Towards improving quality of the evidence base for medical decision-making

Jansen, M.S.

Chapter 8

Summary and general discussion

Summary

This thesis aimed to address challenges in the evidence base for medical decision-making, focusing on various stages within the research enterprise, from designing and executing studies to their translation in decision-making. Specifically, we examined potential optimism in sample size calculations of trial protocols and the role of research ethics committees (RECs) in evaluating them, identified predictors of premature trial termination, and assessed publication rates of small trials as well as potential predictors. Additionally, we identified factors that determine the need for real-world evidence (RWE) to inform market approval and reimbursement decisions. By tackling the abovementioned issues, this thesis aimed to contribute to improving the quality and efficiency of medical research.

The first area of investigation in this thesis concerned sample size calculations in clinical trials. Realistic assumptions regarding expected treatment effects are important to design adequately powered studies and prevent inconclusive results. However, in line with previous literature,[1-4] we found that hypothesised effect sizes in sample size calculations of clinical trials were often optimistic, with the majority of trials (80%) finding treatment effects that were smaller than expected. While RECs review sample size calculations, our findings indicate that requests for justification of assumptions regarding effect sizes are limited. About 1 in 5 trials received REC comments on sample size, and these primarily concerned calculation errors rather than the plausibility of hypothesised effect sizes. When modifications were made during ethics review, they led to a relevant reduction of the hypothesised effect size for only one trial. Notably, industry sponsored trials appeared to have very little to no overestimation, in contrast to investigator initiated trials.

Another important issue contributing to inconclusive results and research waste is early trial termination. Similar to other studies,[5,6] we found that approximately one-third of clinical trials terminate prematurely, with recruitment failure being the most common reason. Several characteristics were associated with early termination, including multicentre designs and investigator sponsorship, the latter of which was specifically linked to recruitment failure. Possibly, academic investigators may have less access to resources and dedicated personnel to optimise recruitment, and may be more optimistic in estimating recruitment rates due to lack of experience and empirical data / pilot data compared to industry. In a qualitative study, overly optimistic recruitment rates, narrow eligibility criteria, lack of engagement and cooperation between the trial team and recruiters, lack of competence and experience, and insufficient initial funding were most often mentioned by various stakeholders as reasons for recruitment failure.[7] Interestingly, we found that elements of the ethics review process, such as the number and type of comments raised by REC reviewers, appeared predictive of early trial termination. This suggests that ethics review could serve as an early warning system to identify trials at high risk of failing, potentially allowing for preventive measures such as feasibility assessments or study design modifications.

We also examined the dissemination of clinical trial results, particularly focusing on small trials (cut-off arbitrarily set at fewer than 150 participants), which may be at higher risk of non-publication. Our research found that 30% of these trials did not disseminate their results (neither in a peer-reviewed journal or on the trial record), and that publication rates appear to stagnate after 5 years of follow-up since trial completion. The median time-to-publication of 3.4 years in our study of small trials was longer compared to a recent large meta-analysis investigating publication rates in trials of all sizes (median time-to-publication of 2.1 years).[8] Notably, we found that small trials that were not able to recruit their planned sample size, had an increased risk of not disseminating their results. These findings illustrate how issues during one research stage (i.e., trial planning and conduct) likely affect subsequent stages (e.g., publication), and further emphasises the need to reduce premature trial terminations.

Finally, we investigated the evolving role and need for RWE in regulatory and health technology assessment (HTA) decision-making. Through literature reviews and stakeholder interviews, we identified factors that increase the necessity and desirability of RWE to inform market approval and reimbursement decisions,

respectively. The need for RWE depended on two pillars; namely (1) the questions needing to be addressed to facilitate decision-making, and (2) contextual factors related to the feasibility of traditional trials, their ethics considerations, limitations of available evidence, and disease-specific circumstances. We found that the contextual factors influencing the need for RWE were similar for market approval and reimbursement decisions, but factors related to the first pillar differed. Reimbursement decisions typically require a broader scope of questions needing to be addressed in comparison to market approval decisions, such as those related to comparative effectiveness, costs, treatment implementation and quality of care. This leads to different evidence requirements between both decision-making domains, and likely an inherently greater need for RWE for HTA bodies. The current findings could help streamline evidence generation for both decision-making domains, and provide a starting point for weighing different types of evidence in decision-making processes.

Strengths and limitations

While each study of this thesis has its own strengths and weaknesses, there are also several that relate to the thesis as a whole. A key strength is its broad scope, examining multiple stages of the medical research enterprise, which contributes to a comprehensive perspective on the challenges in the evidence base for medical decision-making. We also made use of unique data sources, including stakeholder perspectives and RECs documents, the latter of which has been relatively underutilised in research and provided novel insights. Moreover, we integrated diverse methodologies, including literature reviews, stakeholder interviews, publicly available data (e.g., trial registries and publications), and confidential REC data.

