardised checklist. A high level of inter-rater agreement was found ($\kappa = 0.93$).

Conclusion. The standardised set of guidelines outlined in this report should assist future researchers in conducting more transparent and methodologically rigorous metareviews.

Key words: review, meta-analysis (MeSH).

Singh JP. Development of the Metareview Assessment of Reporting Quality (MARQ) Checklist. Rev. Fac. Med. 2012; 60: 325-332.

| [Resumen](#) |

Antecedentes. Las revisiones sistemáticas y metaanálisis a menudo llegan a conclusiones contradictorias sobre cuestiones fundamentales y tienen una serie de limitaciones metodológicas potencialmente importantes. Una metarevisión representa una aproximación a una investigación descriptiva de estos temas en la literatura de revisión, lo que implica una revisión sistemática de las revisiones publicadas anteriormente. Las metarevisiones informan sobre las áreas que las revisiones sistemáticas y metaanálisis han cubierto, investigando la calidad metodológica de dichos estudios, y comparando los métodos para informar resultados con las normas recomendadas en el campo de las revisiones sistemáticas y poner de relieve áreas

J. P. Singh.

Department of Mental Health Law and Policy, University of South Florida, 13301 Bruce B Downs Blvd, MHC 2615, Tampa, FL, 33612, USA. Institute of Health Sciences, Molde University College, Molde, Norway

Correspondence: jaysingh@usf.edu

que se podrían beneficiar de futuras investigaciones.

Objetivo. Examinar críticamente la calidad de la información disponible de las metarevisiones médicas y fomentar el uso de tal enfoque innovador para desarrollar un instrumento que evalúe la calidad metodológica de las metarevisiones.

Material y métodos. En las bases de datos PsycINFO, MEDLINE, EMBASE, y CINAHL se buscaron metarevisiones médicas hasta el 11 de febrero de 2012. Para complementar la búsqueda fueron usadas referencias relativas a los informes identificados y las bibliografías anotadas.

Resultados. Cuatro metarevisiones cumplieron los criterios de inclusión las cuales fueron identificadas y analizadas de forma descriptiva. El primer conjunto de directrices de la lista de verificación estandarizados metareview (Lista de verificación de la Evaluación de los informes de calidad de las Metarevisiones (MARQ)), uso listas de control de calidad desarrollados por los estudios primarios y revisiones como modelos, se introdujo para permitir un registro transparente y coherente de la metodología de las metarevisiones. Un promedio de 15 (SD=3) MARQ criterios se cumplieron cuando se aplicaron a las cuatro metarevisiones identificadas durante la búsqueda sistemática. Esto indica un nivel moderado de calidad de los informes que deben ser mejorados en las aplicaciones posteriores de la metodología utilizando la lista de verificación estandarizada. Se encontró un alto nivel de acuerdo interevaluadores ($\kappa=0,93$).

Conclusión. El conjunto estandarizado de directrices que se describen en este informe debería ayudar a los futuros investigadores la realización de metarevisiones más transparentes y rigurosas metodológicamente.

Palabras clave: revisión, metanálisis (DeCS).

.....
Singh JP. Desarrollo de la Evaluación de las Metareview y Reportes de la Calidad (MARQ) Lista de Chequeo. Rev. Fac. Med. 2012; 60: 325-332.

Background

As the number of primary studies in a literature grows, reviews of the field will be published. Reviews are helpful in that they allow large quantities of information to be quickly assimilated by readers, be they researchers, clinicians, policymakers, or non-professionals (1). The contemporary review literature contains three kinds of reviews, each with a different methodology and a unique set of strengths and weaknesses: narrative reviews, systematic reviews, and meta-analyses. With some fields having had dozens or even hundreds of reviews published (2), a fourth type of review has recently emerged, that of a metareview, or a “review of reviews”. Each of these four approaches to reviewing has a different methodology and

a unique set of strengths and weaknesses. Understanding these differences can clarify the reliability and credibility of different reviews for decision-makers, who often use reviews as an empirical base to influence practice and policy (3).

