Ruling out risks in medical research

Anttila, Sten; Persson, Johannes; Rosén, Måns; Vareman, Niklas; Vitols, Sigurd; Sahlin, Nils-Eric

Published in:
Journal of Risk Research

DOI:
10.1080/13669877.2019.1586750

2019

Document Version:
Publisher's PDF, also known as Version of record

Routledge
Taylor & Francis Group

Journal of Risk Research

ISSN: 1366-9877 (Print) 1466-4461 (Online) Journal homepage: https://www.tandfonline.com/loi/rjrr20

Ruling out risks in medical research

Sten Anttila, Johannes Persson, Måns Rosén, Niklas Vareman, Sigurd Vitols & Nils-Eric Sahlin

To cite this article: Sten Anttila, Johannes Persson, Måns Rosén, Niklas Vareman, Sigurd Vitols & Nils-Eric Sahlin (2019): Ruling out risks in medical research, Journal of Risk Research, DOI: 10.1080/13669877.2019.1586750

To link to this article: https://doi.org/10.1080/13669877.2019.1586750

Published online: 19 Mar 2019.

JOURNAL OF RISK RESEARCH
https://doi.org/10.1080/13669877.2019.1586750

Ruling out risks in medical research

Sten Anttila[a], Johannes Persson[b], Måns Rosén[c], Niklas Vareman[d], Sigurd Vitols[e] and Nils-Eric Sahlin[a,f]

[a]Medical Ethics, Lunds Universitet, Lund, Sweden; [b]Department of Philosophy, Lunds Universitet, Lund, Sweden; [c]Department of Learning, Informatics, Management, and Ethics, Karolinska Institutet, Stockholm, Sweden; [d]Medical Ethics, Lunds Universitet, Lund, Sweden; [e]Department of Medicine, Clinical Pharmacology, Karolinska Institutet, Stockholm, Sweden; [f]Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

In medical research, it is not unusual that risks are ruled out without any specification the exact risk that was ruled out. This makes it difficult to balance expected health benefits and risk of harm when choosing between alternative treatment options. International guidelines for reporting medical research results are sufficiently specific when it comes to establishing health benefits. However, there is a lack of standards for reporting on ruling out risks. We argue that transparency is needed, as in the case of non-inferiority trials. The Consolidated Standards of Reporting Trials and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statements should be revised accordingly.

ARTICLE HISTORY
Received 23 March 2017
Accepted 19 October 2018

KEYWORDS
Ruling out risks; value judgment; transparency; risk margin; Consolidated Standards of Reporting Trials (CONSORT); Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

Introduction

In any rational decision procedure, there are two questions that have to be settled. First, what is the possible gain involved in the decision? Second, what are the risks? To decide on the best possible, or at least an acceptable, line of action both these questions need to be addressed. Furthermore, in any evidence-based practice, transparent and accurate research reporting is crucial when addressing these two questions.

However, in medical research, there is a peculiar focus on gains. Guidelines for reporting research results are well formulated, specific, and rigorous when the research question concerns the effectiveness of a treatment. But this is not the case when it comes to the risk of harm. Here, instead, specific guideline standards are lacking. This opens up for misleading conclusions. A common statistical mistake is to conclude that a nonsignificant result regarding harm means that there is no difference between treatment and control, and that the risk of harm can, therefore, be ruled out. This, in turn, makes it difficult to balance possible health benefits with the risk of harm in decision making.

Our *purpose* is to clarify the above problem of reporting medical research on whether a risk of harm can be ruled out. We also suggest a way to improve the situation within the framework of established guidelines for reporting medical research. A systematic review, concerning folic

CONTACT Sten Anttila anttila@sbu.se Lunds Universitet, Medical Ethics, Lund, 221 84, Sweden.

acid and possible cancer risks (Vollset et al. 2013), illustrates our arguments. However, the problems found are not restricted to this particular study. Similar examples of poor justification for ruling out risks can be found in many other studies, for example, published by the prestigious Cochrane Library (Akl 2014; Leone et al. 2016).