However, some limitations should be acknowledged. As with all research, our analyses were limited by the quality of the underlying data. Some data sources contained or may have contained missing data (e.g., trial records, trial completion status, stakeholders responses), and some were subject to potential misclassification, subjective interpretation (e.g., thematic analysis of interviews, categorisation of REC comments) or limited generalisability (comments of one Dutch REC). Additionally, while we identified potential predictors of trial outcomes (e.g., early termination, publication), these do not imply causality. Most studies were exploratory in nature, with no adjustment for multiple testing and fairly small sample sizes, which may have resulted in false-positive and false-negative findings, and should thus be confirmed in future studies. And although we discussed the role of RWE in decision-making, we did not solve one of epidemiology's ultimate challenges: how to be certain that a non-randomised study yields valid results.

Recommendations

Improving the quality of the evidence base for medical decision-making and reducing research waste requires a concerted effort from all stakeholders involved in the research enterprise. As highlighted in prior calls for reform,[9-14] a cultural shift is necessary to move away from an emphasis on research quantity towards promoting research that is clinically relevant, methodologically robust, and accessible. Relying solely on individual researchers to self-regulate is apparently insufficient. Instead, external incentives are necessary to move towards these goals. Entities with gatekeeping functions, including research funders, RECs, journal editors, and authorities such as regulatory agencies, HTA bodies, and governments, could provide these external incentives. In this thesis, we primarily focused on RECs, regulatory agencies, and HTA bodies.

Some external incentives have already led to positive developments. To improve identifiability and transparency of clinical trials and mitigate publication bias, trial registries have been established since the early 2000s, with trial registration mandated by various gatekeepers, including certain funders, RECs, regulatory authorities (e.g., EMA, FDA – for regulated interventions such as medicines), and journal

editors.[15-17] Trial records serve as a valuable resource for identifying trials and accessing their results, regardless of publication status.[18] Similarly, online repositories (e.g., medRxiv, osf.io) provide platforms for preregistering study protocols, disseminating findings, and sharing data across all study types on a voluntary basis. Although often criticised for a lack of peer review, online repositories function essentially similarly to trial registries. Nonetheless, certain challenges remain. Trial registration requirements are primarily enforced by regulatory authorities and only for regulated interventions, leading to gaps in registration for non-regulated interventions (e.g., surgical or behavioural studies).[17,19] Additionally, retrospective registration, outdated trial records, poor data quality including missing protocol and outcome information, are widespread issues that undermine the usefulness of these registries.[20] Stricter reinforcement (e.g., beyond regulated interventions), as well as broadening of registration requirements (e.g., update of the trial record during potential amendments and upon study completion) may be necessary to overcome these issues.

Further structural changes are necessary to reduce research waste and improve the quality of the evidence base for medical decision-making. The following recommendations, largely informed by the findings in this thesis, outline potential improvements across several key gatekeepers and research stages.

Research Ethics Committees

Although overestimation of treatment effects in sample size calculations is common, and justifications for chosen parameters are often missing from protocols, REC reviewers appear to rarely challenge or request justifications for chosen parameters. To help reduce underpowered studies and inconclusive results, RECs may consider to put increased emphasis on the plausibility of hypothesised effect sizes in trial protocols, particularly for investigator-initiated trials.

Similarly, trial feasibility is currently not a key focus of ethics review, while early terminations, particularly due to recruitment failure, are frequent. RECs may consider to play a more proactive role in assessing feasibility – for example, by critically appraising the plausibility of recruitment projections, eligibility criteria, and resource availability (e.g., funding, experienced personnel). If not clearly reported in the protocol, RECs may request estimates of the available population, expected recruitment rates, and supporting data. Increased emphasis on these aspects might be helpful in reducing preventable early terminations.

Moreover, although Dutch RECs require annual study updates and final result submissions within one year of study completion, adherence to these requirements appears low – we found that less than half of the trials investigated in this thesis were compliant. Currently, it seems investigators are only sent reminders incidentally. Furthermore, while trial protocols must contain a section on public results dissemination, there is no follow-up. RECs should actively urge investigators to comply with these rules on a systematic basis, including regarding the public dissemination of study results. Automated systems could facilitate this without increasing workload for RECs. Proactive follow-up could help improve adherence and might increase timely publication of study results on at least a trial record.

Funders

Funding mechanisms play a critical role in research quality and feasibility. As outlined earlier, a shift is necessary from an emphasis on research quantity towards promoting research that is clinically relevant, methodologically robust, and accessible. In alignment with these goals, and to improve the likelihood that studies will succeed, funders should consider funding fewer studies and incorporating stricter criteria for grant allocation, but increase the financial support per study.

