Using Technology to Increase the Reliability of Systematic Reviews

Authors:
Sandra Zelman Lewis, PhD
Gerald Borok, PhD, MPH
Todd Feinman, MD
Melanie Golob, MS
Karin Lawson-Remer, MPH

Karin Lawson-Remer, MPH
Vice President, Clinical Affairs
Doctor Evidence, LLC
klawsonr@doctorevidence.com
growthevidence.com

Presented at 11th G-I-N (Guideline International Network) Annual Conference
Melbourne, Australia
August 20-23, 2014

Disclosure of Interests

- Karin Lawson-Remer, MPH
- Vice President, Clinical Affairs
  - Employed by Doctor Evidence, LLC, a specialty evidence-based medicine software platform and company with a mission to provide stakeholders across the healthcare ecosystem with the most timely and accurate relevant medical evidence and related analytics.
- Some projects funded by industry, but all work performed by in-house, highly trained analysts using methodologically rigorous processes based on IOM standards.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Presentation Outline

- Background
- Objectives
- Methods
- Results
- Discussion
- Implications for Guideline Development

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Background

- Guidelines pivot on the evolving evidence; thus it is critical that all supporting evidence reviews be complete and reliable.
  - Systematic Reviews are increasingly relied upon in part due to limited resources as the supporting evidence for guidelines
  - Large number of new SRs being published (estimated at 2500 per year) and will only grow in their role supporting guideline development. [1]
  - Errors in unreliable systematic reviews (SRs) could add bias and/or harm when buttressing guidelines.
  - There is a small, but growing body of literature examining the quality and methodological rigor of systematic reviews.

[1] Moher D, et al. Epidemiology and reporting characteristics of systematic reviews. 2007. PLoS Med 4(3);e78. doi:10.1371/journal. pmed. 0040078

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Background

- An evaluation of 300 SRs found that the quality of reporting in many of the SRs reviewed was disappointing and inconsistent, especially in non-Cochrane reviews. [1]
  - Years searched: 2/3
  - Quality Assessment Information: 2/3
  - Funding sources not reported: 1/3
  - Publication bias assessed: < 1/4
  - Study design eligibility criteria/restriction: many did not state
  - Potential harms (in addition to benefits) of the intervention not mentioned: 1/4
  - Primary outcome reported: 1/3

[1] Moher D, et al. Epidemiology and reporting characteristics of systematic reviews. 2007. PLoS Med 4(3);e78. doi:10.1371/journal. pmed. 0040078

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Objectives

- The focus of this research is to address these issues: [1]
  - How rigorously are systematic reviews being performed?
  - How consistent are they in reporting their methods and their results?
- This research compares literature searching and data extraction methods to technology-based processes, specifically addressing these deficiencies:
  - Missing studies despite reported inclusion/exclusion criteria
  - Extraction errors
  - Inappropriate inclusion/exclusion criteria (i.e., too narrow/too broad, eliminating non-RCTs)
  - Incorrect input values for meta-analyses and impact on conclusions
  - Non-transparent study and datapoints identification
  - Identifying relevant studies since SR publication

[1] The PLoS Medicine Editors. Many reviews are systematic but some are more transparent and completely reported than others: Editorial. 2007. PLoS Med 4(3);e147. doi:10.1371/journal. pmed. 0040147

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Methods

- We audited three published SRs and meta-analyses in different therapeutic areas to determine search completeness, extraction accuracy, and transparency.
- SRs were selected as part of validation projects to compare published SRs vs. the Doctor Evidence technology-based platform.
- Published SR results will be compared with those generated from technology-supported methods employing identical PICOs. Discrepancies and underlying causes will be identified, evaluated, and reported separately for each process, then compared across methods.
- SRs were assessed for quality based on AMSTAR checklist

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

AMSTAR Results

Errors found in 4 different AMSTAR [1] criteria across the 3 SRs:

- Orme – 8/11 (No for 2, 5, 11)
  - #2: Was there duplicate study selection and data extraction?
  - #5: Was a list of studies (included and excluded) provided?
  - #11: Was the conflict of interest included?

- Wang – 8/11 (No for 2, 5, 11)
  - #2: Was there duplicate study selection and data extraction?
  - #5: Was a list of studies (included and excluded) provided?
  - #11: Was the conflict of interest included?

- Bach – 10/11 (No for 10)
  - #10: Was the likelihood of publication bias assessed?