Narrative Reviews

Narrative reviews summarize the available literature on a given topic from the theoretical and experiential perspective of the reviewer (4). A primary strength of narrative reviews is that they may cover a broad variety of issues concerning a particular subject. However, narrative reviews may be strongly influenced by the viewpoint of their authors, as reviewers may take sides on a controversial issue. If a particular piece of research does not support the authors’ viewpoint, they may choose to exclude it rather than present it and appraise its validity (or lack thereof). Therefore, a weakness of narrative reviews is that they can be subjective representations of the literature on a topic. To obtain an objective overview of the available literature, all works identified using a systematic search that meet a set of pre-specified inclusion and exclusion criteria must be included. Without taking such a transparent and systematic approach, a review is not considered reproducible (4).

Systematic Reviews without Meta-analysis

Using a systematic search strategy and pre-defined inclusion and exclusion criteria to identify eligible studies, systematic reviews address the potential selection biases of narrative reviews. As reproducible systematic searches are used, readers of systematic reviews may be confident that a representative sample of work on a given topic has been included. Systematic reviews allow researchers to evaluate the consistency of results from primary studies. If consistent findings are reported by multiple studies, it strengthens these findings’ credibility. If inconsistent findings are discovered, the reviewer can explore descriptively why such discrepancies occurred. In addition to identifying patterns in the literature, systematic reviews also allow researchers to identify gaps that future research may address. If statistical, clinical, or methodological heterogeneity between studies is high (5), then quantitatively combining the results of primary studies may bias summary effect estimates and could be inappropriate (6). In such cases, systematic reviewing would be the methodology of choice. However, if studies are sufficiently similar and results are not quantitatively combined, systematic reviews will be limited by their inability to either calculate summary effect estimates or statistically investigate sources of heterogeneity.

Meta-analyses

Meta-analytic methodology maintains the strengths of systematic reviews while allowing for the statistical combination of primary study results. Researchers conducting meta-analyses use systematic searches and apply pre-specified inclusion and exclusion criteria to identify studies of interest. Provided that the studies are sufficiently similar (either in theory or using a statistical test of heterogeneity), effect sizes, tabular data, or individual participant data from the identified studies is then collected and quantitatively synthesized. In addition to calculating summary effect estimates, meta-analyses also allow researchers to statistically investigate the influence of demographic factors (e.g., participant age) and study design characteristics (e.g., length of follow-up) on effect size. Potential weaknesses of meta-analytic methodology include the combination of studies that measure different outcomes in different populations (i.e., the “apples and oranges” problem (7)), the combination of studies of varying quality (i.e., the “garbage in, garbage out” problem (7)), and the analysis of a non-representative group of studies due to publication bias (i.e., the “file drawer problem” (8)). These problems are often present in meta-analyses of observational studies but may be less problematic in reviews of randomized controlled trials (RCT), particularly in relation to combining high and low quality studies.

The reporting quality of studies included in systematic reviews and meta-analyses can be measured using standardized checklists such as the Standards for Reporting of Diagnostic Accuracy Studies (STARD) Statement (9), the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) Statement (10), the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (11), the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) Statement (12), or the Consolidated Standards of Reporting Trials (CONSORT) Statement (13). Further, the quality of systematic reviews and meta-analyses, themselves, can be assessed using published checklists such as the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Statement (14), the Quality of Reporting of Meta-analyses (QUOROM) Statement (15), or the Meta-analysis of Observational Studies in Epidemiology (MOOSE) Statement (16).

Metareviews

To examine the consistency of review findings and to descriptively compare their methods of reporting results with recommended standards in the field of systematic reviewing, one novel approach is to conduct a metareview, a systematic review of systematic reviews and meta-analyses. (Narrative

reviews are not included, as they are especially vulnerable to selection bias (17). In his recent editorial on the applications of metareview, Delgado-Rodriguez (18) identified four major strengths of this approach: First, metareviews allow researchers to investigate the general quality of the review literature on a given topic. This may be accomplished using a standardized quality reporting checklist. Second, metareviews can advance researchers’ understanding of heterogeneity. If consistent evidence of between-study variability is found, a metareview can suggest that sources of heterogeneity be investigated using methods such as subgroup analysis or meta-regression regardless of the statistical significance of a review’s findings (19). Third, by using the review as the unit of analysis, the consistency of publication bias findings can be explored. As reviews may operationalize the term “publication” differently (e.g., published in a peer-reviewed journal versus not, or available to the public in published form [journal article or government report] versus not), metareviews can qualify findings of publication bias. Finally, metareviews can identify which outcome statistics are commonly used to summarize study findings and describe their strengths and weaknesses. Thus, metareviews can be used to investigate key benchmarks of review quality.