Regulators and HTA Bodies

Regulatory and HTA decision-making increasingly relies on RWE, and authorities, including EMA, continue to develop important guidance on methodological and data considerations for its use.[21,22] While

EMA is exploring use cases of regulatory-led RWE studies in its RWE framework,[23] the specific circumstances in which RWE is desired by decision-makers, as well as in what form, remain insufficiently defined in current guidance documents aimed at sponsors, particularly in the context of market approval decisions. Potentially utilising this thesis' findings, regulatory authorities should outline these circumstances so that sponsors may better respond to decision-makers evidence requirements. This may similarly include circumstances in which RWE studies may not be sufficient or desired to inform market approval and/or reimbursement decisions.

Additionally, the framework developed in this thesis for assessing the need for RWE may have broader applications beyond regulatory and HTA decision-making. Other healthcare policy areas, such as screening programs, also often require integration of multiple evidence types (e.g., RCTs, non-randomised RWE studies). Expanding this framework to other domains may be helpful to determine evidence requirements and improve transparency and consistency in evidence appraisal across decision-making areas.

Future research

The findings of this thesis provide valuable insights that contribute towards improving the quality of the evidence base for medical decision-making. However, most studies were exploratory in nature and had limited generalisability (e.g., focusing on the ethics review process of a single Dutch REC), necessitating future research to validate these findings in broader contexts. Additionally, while we identified factors determining the need for RWE in decision-making, these represent only one part of the complex process of weighing different types of evidence. As such, several areas require further investigation.

Firstly, the studies on sample size calculations and early trial termination could be replicated on a larger scale, involving multiple RECs across different regions and countries, to validate and expand upon our findings. The results of these studies could inform the development of predictive models to assess the probability of early trial termination at time of ethics review. Such models could help RECs identify high-risk studies and implement targeted preventive measures (e.g., feasibility checks, recommendations for study design modifications or acquiring additional funding).

Secondly, the development of predictive models could also be extended to other outcomes, such as delays in trial execution or non-publication of study results. Similarly, this approach could be applied earlier in the research process, at the funding stage. If elements of research proposals and their review process for funding are predictive of future trial success, funders could consider integrating these findings into grant allocation strategies (e.g., grant size, timing of disbursement, follow-up or management strategies) to help reduce research waste.

Thirdly, qualitative studies could provide further insight into how academic investigators determine hypothesised effect sizes used in sample size calculations. The protocols assessed in this thesis only provide a limited view on this process. Anecdotal evidence suggests that researchers adapt sample size parameters in order to achieve a feasible sample size, rather than according to clinical relevance, leading to sample sizes that are practically feasible but from an ethical point of view are too small. Qualitative studies may be useful to determine whether this phenomenon, also referred to as "sample size samba",[24] indeed occurs, and whether the selection of overly optimistic hypothesised effects is primarily driven by feasibility constraints, optimism bias, or lack of knowledge regarding the potential consequences of overly optimistic parameters. This information could help inform potential strategies to improve sample size calculations of study protocols (e.g., educational interventions).

Fourthly, further research is needed on the role of RWE in regulatory and HTA decision-making. While this thesis identified factors determining the need for RWE, their relative importance is unclear. It is likely

that certain factors are more influential than others, and factors likely interact with each other. Future studies could assess their relative importance to further clarify the need for RWE in decision-making.

Finally, our findings suggest there is no formal framework used by regulators to appraise heterogeneous types of evidence. Future studies could further investigate how decision-making is currently done by regulatory authorities. Are different types of evidence indeed weighed implicitly, and what are potential consequences for resulting decisions (e.g., inconsistencies, transparency)? This research could help build towards the development of a framework for evidence appraisal, which integrates relevant aspects for weighing (e.g., methodological quality, data quality, need for RWE, consequences decision), to increase consistency and transparency in decision-making.

Conclusion

This thesis investigated several challenges in the research enterprise that undermine the quality of the evidence base for medical decision-making. We focused on various stages within the research enterprise, from designing and executing studies towards their translation in decision-making. Treatment effects in sample size calculations are often overestimated, early trial termination – particularly due to recruitment failure – is common, and small trials frequently go unpublished. As RECs review trial protocols before initiation, they have the potential to influence trial outcomes. Our findings suggest RECs could take a more proactive role in scrutinising the plausibility of hypothesised effects in sample size calculations and encourage result dissemination. Moreover, ethics review characteristics appear predictive of premature trial termination, highlighting an opportunity to identify high-risk trials early. Building upon the findings in this thesis, future studies could explore the development of prediction models for trial termination, allowing for potential targeted preventive measures to reduce this risk. Finally, we identified factors that determine the need for RWE in regulatory and HTA decision-making. These findings could be used to help clarify evidence requirements by decision-makers, and might serve as a first step towards a decision-making framework to systematically evaluate heterogeneous types of evidence.