Communicating risk from risk assessment to risk management

The basic idea of risk analysis is that there is a scientific part: risk assessment, and a decision part: risk management, combined by risk communication (expressed, e.g. in the EU food law: https://ec.europa.eu/food/safety/general_food_law/principles_en). A democratic principle demands that the values of experts should not influence the decisions made, that is, the risk assessors should not decide on the acceptability or non-acceptability or risks. Ideally they should only provide the facts: what the risks are for different scenarios and exposures. It is up to the decision makers (in high level policy decisions, these are elected representatives of the public that are supposed to express the values that a majority of the public agrees with) to judge what risks are acceptable or what not. This is not to say that only the decision makers make value judgements, while the risk assessors make none; a risk assessment involves values of many kinds (Vareman and Persson, 2010). Rather, the separation is one of the responsibilities, and as such, it is not a question about transparency, but it is a division of labor. However, if, for some reason, risk assessors decide on the acceptability of risks they should, one may argue, provide the reasons for accepting the risk and this is obviously a case of transparency. Have they simply thought the risk so small that it is negligible no matter what, or have they made a trade-off with the benefits involved? One can assume that both the policy makers and the public would be interested to know.

The case of ruling out risk falls into one, or perhaps both, of these problematic situations. Although we suspect that the most reasonable way to deal with this kind of risk assessment is to view it as of the former kind above—that the decision whether a risk should be ruled out or not is left to decision makers—we will accept that in reality, systematic reviewers making meta analyses do rule out risk and consequently we will treat it as a matter of transparency.

Transparency

It is generally assumed that transparent reporting from decision-making institutions is necessary for maintaining public trust in these institutions (Cowhey 1993, Ball 2009, Way, 2017). There is a rich set of examples of situations where adverse consequences of a decision has led to diminished trust as it turns out that decision makers have had more knowledge of the risks than what has been reported to the public (Millstone and van Zwanenberg, 2000).

Could there be too much transparency? A study by Löfstedt and Way (2016) asks whether the disclosure of raw data ("fishbowl transparency") is the kind of transparency that informs the public in the right way. Without contextualization, it is not clear that the knowledge of such data leads to more informed decisions by the public, and neither that trust in the decision making institutions increases (Löfstedt and Way, 2016). However, one can assume that if the demand for open raw data is driven by a conviction then such openness leads to better research—and more informed policy decisions made by decision makers familiar with the science—a case might be made (Goldacre 2013). There may be a difference in what relevant transparency includes if the communication is from decision makers to the public or if it goes from scientists to decision makers (Fischhoff and Scheufele 2013). A basic insight is that there should be an idea about how the disclosed information will inform decision making, not just that any information will do it.

Guidelines for reporting medical research results

To promote accurate and transparent reporting of medical research results, several guidelines have been developed. The most important are freely available on Equator Network (www.equator-network.org). Two such influential reporting guidelines, including extensions, are of particular relevance here: the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (www.prisma-statement.org) and the Consolidated Standards of Reporting Trials (CONSORT) statement (www.consort-statement.org).

PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. The focus is on the reviews of randomized and controlled trials (RCTs), but PRISMA can also be applied to systematic reviews of other types of research. It should be noted that the Cochrane Collaboration—which can be conceived as the central organizational body of the paradigm of Evidence-Based Medicine—has endorsed PRISMA since it was released in 2009. Furthermore, the methodological standards, as state by Methodological Expectations of Cochrane Intervention Reviews, are compliant with PRISMA (http://methods.cochrane.org/mecir).

CONSORT encompasses a variety of initiatives within the CONSORT group. One aim is to alleviate the problems arising from inadequate reporting of randomized controlled trials. The main product of CONSORT is the CONSORT Statement, an evidence-based, minimum set of recommendations for reporting randomized trials. It is a standard for preparing reports of results to ensure complete and transparent reporting, and thereby facilitating critical appraisal and interpretation. The extensions of the CONSORT statement give additional guidance for RCTs with specific designs, data, and interventions.