[1] Shea et al. BMC Medical Research Methodology 2007 7:10 doi:10.1186/1471-2288-7-10

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

Results

Error types and frequencies* found in systematic review meta-analyses

Error type	Bach 2012	Orme 2012	Wang 2012
Conflicting Outcome Definitions**			20/8
Inconsistently Combining Tx Group Data			2/8
Improperly Combining Outcomes			1/8
Lack of Citation for Unpublished Data**			5/8
Incorrect Denominators		31/120	
Mislabeled Study		12/120	
Incorrect Input Values		12/120	
Miscalculation of Odds/Risk Ratios	1/6	16/120	

*Individual instances of that error within each meta-analysis (units are # of errors/total metas analyzed for that review)

** Multiple times per meta for some metas

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION

GROWTH

Results

After correcting errors identified in the SRs and re-running meta-analyses on the DRE platform, there were statistically significant changes for each of the three SRs.

- **Orme SR**: two treatment comparison findings which were non-significant in the SR would have been statistically significant
  - Rituximab combination therapy was significantly better than DMARDs alone for ACR 70.
  - Etanercept was significantly better than adalimumab for ACR 20.
  - Potential impact on article Abstract: would have changed one of the four results statements (non-significant to significant), and added a fifth results statement.
- **Bach SR**: one mortality outcome which was statistically significant in the SR would have been non-significant:
  - All-cause mortality: Bach SR RR result (RR 1.19, 95% CI 1.01-1.40) did not match DRE result (RR 1.45, 95% CI 0.99-2.14) for DLCST RCT (Saghir 2012).
    - This error did not change the all-cause mortality overall result due to relatively small enrolled patient population in the DLCST RCT vs. the considerably larger NLST RCT.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

Results

- **Wang SR**: three meta results which were reported as statistically significant in SR (underlined) would have been non-significant.
  - Figure 3. Moderate to Severe COPD Exacerbations Meta-Analysis
    - 1.1.1 Salmeterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.79 (0.70 to 0.90) to 0.80 (0.70 to 0.91)
    - 1.1.2 Formoterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.83 (0.73 to 0.96) to 0.86 (0.71 to 1.04)
    - 1.1.3 Indacaterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.82 (0.69 to 0.97) to 0.79 (0.61 to 1.02)
    - **Overall LABAs vs. Placebo: Fixed effect OR (95% CI) changed from 0.81 (0.75 to 0.88) to 0.79 (0.68 to 0.92)**
  - Figure 4. Severe COPD Exacerbations/Withdrawals due to Exacerbations Meta-Analysis
    - 1.2.1 Salmeterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.66 (0.49 to 0.89) to 0.75 (0.51 to 1.11)
    - 1.2.2 Formoterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.85 (0.68 to 1.06) to 0.59 (0.24 to 1.42)
    - 1.2.3 Indacaterol vs. Placebo: Fixed effect OR (95% CI) changed from 0.42 (0.21 to 0.83) to not available because studies were excluded
    - **Overall LABAs vs. Placebo: Fixed effect OR (95% CI) changed from 0.74 (0.63 to 0.88) to 0.68 (0.52 to 0.90)**
  - Correction of these errors did not change the statistical significance of the overall LABA vs. Placebo OR.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Discussion

- Published SRs were found to have errors that could have been avoided through more efficient and comprehensive technology-assisted processes.
  - First 3 SRs reviewed all contained at least one error
  - 3 distinct and different diseases, resulting in multiple types of errors
  - Most common types of errors:
    1. Incorrect data input for meta analyses
    2. Mislabeling or improper definitions (source study misinterpretation)
    3. Miscalculation of odds/risk ratios
- Technology-supported dual extraction + QC process and integrated analytic software in the same IT platform explains:
  - Technology-supported analyses result in fewer errors than the published SRs
  - Ability to identify the errors in the published SRs
- Research is continuing with more SRs with plans to publish results

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Implications

- Guideline development complexities and costs inevitably lead to reliance on existing SRs rather than full analyses of primary studies
- Can't trust that all SRs are error-free
- Cognizant of SRs that might change the direction or strength of recommendation, potentially due to analytical errors
- When multiple SRs on given PICO, guideline developers should look to see if results are consistent. Inconsistencies, and especially outliers, should be examined more closely for potential errors.
- Research makes one wonder if errors in SRs could rise to the level of adversely impacting patient outcomes
- Technological solutions for assessing primary literature could enhance quality and reduce the error rate. Efficiencies now available in these platforms eliminate the need for reliance on short cuts.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Acknowledgements

Karin Lawson-Remer, MPH
Vice President, Clinical Affairs
klawsonr@doctorevidence.com