References

1. Djulbegovic B, Kumar A, Magazin A, et al. Optimism bias leads to inconclusive results-an empirical study. *J Clin Epidemiol*. Jun 2011;64(6):583-93. doi:10.1016/j.jclinepi.2010.09.007

2. Zakeri K, Noticewala S, Vitzthum L, Sojourner E, Shen H, Mell L. 'Optimism bias' in contemporary national clinical trial network phase III trials: are we improving? *Ann Oncol*. Oct 1 2018;29(10):2135-2139. doi:10.1093/annonc/mdy340

3. Olivier CB, Struss L, Sunnen N, et al. Accuracy of Event Rate and Effect Size Estimation in Major Cardiovascular Trials: A Systematic Review. *JAMA Netw Open*. Apr 1 2024;7(4):e248818. doi:10.1001/jamanetworkopen.2024.8818

4. van Zwet E, Gelman A, Greenland S, Imbens G, Schwab S, Goodman SN. A New Look at P Values for Randomized Clinical Trials. *NEJM Evid*. Jan 2024;3(1):EVIDoa2300003. doi:10.1056/EVIDoa2300003

5. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA*. Mar 12 2014;311(10):1045-51. doi:10.1001/jama.2014.1361

6. Speich B, Gryaznov D, Busse JW, et al. Nonregistration, discontinuation, and nonpublication of randomized trials: A repeated metaresearch analysis. *PLoS Med*. Apr 2022;19(4):e1003980. doi:10.1371/journal.pmed.1003980

7. Briel M, Elger BS, McLennan S, Schandelmaier S, von Elm E, Satalkar P. Exploring reasons for recruitment failure in clinical trials: a qualitative study with clinical trial stakeholders in Switzerland, Germany, and Canada. *Trials*. Nov 25 2021;22(1):844. doi:10.1186/s13063-021-05818-0

8. Showell MG, Cole S, Clarke MJ, DeVito NJ, Farquhar C, Jordan V. Time to publication for results of clinical trials. *Cochrane Database Syst Rev*. Nov 27 2024;11(11):MR000011. doi:10.1002/14651858.MR000011.pub3

9. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. Jul 4 2009;374(9683):86-9. doi:10.1016/S0140-6736(09)60329-9

10. Al-Shahi Salman R, Beller E, Kagan J, et al. Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. Jan 11 2014;383(9912):176-85. doi:10.1016/S0140-6736(13)62297-7

11. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. Jan 11 2014;383(9912):156-65. doi:10.1016/S0140-6736(13)62229-1

12. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. Jan 11 2014;383(9912):166-75. doi:10.1016/S0140-6736(13)62227-8

13. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. Jan 18 2014;383(9913):267-76. doi:10.1016/S0140-6736(13)62228-X

14. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. Jan 18 2014;383(9913):257-66. doi:10.1016/S0140-6736(13)62296-5

15. Dickersin K, Rennie D. The evolution of trial registries and their use to assess the clinical trial enterprise. *JAMA*. May 2 2012;307(17):1861-4. doi:10.1001/jama.2012.4230

16. De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet*. Sep 11-17 2004;364(9438):911-2. doi:10.1016/S0140-6736(04)17034-7

17. Dal-Re R, Bracken MB, Ioannidis JP. Call to improve transparency of trials of non-regulated interventions. *BMJ*. Mar 27 2015;350:h1323. doi:10.1136/bmj.h1323

18. DeVito NJ, Morley J, Smith JA, Drysdale H, Goldacre B, Heneghan C. Availability of results of clinical trials registered on EU Clinical Trials Register: cross sectional audit study. *BMJ Med.* 2024;3(1):e000738. doi:10.1136/bmjmed-2023-000738

19. Azar M, Riehm KE, Saadat N, et al. Evaluation of Journal Registration Policies and Prospective Registration of Randomized Clinical Trials of Nonregulated Health Care Interventions. *JAMA Intern Med.* May 1 2019;179(5):624-632. doi:10.1001/jamainternmed.2018.8009

20. DeVito NJ, Goldacre B. Trends and variation in data quality and availability on the European Union Clinical Trials Register: A cross-sectional study. *Clin Trials.* Apr 2022;19(2):172-183. doi:10.1177/17407745211073483

21. European Medicines Agency. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence. Draft. 2024. Accessed March 14, 2025. https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence_en.pdf

22. European Medicines Agency. Guideline on registry-based studies. 2021. Accessed March 14, 2025. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf

23. European Medicines Agency. Real-world evidence framework to support EU regulatory decision-making. Second report on the experience gained with regulator-led studies from February 2023 to February 2024. 2024. Accessed March 14, 2025. https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-2nd-report-experience-gained-regulator-led-studies-february-2023-february-2024_en.pdf

24. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet.* Apr 9-15 2005;365(9467):1348-53. doi:10.1016/S0140-6736(05)61034-3