The CONSORT statement (Moher et al. 2010) establishes that "Without transparent reporting, readers cannot judge the reliability and validity of trial findings nor extract information for systematic reviews." In a similar vein, the PRISMA statement (Liberati et al. 2009) affirms that "systematic reviews should be reported fully and transparently to allow readers to assess the strengths and weaknesses of the investigation." These documents have brought an improved structure to the reporting of treatment effects. This has made it possible for the reader to assess whether the results can be trusted as well as to decide whether the conclusions are valid and relevant. However, when the research question is about ruling out risks of harmful effects, these statements, including extension, and other relevant guidelines for reporting (Ioannidis et al. 2004; Lang and Altman 2013; Liberati et al. 2009; Moher et al. 2010; Zorzela et al. 2016) are at a loss. Therefore, we suggest that the CONSORT and PRISMA statements should be revised to deliver improved transparency about specifying and justifying risk margins.

An illustration

Flour is fortified with folic acid in some countries to lower the incidence of neural tube defects. In some countries, decision makers have hesitated to do this, since there are concerns about the cancer risks of folic acid. In the systematic review by Vollset et al., it is concluded that the "… meta-analysis rules out moderate increases in overall cancer incidence from folic acid supplementation" (p. 1 035, our emphasis) (Vollset et al. 2013). A higher incidence rate ratio (RR = 1.06) is indicated for those exposed when compared with placebo, but the confidence interval ($CI_{95\%}$) covered "no difference" (0.99–1.13).

However, in addressing the "ruling out" of a "moderate risk," the authors need to be specific. No risk that lies within the CI can be ruled out. This implies that the authors must assume that a "moderate risk" is RR $>$ 1.13, but this is neither explicitly stated nor justified. This is the kind of case we are interested in: one in which inexplicit risk margins with epistemic consequences are at play.

The conclusions made by Vollset et al. have not been uncontested. For example, Miller and Ulrich (2013; p. 974–5), in a critical commentary, argued that their findings can lead to "… excess intake of folic acid, either from supplements or through fortification", and that data should be

viewed with caution. Miller and Ullrich pointed out that the level of statistical significance ($p < .05$) is arbitrary, that the lower CI bound was close to 1.00, and that the overall RR increased with 6%. The results in Vollset et al. are in other words fragile. This illustrates the necessity to be clear regarding the reasons for ruling out risks.

One mistake is that of wrongly "ruling out" the presence of a relevant ("moderate") RR. Another is that of concluding that an estimated and relevant RR is present while it in reality is not. In the review, a conventional $CI_{95\%}$ was selected for the pooled outcome. Given the result, this means that only a risk higher than 1.13 can be ruled out. It is important to see that Vollset et al. here in effect, although not explicitly, have made a value judgment that a "moderate risk" must be higher than 1.13. This may be a perfectly decent margin but it needs to be justified. We do not question the conclusions as such drawn by Vollset et al., but question only their lack of transparency.

Methodological choices of this kind should be explicitly stated, explained, and justified, since they may alter results, conclusions, and in some cases, policy decisions. The lack of such transparency makes it difficult for the decision makers to balance benefit and harm. This is because it is not clear why risks were ruled out. Was it because there is a pre-set level of moderate risk that lies somewhere above 1.13?; or was it because the article authors themselves judged this to be a reasonable level? In the first case, it should be clearly stated that there is such a pre-set level. In the second case, there needs to be a justification of the choice of level. Since we have not found any evidence of a pre-set level, we assume that the article authors have made their own valuation of what risks can be accepted relative to the positive effects of the treatment.

Lessons from non-inferiority trials

A problem of deciding on non-inferiority of a new treatment compared with an old one is similar in nature to the one met in the ruling out of risk (Walker and Nowacki 2011). In this case, the need for transparency is acknowledged explicitly in the CONSORT statement (Piaggio et al. 2010, 2012). The margin of non-inferiority must be specified and justified (see item 7a). Non-inferiority is illustrated in a simplified way in Figure 1. The vertical line represents "no difference" between alternative treatments. Horizontal lines are CIs with point estimates shown as black squares. The NI margin indicated by Δ and R is the region of appreciable harm with no clinical relevance. A point estimate located to the left of "no difference" indicates that the evaluated treatment is favored.

