Systematic Review Variants - The Good, the Bad and the Ugly

Barbara Kitchenham

David Budgen

Lech Madeyski

Sebastian Pizard

Background – 1

- Systematic reviews in SE are usually
  - *Standard SRs*: often using qualitative synthesis
  - *Mapping studies*: identifying and classifying the literature on some (usually broad) topic
  - *Tertiary studies* which are systematic reviews of systematic reviews
    - Often investigating the SR methodology used by SE researchers

Background - 2

- Recently three other SR variants have been introduced to SE researchers:
  - *Multivocal (MV) reviews*
  - *Grey Literature* (GL) *reviews*
  - *Rapid Reviews (RRs)*
- There are good justifications for these types of review, BUT
  - They make changes to the standard SR process
  - SE researchers need to understand not just benefits but also potential risks and how to minimize them

Background - 3

- SRs are a form of literature with a highly formalized review and reporting process
  - Optimized to support the delivery of *trustworthy* answers to research questions and recommendations for practice:
    - Based on a comprehensive body of well-conducted empirical research and a valid aggregation process
    - Resilient to human error
    - Capable of being reproduced by other researchers
    - Easily updated when new evidence becomes available
    - Assessed for strength of evidence and limitations
- Changes to the process risk undermining SR goals

Background - 4

- This lecture will discuss these review types in terms of
  - How they differ from current SR reviews
  - The implications of those differences in the context of SE in terms of
    - Good practice
      - Minimize the risks caused by changing the standard SR process
    - Bad practice
      - Reduce the scientific value of the review results
    - Ugly practice
      - Undermine the goals or principles of evidence-based SE

Justification for MV reviews and GL Reviews

- Seek to increase the relevance of SRs
  - By including informally published but current results
  - Particularly results reported by practitioners
- Introduced to SE researchers by Garousi et al. in 2016[1], 2019[2] & 2020[3]
  - Adjustments proposed by Kitchenham, Madeyski and Budgen[4]

---

1. Garousi, V. and Mäntylä, M.V., 2016. When and what to automate in software testing? A multi-vocal literature review. *Information and Software Technology*, *76*, pp.92-117.
2. Garousi, V., Borg, M. and Oivo, M., 2020. Practical relevance of software engineering research: synthesizing the community's voice. *Empirical Software Engineering*, *25*, pp.1687-1754.
3. Garousi, V., Felderer, M. and Mäntylä, M.V., 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology*, *106*, pp.101-121.
4. Kitchenham, B., Madeyski, L. and Budgen, D., 2022. How should software engineering secondary studies include grey material?. *IEEE Transactions on Software Engineering*, *49*(2), pp.872-882.

Justification for Rapid Reviews

- Seek to decrease the time & effort needed to produce evidence-based recommendations
    - Usually by omitting or simplifying stages in the SR process
- Introduced to SE researchers by Cartaxo et al. in 2018[5,6]
- Currently being investigated by Pizard, Lezama, Garcia, Vallespir and Kitchenham
    - Citation review and partial replication of Cartaxo's study

5. Cartaxo, B., Pinto, G. and Soares, S., 2018, June. The role of rapid reviews in supporting decision-making in software engineering practice. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (pp. 24-34).
6. Cartaxo, B., Pinto, G. and Soares, S., 2018. Towards a model to transfer knowledge from software engineering research to practice. *Information and Software Technology*, *97*, pp.80-82.

MV and GL Reviews

- *Multivocal reviews*
  - Aim to increase relevance to practitioners
    - Including new ideas and knowledge not available in the formal literature
    - Including the views of diverse authors
      - Not just academics
      - Practitioners, journalists, government policy makers, independent research and development firms, etc.
  - Garousi et al. propose
    - Use of Grey Literature to address these aims
    - Guidelines for Multivocal & Grey Literature reviews[2]

2. Garousi, V., Felderer, M. and Mäntylä, M.V., 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology*, *106*, pp.101-121

Deviation from SRs

- Definition of *grey literature*
  - Current SR standards recommend including grey literature
    - Assume the definition developed by library scientists (most recent is the Prague definition[7])
      - Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are ***protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories***, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body
    - MV & GL Reviews widen this definition

7. Schöpfel, J., 2010, December. Towards a Prague definition of grey literature. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010* (pp. 11-26).