Being a primarily descriptive methodology, metareviews share some of the limitations of systematic reviews, in that summary effect estimates cannot be calculated and sources of between-review heterogeneity cannot be statistically investigated. There is no accepted method of quantitatively synthesizing the findings of meta-analyses, a limitation inherent to the inclusion of overlapping studies by multiple reviews. Relatedly, metareviews are limited in that it can be difficult to directly compare review findings, as reviews may have identified studies using systematic searches with different inclusion and exclusion criteria. In addition, reviews may have explored treatment efficacy or instrument utility in different populations using different outcome measures. Due to these potential limitations, metareviews may best be viewed as investigations into the overall quality of review literatures and as thematic analyses identifying key uncertainties that warrant further exploration.

Aims

The objective of the present report was to systematically review metareviews of the medical sciences literature. To encourage the use of this innovative approach, the first standardized reporting quality checklist will be introduced to promote a transparent and consistent reporting of metareview methodology.

Methods

PsycINFO, MEDLINE, EMBASE, CINAHL, and ProQuest were used to identify published or unpublished metareviews of the medical sciences literature as of February 11, 2012 using the keyword metareview. References of identified reports and annotated bibliographies were used to supplement the search. Using this search strategy and excluding duplicates, narrative and systematic reviews, meta-analyses, and records that were not concerned with a medical field, four records remained.

Results and Discussion

The systematic search identified four published metareviews of the medical sciences literature as of February 11, 2012. Topics covered by these reports included: (1) the short-term effectiveness and safety of antidepressants for treating depression (20); (2) interventions in key areas of liaison psychiatry (21); (3) the epidemiology and reporting characteristics of systematic reviews in the field of medicine (22); and (4) the utility of violence risk assessment in forensic psychiatry (17).

Cipriani et al., 2007

Cipriani and colleagues (20) conducted a metareview of the short-term effectiveness and safety of antidepressants used in the acute phase treatment of major depression. The researchers systematically searched eight online databases to identify reviews of short-term pharmacologic interventions for depression that used antidepressants as part of treatment. Only reviews of randomized controlled trials were included. The metareview identified 1 relevant systematic review and 11 meta-analyses. While the reviews provided consistent evidence that antidepressants are effective in treating major depression in primary care settings, the authors concluded that there remains considerable uncertainty concerning the health-related effects of such medication. For example, maternal selective serotonin reuptake inhibitor (SSRI) usage during breast-feeding was not found to have any negative effects on infants, though there was some evidence of a relationship between maternal SSRI usage and pregnancy-related complications. The authors discussed the limitations of previous reviews of the antidepressant literature and suggested areas in need of further research.

Ruddy & House, 2005

Ruddy and House (21) conducted a meta-review to investigate interventions for clinical problems likely to be treated by

liaison psychiatric services. The authors systematically searched six databases for systematic reviews and meta-analyses concerning interventions designed to treat psychological problems resulting from a physical illness, somatoform disorders, or self-harming behavior. Using pre-specified inclusion and exclusion criteria, 51 relevant systematic reviews and 14 relevant meta-analyses were identified. The researchers appraised the quality of the collected reviews using a checklist published by Oxman and Guyatt (23). This checklist assesses whether reviewers clearly stated their aims and methods, conducted a systematic search, specified their inclusion and exclusion criteria, assessed the validity of the included primary studies in a way that was free from bias, ensured that primary studies were quantitatively synthesized in an appropriate manner, and discussed inconsistent findings (23). The inter-rater reliability of ratings of review quality was assessed using a random sample of the included reviews, which were rated by two of the metareview's authors.

The metareview concluded that much of the clinical practice of liaison psychiatry is based on low-quality research evidence. Further, reviews often came to conflicting conclusions about which form of treatment is most effective for different problems. The authors discussed the limitations of the identified reviews and the need for more service-oriented research focusing on common problem areas in clinical practice.