In case A, the evaluated treatment is *superior* to the control alternative. In B, "no difference" is included but not Δ, which is why the treatment is *non-inferior* to the control, and it can be "ruled out" that the treatment is *inferior* to the control. Result C is *inconclusive*, since "no

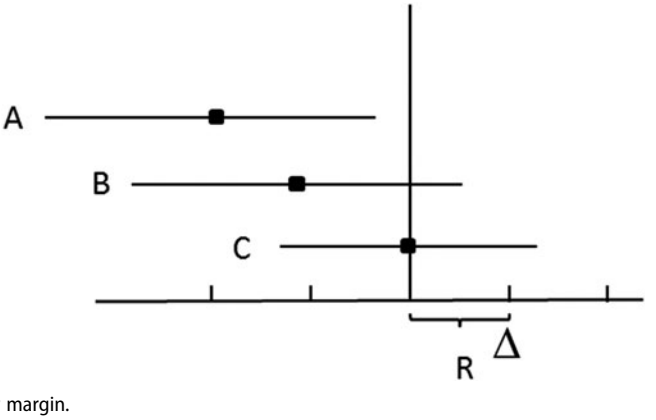

Figure 1. Non-inferiority margin.

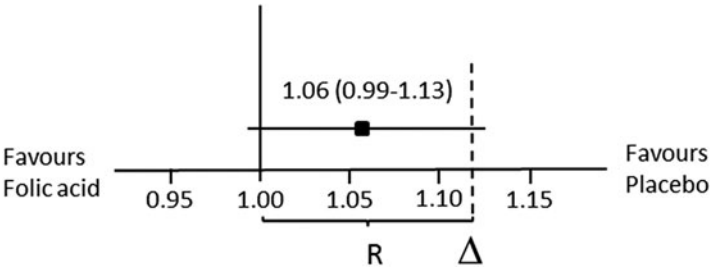

Figure 2. Possible margin of acceptable risk.

difference" and Δ are included in the CI, meaning that neither *superiority* nor *inferiority* can be "ruled out."

Although the folic acid case (Vollset et al. 2013) is not overtly described as a non-inferiority problem, Δ can be viewed as a risk margin below which risks are clinically irrelevant. To rule out the risk, in the case of folic acid, would then be similar to conclude that folic acid supplementation is non-inferior to placebo. Consequently, whether a risk can be ruled out is determined by the level of the CI and the choice of Δ. If a $CI_{95\%}$ and a $\Delta = 1.14$ are chosen, the relevant risk can be ruled out in the review since the upper CI bound is below 1.14. On the other hand, $\Delta = 1.12$ would mean that the risk cannot be ruled out since the CI would include Δ. This is illustrated in Figure 2.

If the folic acid review was an application of non-inferiority, the choice of margin would have been explained and justified according to the CONSORT statement regarding non-inferiority trials (Piaggio et al. 2010). However, interpreted as a question of ruling out a risk, this transparency is not required in other relevant guideline documents (Ioannidis et al. 2004; Lang and Altman 2013; Liberati et al. 2009; Moher et al. 2010; Zorzela et al. 2016). Furthermore, ruling out a risk of harm is generally not understood as a non-inferiority problem, and is not mentioned in the CONSORT statement on non-inferiority (Piaggio et al. 2012).

However, to specify and justify the risk margin, when ruling out a risk, is just as important as being transparent when it comes to calculation of statistical power regarding a beneficial treatment outcome. While power is well-covered in most relevant guidelines (Ioannidis et al. 2004; Lang and Altman 2013; Liberati et al. 2009; Moher et al. 2010; Zorzela et al. 2016), including *NI* guidelines (Piaggio et al. 2010, 2012), the justification of risk margins is not.

In most studies of treatment effects, PRISMA (Liberati et al. 2009) and CONSORT (Moher et al. 2010) statements are used by authors and editors, also when ruling out risks. This is a problem, since it is not imperative in these guidelines to provide detailed information justifying the choice of relevant risk margins. We have illustrated this problem with the folic acid case (Vollset et al. 2013).

Discussion

No one in their right mind would tolerate a team of researchers who allow their preferences for a certain outcome to guide their methodological choices so that the preferred outcome is guaranteed. This is considered to be an illegitimate way of making value judgments in the research process. It is liable to invite the charge of scientific misconduct. With or without such an illegitimate intrusion of preferences, value-based methodological choices are problematic and need to be understood.