Good Aspects

- Identifying the value of social media posts
  - Source of new ideas/topics
  - Source of information about practical issues
- Identifying approaches to
  - Find relevant social media posts
  - Assess the credibility of such material

Bad Aspects - 1

- No appreciation that the "unit" of an SR is an empirical study, not simply a piece of textual material
- No well-formulated definition of grey literature
  - Treating all forms social media & internet post as "grey literature", ignores the important differences
  - Shades of Grey Model is not easy to use[8]
    - Output control & Source expertese

8. Adams, R.J., Smart, P. and Huff, A.S., 2017. Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews*, *19*(4), pp.432-454.

Bad Aspects - 2

- Ignores the critical issues of the Prague definition that support SR **reproducibility & future updates**
  - Grey Literature is worthy of being ***collected and preserved***
  - Garousi and Mäntylä (2016)[3] cited 46 internet articles and white papers using URL addresses, but on 25th May 2021 only 19 were still accessible (Kitchenham et al.[4])
    - Use of the Wayback machine doesn't solve the problem

3. Garousi, V., Felderer, M. and Mäntylä, M.V., 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology*, *106*, pp.101-121.
4. Kitchenham, B., Madeyski, L. and Budgen, D., 2022. How should software engineering secondary studies include grey material?. *IEEE Transactions on Software Engineering*, *49*(2), pp.872-882.

Ugly Aspects

- Reporting MV and GL results without:
  - Distinguishing the source of the evidence
  - Providing separate syntheses
- Prevents
  - Rational assessment of the quality of evidence related to a specific result/recommendation
  - Reproducibility
    - when social media posts disappear
  - Valid updates to results/recommendations

Suggested Changes to MV Guidelines

- Use a model that better defines different *types of information*[9]
  - Researchers need to understand what they are talking about
  - Need to understand the value of, and limitations of, different types of social media post

9. Adams, J., Hillier-Brown, F.C., Moore, H.J., Lake, A.A., Araujo-Soares, V., White, M. and Summerbell, C., 2016. Searching and synthesising 'grey literature'and 'grey information'in public health: critical reflections on three case studies. Systematic Reviews, 5 (1), 164

Information Sources

Formally published Information
(White literature)

Books, book chapters , trade, academic and professional journals, magazines, and Conference/workshop proceedings

Informally published Information
(Grey literature Conforming to Prague definition)

Technical reports, theses, white papers, preprints or supplementary materials (e.g., ArXiv.org, zenodo, Figshare, PROSPERO)

Use as a source of primary studies i.e., reports of an empirical study

Admissible for an SR

Self-Published Information
(Social Media posts)

Social media posts ,Blogs, Vlogs, Tweets, Q&A Fora, Wikis, Predatory publishing sources, Online discussion groups

Use like personal opinion surveys and for new research ideas

Unpublished Information
(Personal Communications)

E-mails, memos, meeting notes

Use to support qualitative industry case studies (e.g., data triangulation)

Inadmissible for an SR

Suggested Changes-2

- Treat grey literature conforming with the Prague definition the same as other primary studies
  - If it passes the eligibility criteria, grey literature can be treated just like any other primary study in a SR
    - Subject to the same risk of bias (quality) assessments
    - Aggregated with other primary studies
  - The results/recommendations from the aggregation can be
    - Assessed for validity & strength of evidence
    - Reproduced
    - Updated when new evidence is found

Suggested Changes - 3

- Treat Social Media posts that do **not** conform with the Prague definition as a different type of evidence
  - Essentially a personal opinion survey
    - Without any valid sampling process
    - Subject to risk of personal bias
  - Aggregate separately from SR primary studies
  - Use to compare and contrast with the results of the SR
  - Use to suggest areas for future research

Rapid Reviews

- RRs (aka *focused reviews*) have been used in other domains for many years
  - Urgency surrounding COVID increased their importance
    - More than 3000 RRs published[10]
    - Has encouraged more research into the RR process
- Cochrane Rapid Review Methods Group proposed the following definition of an RR[11]:
  - **A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting various methods to produce evidence for stakeholders in a resource-efficient manner**
  - However only about 50% of healthcare RRS conform[12]

---

10. Tricco, A.C., Straus, S.E., Ghaffar, A. and Langlois, E.V., 2022. Rapid reviews for health policy and systems decision-making: more important than ever before. *Systematic Reviews*, *11*(1), p.153.
11. Hamel, C., Michaud, A., Thuku, M., Skidmore, B., Stevens, A., Nussbaumer-Streit, B. and Garritty, C., 2021. Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *Journal of Clinical Epidemiology*, *129*, pp.74-85.
12. Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E. and Boyer, L., 2023. Rapid literature review: definition and methodology. *Journal of Market Access & Health Policy*, *11*(1), p.2241234.