Moher et al., 2007

Moher and colleagues (22) conducted a metareview investigating the quality of systematic reviews and meta-analyses of the medical literature. The aim of the metareview was to identify a cross-sectional sample of recently published reviews and to examine their epidemiological, descriptive, and reporting characteristics. The authors systematically searched for reviews that were published in English and indexed on the electronic database, MEDLINE, in November of 2004. Using pre-specified inclusion and exclusion criteria, 139 systematic reviews and 161 meta-analyses were identified. Information was extracted on methodological characteristics of these reviews. In addition, the authors described the journals in which published reviews appeared (e.g., impact factors). A random sample of the reviews was chosen and coded by a second examiner to test inter-rater reliability.

The metareview concluded that the quality and consistency of systematic reviews of the medical literature vary considerably. Review findings were often conflicting, even when the same literature was reviewed. Over one-third of the included reviews did not include replicable

search strategies, did not exclude duplicate studies or overlapping samples, did not investigate sources of heterogeneity, and did not assess evidence of publication bias. The authors also discussed the limitations of the included reviews and noted a series of high quality reviews that could be used as models for the field.

Singh & Fazel, 2010

Singh and Fazel (2010) conducted the first summary overview of systematic reviews and meta-analyses of the field of forensic risk assessment. The aim of the metareview was to explore the methodological quality of previous reviews and to descriptively analyze their findings in order to identify key uncertainties regarding the prediction of future criminal behavior. Epidemiological, descriptive, and reporting characteristics were extracted from 9 systematic reviews and 31 meta-analyses that were identified through a systematic search. The quality of the included reviews was investigated using the PRISMA Statement, a 27-item checklist of review characteristics designed to enable a transparent and consistent reporting of results. The inter-rater reliability of the data extraction was established using a random sample of the included reviews.

The methodological quality of the identified systematic reviews and meta-analyses was generally poor. The average review met only two-thirds of PRISMA criteria with few reviews reporting a replicable search strategy, approximately half of the reviews not excluding overlapping samples or investigating sources of clinical or methodological heterogeneity, and a third of the reviews not assessing publication bias. Further, the reviews reported a narrow range of effect sizes and often included a mixture of both commonly and uncommonly used risk assessment tools, making it difficult to draw conclusions about the general utility of those measures with the greatest clinical impact. The metareview also found that previous reviews of the forensic risk assessment literature have come to conflicting conclusions on a number of issues, including the comparative predictive validity of individual risk assessment tools, the efficacy of actuarial instruments versus structured clinical judgment, the influence of demographic factors and study design characteristics on predictive validity, and the relative strength of association of individual risk factors for recidivism. Finally, the metareview concluded that the most commonly cited systematic reviews of the field were between 8 to 14 years old, suggesting that clinicians and policymakers' views

of risk assessment may be based on outdated literature. This finding highlighted the responsibility of those who publish influential systematic reviews and meta-analyses to publish updates.

Development of the MARQ Checklist

The main benefit of metareview methodology is its usefulness in conducting thematic analyses to identify major uncertainties that warrant further exploration. By comparing reviews' reporting characteristics against standardized quality checklists such as the PRISMA Statement, metareviews also provide fields with high standards that future systematic reviews and meta-analyses can follow. According to a recently developed checklist designed by expert systematic reviewers (24) as well as guidelines published by the Cochrane Collaboration (25), literature reviews should assess, at a minimum, whether included reviews state their objectives a priori, report a reproducible search strategy, include unpublished studies, provide a list of included (and, if applicable, excluded) studies, summarize the sample and design characteristics of included studies, conduct an inter-rater reliability check to assess the consistency of the data extraction process, investigate sources of heterogeneity, assess evidence of publication bias, and disclose conflicts of interest. As part of this process, metareviews should attempt to develop a standard list of clinical and methodological covariates that subsequent reviews can investigate as potential sources of between-study heterogeneity (5).