Sandra Zelman Lewis, PhD
Chief Guidelines Officer
slewis@doctorevidence.com

Gerald Borok, PhD, MPH
Sr. Director, Client Solutions
gborok@doctorevidence.com

Todd Feinman, MD
CMO & Founder
tf@doctorevidence.com

Melanie Golob, MS
Senior Guidelines Associate
mgolob@doctorevidence.com

www.doctorevidence.com
http://growthevidence.com/

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Appendix

- Additional information

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

Inclusionary Criteria for SR Publications

- Orme et al. Systematic review and network meta-analysis of combination and monotherapy treatments in disease-modifying antirheumatic drug-experienced patients with rheumatoid arthritis: analysis of American College of Rheumatology criteria scores 20, 50, and 70. Biologics: Targets & Therapy, 2012; 6: 429-464.
  - Included by client as part of literature review of rheumatoid arthritis because met these criteria:
    - Comprehensive set of RCTs included in SR
    - Inclusion of full array of biologics
    - Thorough description of methodology
    - Recently published
- Wang et al. Effect of Long-Acting Beta-Agonists on the Frequency of COPD Exacerbations: A Meta-Analysis. Journal of Clinical Pharmacy & Therapeutics, 2012; 37: 204-211.
  - Included by client as part of development of clinical practice guideline by their professional society.
- Bach et al. Benefits and Harms of CT Screening for Lung Cancer: A Systematic Review. Journal of the American Medical Association, 2012; 307: 2418-2429.
  - Included by client as part of evaluation of guidelines (USPSTF and professional societies) for lung cancer screening using low-dose computed tomography because:
    - Developed as the SR for a multi-society collaborative initiative involving the American Cancer Society, American College of Chest Physicians, American Society of Clinical Oncology, and National Comprehensive Cancer Network to create the foundation for development of an evidence-based clinical guideline.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

Results Section of Abstract in RA SR publication

- DRE identified two treatment comparison findings which were non -significant in the SR, but would have been statistically significant if analyzed correctly.
  - Rituximab combination therapy was significantly better than DMARDs alone for ACR 70 – highlighted by SR as an "exception" in Abstract.
  - Etanercept was significantly better than Adalimumab for ACR 20 – not included in Abstract because not statistically significant.
- **Results section of Abstract:** "The systematic review identified 10,625 citations, and after a review of 2450 full-text papers, there were 29 and 14 eligible studies for the combination and monotherapy meta-analyses, respectively. In the combination analysis, all licensed bDMARD combinations had significantly higher odds of ACR 20/50/70 compared to DMARDs alone, **except for the rituximab comparison, which did not reach significance for the ACR 70 outcome (based on the 95% credible interval).** The etanercept combination was significantly better than the tumor necrosis factor-α inhibitors adalimumab and infliximab in improving ACR 20/50/70 outcomes, with no significant differences between the etanercept combination and certolizumab pegol or tocilizumab. Licensed-dose etanercept, adalimumab, and tocilizumab monotherapy were significantly better than placebo in improving ACR 20/50/70 outcomes. Sensitivity analysis indicated that including studies outside the target population could affect the results."

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

AMSTAR Checklist

AMSTAR Checklist

Article Name: _____

Response choices: Yes No Can't answer Not applicable **Total possible score**: 11

1. Was an 'a priori' design provided?

The research question and inclusion criteria should be established before the conduct of the review.

Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes."

2. Was there duplicate study selection and data extraction?

There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

Note: 2 people do study selection, 2 people do data extraction, consensus process or one person checks the other's work.

3. Was a comprehensive literature search performed?

At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

Note: If at least 2 sources + one supplementary strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as supplementary).

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

AMSTAR Checklist

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?

The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

Note: If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SINGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.

5. Was a list of studies (included and excluded) provided?

A list of included and excluded studies should be provided.

Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select "no."

6. Were the characteristics of the included studies provided?

In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

Note: Acceptable if not in table format as long as they are described as above.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH

AMSTAR Checklist

7. Was the scientific quality of the included studies assessed and documented?

'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?

The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

Note: Might say something such as "the results should be interpreted with caution due to poor quality of included studies." Cannot score "yes" for this question if scored "no" for question 7.

9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, I2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).

Note: Indicate "yes" if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions.

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION
GROWTH

AMSTAR Checklist

10. Was the likelihood of publication bias assessed?

An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olken).

Note: If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.

11. Was the conflict of interest included?

Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

Note: To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.

Shea *et al. BMC Medical Research Methodology* 2007 7:10 doi:10.1186/1471-2288-7-10

GUIDELINE & RESEARCH ORGANIZATIONS WORLDWIDE for TRANSPARENCY & HARMONIZATION GROWTH