Reporting RRs in Healthcare

- Recommendations, which assume a specified knowledge user(s)[13] :
  - Work from protocol
  - Accurately and transparently document all steps
  - Use clear language that will be understandable to knowledge users
    - Avoid the use of jargon or technical terms, except where essential
    - Some technical terms may have a different definition in everyday usage
  - Provide enough detail to reproduce the review
  - Summarize the methodological strengths and weaknesses
    - Use language designed to help non-experts interpret and judge the value of the review;
  - Consider the needs of the knowledge user
    - Discuss their time frames,
    - Define the type of report they require
  - Communicate with the knowledge users
    - Preferably throughout the review process,
    - At a minimum discuss communication requirements in advance

13. Kelly, S.E., McGowan, J., Barnhardt, K. and Straus, S.E., 2022. Paper 4: a review of reporting and disseminating approaches for rapid reviews in health policy and systems research. *Systematic Reviews*, *11*(1), p.152.

Initial Rapid Review in SE Domain

- Cartaxo et al.[5] emphasized the following characteristics to be important for SE researchers:
  - Timely results and reduced costs
  - Collaboration with practitioners
  - Presenting results of an RR in formats that appeal to practitioners
    - Advocated *Evidence Briefings* to provide a one page summary
- Undertook a case study to investigate RRs in the context of collaboration with a client
  - Client was an R&D company using Agile methods
  - Experienced problems with the Customer role
  - RR was used to identify methods to address the problem
  - Results were reported to client, who adopted some of the recommendations

5. Cartaxo, B., Pinto, G. and Soares, S., 2018, June. The role of rapid reviews in supporting decision-making in software engineering practice. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (pp. 24-34).

SE RRs

- Using citation analysis, Pizard found 22 papers reporting RRs (search end date Nov 2023)
  - 13 did not work with a specific knowledge user
  - 19 did not report using a protocol
    - 6 referred to the “RR protocol” but meant the generic RR process
  - 21 did not do any quality assessment (RoB)
    - 1 omitted studies “without sufficient detail or with unclear explanations”
    - None assessed the strength & weakness of recommendations
  - 6 did not mention limitations of the RR
  - 18 did not report synthesis adequately
  - 15 did not specify the dates of the search
  - 7 did not report the primary studies they used
  - 4 used only a subset of papers found
  - 3 added papers without explanation

Good Aspects - 1

- RRs can support collaboration with practitioners
  - Pizard replicated Cartaxos's study
    - Clients had no experience of SE research methods
      - Small IT company producing Digital out-of-home advertising product
    - Client using agile methods experiencing problems with KM
    - RR used to identify methods to address problem
    - Recommendations were reported to the clients who:
      - Had confidence that the RR recommendations were sound
      - Reported that they adopted some of the recommendations
  - Citation analysis found 3 of 22 papers
    - Reported collaboration between industry (R&D groups) & review authors
- RRs also support SE research projects
  - In 5 papers the review authors were the knowledge users

Good Aspects - 2

- RR's can reduce effort and/or timescales
  - Cartaxo reduced effort & timescales
    - 6 elapsed working days[5]
    - Did not report how much effort from his co-authors
  - Pizard reduced effort
    - 3 months elapsed but took approx 150 working hours
    - Researchers could only work part-time

5. Cartaxo, B., Pinto, G. and Soares, S., 2018. Towards a model to transfer knowledge from software engineering research to practice. *Information and Software Technology*, *97*, pp.80-82.

Bad Aspects -1

- Treating Evidence Briefings(EB) & RR as related issues
  - Cartaxo's case studies & Pizard's case studies
    - Both presented clients with an EB
    - Both case studies noted that
      - Clients required a meeting to fully understand the results
      - An EB without explanation or supplementary material is insufficient
  - Other SE RRs
    - Only two other studies relied on Evidence Briefings
    - All others recognized the need for more details in a formally published article

Bad Aspects - 2

- Using the process defined by Cartaxo without reflection
  - Cartaxo's process suited his *research* goals
    - Searched only Scopus
    - Excluded secondary studies
    - Excluded assessment of primary studies
    - Did not consider strength of evidence
- However, such decisions may not be appropriate on other circumstances, reviewers may need to
  - Search recent specialist workshops/conferences for new results
  - Include existing SRs to benefit from tested search strings & known primary studies
  - Assess primary study methodological quality & strength of evidence to assess trustworthiness of recommendations