To encourage a transparent and consistent reporting of metareview methodology, a 20-item set of guidelines—the Metareview Assessment of Reporting Quality (MARQ) Checklist (Table 1)—was developed using previously published checklists as models. The use of similar reporting checklists are recommended at both the primary study (26) and review (27) level by over 200 peer-reviewed journals, suggesting that researchers are familiar with and journal editors are supportive of such guidelines. When applied to the four metareviews identified during the systematic search, an average of 15 (SD = 3) MARQ criteria were met, indicating a moderate level of reporting quality that may be improved in subsequent applications of the methodology by using the standardized checklist. In order to test inter-rater reliability, a research assistant working independently of the author (C.L.) coded each of the four metareviews on the 20 MARQ criteria. A high level of inter-rate agreement was established ($\kappa = 0.93$) [28], supporting the use of the checklist in future research.

Table 1. Metareview Assessment of Reporting Quality (MARQ) Checklist

Item	Criterion	Reported?	Page?
TITLE			
1. Title	Identify the investigation as a metareview.		
ABSTRACT			
2. Structured abstract	Include a structured abstract with the following information: objectives, methods (sources of reviews and brief description of systematic search), results, and conclusion.		
INTRODUCTION			
3. Reasoning	Discuss the reason for the metareview (e.g., conflicting findings of previous reviews or questionable methodology).		
4. Aims	State the purpose of the metareview (e.g., to identify uncertainties and methodological characteristics of a specific review literature).		
METHODS			
5. Source of reviews	List sources used to identify reviews for potential inclusion (e.g., electronic databases, annotated bibliographies, references) and the dates searched.		
6. Systematic search	Provide keywords used to search electronic databases, including limits used.		
7. Review screening	Report the inclusion and exclusion criteria applied to screen reviews (e.g., reviews concerning specific population or intervention).		
8. Data extraction	Describe how information was extracted from reviews (e.g., author-developed coding sheets).		
9. Quality checklist	Indicate whether a standardized checklist of reporting characteristics was used to assess within-review quality and where it can be accessed.		
10. Inter-rater reliability	State whether the inter-rater reliability of review characteristics extracted from reports was measured and present an effect size (e.g., kappa coefficient or intraclass correlation coefficient) with interpretation.		
11. Thematic analysis	Specify any descriptive thematic analyses conducted to investigate the consistency of findings across reviews addressing similar areas.		
RESULTS			
12. Review selection	Report the number of reviews identified prior to screening, assessed for eligibility and included in the metareview, providing a rationale for exclusions.		
13. Review characteristics	Present review epidemiological characteristics (e.g., impact factors, dates published), descriptive characteristics (e.g., diagnostic instruments used, populations investigated), and reporting characteristics (e.g., significant heterogeneity or publication bias found, effect sizes reported) for which data were extracted.		
14. Quality checklist	Provide the average number of reporting characteristics met by the included reviews (with standard deviation) along with the most and least frequently met criteria.		
15. Thematic analysis	Report themes identified across reviews and present, for each, a description of the relevant reviews and their findings.		
DISCUSSION			
16. Summary	Summarize the main themes/uncertainties identified as well as methodological weaknesses that future systematic reviews and meta-analyses could address.		
17. Implications	Discuss the relevance of findings to key professional groups (e.g., researchers, clinicians, and policymakers).		
18. Limitations	Discuss limitations of the employed methodology (e.g., overly- or under-sensitive search strategy, forced exclusion of unobtainable reviews).		
ROLE OF EXTERNAL SOURCES			
19. Conflicts of interest	Describe any conflicts of interest (e.g., source of funding and role of funders).		
20. Acknowledgements	Acknowledge assistance of review authors in obtaining additional information.		

Conclusion

Systematic reviews and meta-analyses often come to conflicting conclusions on key issues and have a number of potentially important methodological limitations, suggesting that their findings should be considered as provisional. To examine discrepancies in the findings and methodological quality of reviews, one novel approach is a metareview, a systematic review of reviews. While standardized checklists have been developed to provide reporting guidelines for primary studies, systematic reviews, and meta-analyses, no such standards have been designed for metareviews. The highly reliable MARQ Checklist introduced in the present report provides researchers with a potentially useful set of criteria that, if followed, will result in transparent and consistent reporting of metareview methodology.

Competing Interests

The author declares no competing interests.

Acknowledgements

The author thanks Christie Leung for her assistance with the inter-rater reliability check.