The unavoidable selection of a significance level (α) is one example of a methodological choice based on a value-judgment, and clearly it has possible epistemic consequences, since α affects the length of the CI. The usually implicit choice of risk margins is another example. As Kelly et al. (2015) emphasize (p. 2), values judgments are integral to evidence-based medicine (EBM), and therefore "the highest standards of EBM require values to be made explicit".

Central to risk analysis, in any area, is of course how, who's, and what values are assigned to possible outcomes. For example, can an outcome be so cataclysmic, so negative, that nothing else counts? In most decision situations, we want to weigh the good against the bad, analyze the pros and the cons. Taking complex decisions, doing serious risk analysis, involves many value considerations, of different kinds and at different decision nodes in the process, besides assigning values to outcomes.

Therefore, it could be argued that no Δ values should be specified, and that only CIs should be reported. This would leave the choice of a proper balance of expected benefit and possible harm to the decision maker. Mulla et al. (2012) contend, with reference to NI-margins (thresholds), the following:

> In interpreting noninferiority thresholds, we will encourage you to use your own judgment rather than accepting that of the investigators, relieving you of the need to decipher what many may experience as obscure statistical reasoning used to define the thresholds.

We do not intend to discourage any local choices of threshold, but whenever the authors make conclusion regarding ruling out risks, transparency and detailed specifications are necessary. This concerns both Δ and α values (determining the level of confidence of the intervals). Such a transparency does and should not stop local decision makers to use their own specifications of Δ. On the contrary, we think that transparency will make implicit value judgments explicit and thereby demystify statistical reasoning, and also encourage local choices of Δ as well as α. Explicit statistical reasoning is rather the opposite of obscurity, but basic statistical literacy is required.

Conclusion

None of the prominent reporting guidelines provide sufficiently specific guidelines for ruling out risks. This opens up for making non-justified, value based, judgements about risk margins and CIs. As is shown in the example of the meta-analysis on folic acid cancer risk, an arbitrarily set level of acceptable risk can have wide reaching consequences when it is taken over to societal decision-making. Unless there is an absolute threshold below which a risk is judged negligible (a *de minimis* risk), the decision to accept or reject a risk is made in relation to possible gains. This is a valuation that society should make. If decision makers simply accept the valuation made by science this could lead to unacceptable risk taking, from society's point of view (Sahlin and Persson, 1994). Transparency is needed here, from science to policy, for decision makers to be able to disentangle uncertainty from implicit or explicit thresholds (Fischhoff and Davis, 2014).

The basic insights of non-inferiority trials should be applied when the medical research question concerns ruling out risks. Like a non-inferiority margin, the smallest acceptable risk should be specified and shown to be justified. To enforce transparent reporting, the PRISMA and CONSORT groups should therefore consider revising or adding some items in their statements.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Riksbankens Jubileumsfond.

ORCID

Johannes Persson http://orcid.org/0000-0003-4568-1850

References

Akl, E., L. A. Kahale, F. Sperati, I. Neumann, N. Labedi, I. Terrenato, M. Barba., et al. 2014. "Low Molecular Weight Heparin Versus Unfractionated Heparin for Perioperative Thromboprophylaxis in Patients with Cancer." *Cochrane Database of Systematic Reviews* 6: 1–80.

Ball, C. 2009. "What Is Transparency?" *Public Integrity* 11 (4):293–307. doi:10.2753/PIN1099-9922110400.

Cowhey, P. F. 1993. "Domestic Institutions and the Credibility of International Commitments: Japan and the United States." *International Organization* 47 (02):299–326. no. doi:10.1017/S0020818300027958.

Fischhoff, B., and D. A. Scheufele. 2013. "The Sciences of Science Communication." *Proceedings of the National Academy of Sciences* 110 (Suppl 3):14031–14109.

Fischhoff, B., and A. Davis. 2014. "Communicating Scientific Uncertainty." *Proceedings of the National Academy of Sciences* 111 (Suppl 4):13664–13671. doi:10.1073/pnas.1317504111.

Goldacre, B. 2013. "Are Clinical Trial Data Shared Sufficiently Today? No." *British Medical Journal* 347:1–3.