Ugly Aspects

- Reporting practice for SE RRs is terrible
  - Failure to report fully reduces the scientific value of the review authors work
  - The review authors should have done all the work necessary to report their work adequately
    - So there is no need to omit details in a review report intended for publication

Recommendations

- Always reflect on the specific changes to the SR process required for a specific RR
- In cases where there is no defined knowledge user
  - Papers claiming to be RRs should be judged by the standard of a mapping study[14]
- When there is a knowledge user
  - The results can be delivered as soon as available
  - Writing the results for formal publication should support the needs of other researchers for
    - Reproducibility
    - Ability to update evidence
  - Follow-up the results of adopting the recommendations
    - To provide industry–based evidence of the benefit of the recommendation(s)

14. Kitchenham, B., Madeyski, L. and Budgen, D., 2022. SEGRESS: Software engineering guidelines for reporting secondary studies. *IEEE Transactions on Software Engineering*, *49*(3), pp.1273-1298.

Conclusions

- Variants of systematic reviews
  - Have been developed to address
    - Personal opinions of people affected by SR recommendations
    - The requirements for quick evidence summaries of new results
- If we adopt variants in SE we need to ensure that we understand them
  - Scientific jargon is not the same in all disciplines, e.g., the meaning of the term “protocol”
  - Limitations of a variant may be based on tacit assumptions unclear to SE researchers

Final Thoughts

- Both RR case studies addressed problems found using Agile methods
  - Relatively mature methods
  - Still exhibit practical problems
- Should we be including such material in SE education?
  - To introduce our students to the practical problems associated with our methods
- Other RRs covered topics of relevance to practitioners, which ought to be of value to SE students e.g.,
  - Object-Relational Mapping Code Smells for Java
  - Model-based security testing in IoT systems

References

1. Garousi, V., Borg, M. and Oivo, M., 2020. Practical relevance of software engineering research: synthesizing the community's voice. *Empirical Software Engineering*, *25*, pp.1687-1754.
2. Garousi, V., Felderer, M. and Mäntylä, M.V., 2019. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and software technology*, *106*, pp.101-121.
3. Garousi, V. and Mäntylä, M.V., 2016. When and what to automate in software testing? A multi-vocal literature review. *Information and Software Technology*, *76*, pp.92-117.
4. Kitchenham, B., Madeyski, L. and Budgen, D., 2022. How should software engineering secondary studies include grey material?. *IEEE Transactions on Software Engineering*, *49*(2), pp.872-882.
5. Cartaxo, B., Pinto, G. and Soares, S., 2018, June. The role of rapid reviews in supporting decision-making in software engineering practice. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (pp. 24-34).
6. Cartaxo, B., Pinto, G. and Soares, S., 2018. Towards a model to transfer knowledge from software engineering research to practice. *Information and Software Technology*, *97*, pp.80-82.

References - Continued

7. Schöpfel, J., 2010, December. Towards a Prague definition of grey literature. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010* (pp. 11-26).
8. Adams, R.J., Smart, P. and Huff, A.S., 2017. Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews*, *19*(4), pp.432-454.
9. Adams, J., Hillier-Brown, F.C., Moore, H.J., Lake, A.A., Araujo-Soares, V., White, M. and Summerbell, C., 2016. Searching and synthesising 'grey literature'and 'grey information'in public health: critical reflections on three case studies. Systematic Reviews, 5 (1), 164.

References - Continued

10. Tricco, A.C., Straus, S.E., Ghaffar, A. and Langlois, E.V., 2022. Rapid reviews for health policy and systems decision-making: more important than ever before. *Systematic Reviews*, *11*(1), p.153.
11. Hamel, C., Michaud, A., Thuku, M., Skidmore, B., Stevens, A., Nussbaumer-Streit, B. and Garritty, C., 2021. Defining rapid reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *Journal of Clinical Epidemiology*, *129*, pp.74-85.
12. Smela, B., Toumi, M., Świerk, K., Francois, C., Biernikiewicz, M., Clay, E. and Boyer, L., 2023. Rapid literature review: definition and methodology. *Journal of Market Access & Health Policy*, *11*(1), p.2241234.
13. Kelly, S.E., McGowan, J., Barnhardt, K. and Straus, S.E., 2022. Paper 4: a review of reporting and disseminating approaches for rapid reviews in health policy and systems research. *Systematic Reviews*, *11*(1), p.152.
14. Kitchenham, B., Madeyski, L. and Budgen, D., 2022. SEGRESS: Software engineering guidelines for reporting secondary studies. *IEEE Transactions on Software Engineering*, *49*(3), pp.1273-1298.