Several paragraphs in the Background section of the present manuscript were published by the author in Elena Grigorenko's (Ed.) *Handbook of Juvenile Forensic Psychology and Psychiatry* (2012) in the chapter, "The History, Development, and Testing of Forensic Risk Assessment Tools" (pp. 217-240). With permission from Springer Science+Business Media B.V., these paragraphs have been adapted to the present manuscript.

References

1. **Higgins J, Green S.** *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.6. Chichester: John Wiley & Sons; 2006.
2. **Hattie J: Visible Learning.** A Synthesis of over 800 Meta-analyses Relating to Achievement. Oxford: Routledge; 2009.
3. **Gendreau P, Goggin C, Smith P.** *Cumulating Knowledge: How Meta-analysis Can Serve the Needs of Correctional Clinicians and Policymakers.* Ottawa: Correctional Service of Canada; 2000.
4. **Cook DJ, Mulrow CD, Haynes RB.** Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997, 126:376-380.
5. **West SL, Gartlehner G, Mansfield AJ, Poole C, Tant E, Lenfestey N, Lux LJ, Amoozegar J, Morton SC, Carey TC, Viswanathan M, Lohr KN.** *Comparative Effectiveness Review Methods: Clinical Heterogeneity.* Research Triangle Park: Agency for Healthcare Research and Quality; 2010.
6. **Deeks J, Higgins J, Altman D.** Analysing and presenting results. In *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.6. Edited by Higgins J, Green S. Chichester: John Wiley & Sons; 2006.
7. **Sharpe D.** Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *ClinPsychol Rev.* 1997, 17:881-901.
8. **Rosenthal R.** The file drawer problem and tolerance for null results. *Psychol Bull* 1979, 86:638-641.
9. **Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glaziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Wet HCW, the STARD Group.** Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ.* 2003, 326:41-44.
10. **Whitting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J: The development of QUADAS.** a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodology.* 2003, 10:doi:10.1186/1471-2288-3-25.
11. **Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M.** Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007, 10:e297.
12. **Des Jarlais DC, Lyles C, Crepaz N, the TREND Group.** Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND Statement. *AJPH.* 2004, 94:361-366.
13. **Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF.** Improving the quality of reporting of randomized controlled trials: the CONSORT Statement. *JAMA.* 1996, 276:637-639.
14. **Moher D, Liberati A, Tetzlaff J, Altman DG.** (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med.* 2009, 6:e1000097.
15. **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. *Lancet.* 1999, 354:1896-1900.
16. **Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB, the MOOSE Group.** Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA.* 2000, 283:2008-2012.
17. **Singh JP, Fazel S.** Forensic risk assessment: a metareview. *Crim Justice Behav.* 2010, 37:965-988.
18. **Delgado-Rodriguez M.** Systematic reviews of meta-analyses: applications and limitations. *J Epidemiol Community Health.* 2006, 60:90-92.

19. **Higgins J, Thompson S, Deeks J, Altman D.** Measuring inconsistency in meta-analyses. *BMJ*. 2003, 327:557-560.
20. **Cipriani A, Geddes JR, Furukawa TA, Barbui C.** Metareview on short-term effectiveness and safety of antidepressants for depression: an evidence-based approach to inform clinical practice. *Can J Psychiatry*. 2007, 52:543-534.
21. **Ruddy R, House A.** Meta-review of high-quality systematic reviews of interventions in key areas of liaison psychiatry. *Br J Psychiatry*. 2005, 187:109-120.
22. **Moher D, Tetzlaff J, Triccol AC, Sampson M, Altman DG.** Epidemiology and reporting characteristics of systematic reviews. *PLoS Med*. 2007, 4:447-455.
23. **Oxman AD, Guyatt GH.** Guidelines for reading literature reviews. *CMAJ*. 1988, 138:697-703.
24. **Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM.** Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodology*. 2007,7;doi:10.1186/1471-2288-7-10.
25. **Becker LA, Oxman AD.** Overviews of reviews. In *Cochrane Handbook for Systematic Reviews of Interventions 5.0.1*. Edited by Higgins J, Green S. Chichester: John Wiley & Sons; 2008.
26. STARD Group: STARD news. [<http://www.stard-statement.org/>]
27. PRISMA Group. PRISMA endorsers. [<http://www.prisma-statement.org/endorsers.htm>]
28. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics*. 1977, 33:159-174.