Ioannidis, J., S. J. W. Evans, P. C. Gøtzsche, R. T. O'Neill, D. G. Altman, K. Schulz, and D. Moher. 2004. "Better Reporting of Harms in Randomized Trials: An Extension of the CONSORT Statement." Annals of Internal Medicine 141 (10):781–788. doi:10.7326/0003-4819-141-10-200411160-00009.

Kelly, M., I. Heath, J. Howick, and T. Greenhalgh. 2015. "The Importance of Values in Evidence-Based Medicine." *BMC Ethics* 16 (69):1–8.

Lang, T., and D. G. Altman. 2013. "*Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines"*. Equator Network, European Association of Science Editors. www.equator-network.

Leone, M., G. Giussani, S. J. Nolan, S. J. Nolan, A. G. Marson, and E. Beghi. 2016. "Immediate Antiepileptic Drug Treatment, Versus Placebo, Deferred, or No Treatment for First Unprovoked Seizure." *Cochrane Database of Systematic Reviews 2016*. 5: 1–47

Liberati, A., D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher. 2009. "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration." *PLoS Medicine* 6 (7):1–28.

Löfstedt, R., and D. Way. 2016. "Transparency and Trust in the European Pharmaceutical Sector: Outcomes From an Experimental Study." *Journal of Risk Research* 19 (9): 1082–1103.

Miller, J. W., and C. M. Ulrich. 2013. "Folic Acid and Cancer-Where Are We Today?" *Lancet* (London, England) 381 (9871):974–976. doi:10.1016/S0140-6736(13)60110-5.

Millstone, E., and P. van Zwanenberg. 2000. "A Crisis of Trust: For Science." *Scientists or for Institutions? Nature Medicine* 6 (12): 1307–1309.

Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. G⊘tzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman. 2010. "CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials." *British Medical Journal* 340: 1–28.

Mulla, S. M., I. A. Scott, C. A. Jackevicius, J. J. You, and G. H. Guyatt. 2012. "How to Use a Noninferiority Trial: Users' Guides to the Medical Literature." *JAMA* 308 (24):2605–2611. doi:10.1001/2012.jama.11235.

Piaggio, G., D. R. Elbourne, S. J. Pocock, S. J. W. Evans, and D. G. Altman. 2010. "eAppendix: Methodological Considerations in Noninferiority Trials." *JAMA 308 (24): 2594*.

Piaggio, G., D. R. Elbourne, S. J. Pocock, S. J. W. Evans, D. G. Altman, and FT Consort Group. 2012. "Reporting of Noninferiority and Equivalence Randomized Trials: Extension of the CONSORT 2010 Statement." *JAMA* 308 (24): 2594–2604. doi:10.1001/jama.2012.87802.

Sahlin, N.-E., and J. Persson. 1994. "Epistemic Risk: The Significance of Knowing What One Does Not Know." In *Future Risks and Risk Management*, edited by B. Brehmer and N.-E. Sahlin, 37–62. Dordrecht: Kluwer Academic.

Vareman, N., and J. Persson. 2010. "Why Separate Risk Assessors and Risk Managers? Further External Values Affecting the Risk Assessor *Qua* Risk Assessor." *Journal of Risk Research* 13 (5):687–700. doi:10.1080/13669871003660759.

Vollset, S., R. Clarke, S. Lewington, M. Ebbing, J. Halsey, E. Lonn, J. Armitage., et al. 2013. "Effects of Folic Acid Supplementation on Overall and Site-Specific Cancer Incidence During the Randomised Trials: Meta-Analyses of Data on 50 000 Individuals." *Lancet* 381 (9871):1029–1036. doi:10.1016/S0140-6736(12)62001-7.

Walker, E., and A. S. Nowacki. 2011. "Understanding Equivalence and Noninferiority Testing." *Journal of General Internal Medicine* 26 (2):192–196. doi:10.1007/s11606-010-1513-8.

Way, D. 2017. "Transparency in risk regulation: the case of the European Medicines Agency." Dissertation, Kings College, London. http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.718623.

Zorzela, L., Y. K. Loke, J. P. Ioannidis, S. Golder, P. Santaguida, D. G. Altman, D. Moher, and S. Vohra. 2016. "PRISMA Harms Checklist: improving Harms Reporting in Systematic Reviews." *British Medical Journal* 352 (157):1–17.