Practical Challenges and Solutions in Meta-Analysis: An Example From the Pain Medicine Literature

Trevor Thompson

University of Greenwich, UK

Discipline

Medicine [D23]

Sub-discipline

Anesthesiology & Pain Medicine [SD-MD-2]

Academic Level

Postgraduate

Contributor Biography

Trevor Thompson in an associate professor in clinical research at the University of Greenwich, London, who specializes in pain research. He has published numerous articles in high-impact clinical journals such as *JAMA Psychiatry*, *Neurology*, and *Pain*. He is an associate editor of *BMC Psychology*, and his work has been covered by media outlets including the *Independent*, *Daily Mail*, and the *Express*.

Published Articles

Thompson, T., Terhune, D. B., Oram, C., Sharangparni, J., Rouf, R., Solmi, M., . . . Stubbs, B. (2019). The effectiveness of hypnosis for pain relief: A systematic review and meta-analysis of 85 controlled experimental trials. *Neuroscience and Biobehavioral Reviews*, *99*, 298–310. https://doi.org/10.1016/j.neubiorev.2019.02.013

Abstract

Meta-analysis can provide a reliable overall estimate of the effect of an intervention by combining all of the available individual study evidence. Although there are an ever-increasing number of published meta-analyses, journal word limits mean that the numerous practical challenges involved in producing the final review are usually not documented. This case study is an attempt to identify these types of challenges using a concrete example of a real meta-analysis examining hypnotic interventions for pain. Specific strategies that can be used to help tackle these challenges are outlined, including preparation of a detailed study protocol, how to approach tricky screening and extraction decisions, computing effect sizes when data are not presented in conventional form, and the importance of assembling a suitable research team. This case does not describe the methodological principles of meta-analysis for which there are many excellent existing texts. Instead, it can be used as a complementary guide to these texts as it highlights the type of practical challenges that we faced and suggests possible ways to combat them that we hope will be useful to others.

Learning Outcomes

By the end of this case study, students should be able to

- Understand why preparing a detailed protocol is critical to a successful review and meta-analysis
- Understand the practical challenges that are likely to be faced during the review process
- Describe a range of remedial strategies that can be used to confront these challenges

Case Study

Project Overview and Context

"Opioid crisis" is one of several factors contributing to the increased interest in nonpharmaceutical treatments for pain. Most people have heard stories of how hypnosis has resulted in dramatic relief from pain at one time or another. I remember reading a newspaper article many years ago of a man (presumably a surgeon) who used self-hypnosis as a method of pain control while he performed a vasectomy on himself—surely the ultimate test of a person's confidence in their abilities. We were interested to explore whether tales of hypnotic analgesia (pain relief) were purely anecdotal or were supported by credible scientific evidence, and we wanted to examine this in studies of experimentally induced pain in healthy people. We also wanted to determine not only *if* hypnosis worked but also to quantify *how well*, to allow some preliminary conclusions on whether hypnosis offered a realistic alternative to medication. A systematic review and meta-analysis was our chosen methodology to do this, as it provides robust conclusions based on a synthesis of all of the available evidence.

This case is not intended to be a step-by-step guide on how to do a systematic review and meta-analysis. This would be beyond the scope of this case and would duplicate material already covered extensively in a range of excellent texts (Borenstein et al., 2009; Card, 2012; Cooper et al., 2009; Higgins & Green, 2008; Lipsey & Wilson, 2001; Littell et al., 2008). Instead, this is a short case study based on our meta-analysis of hypnosis and pain which has the goal of highlighting the type of practical challenges and decisions we faced and that are typical of this type of methodology, but that are rarely documented in published studies due to journal space constraints. The case study is likely to be most beneficial to those with a little exposure to the basic principles of meta-analysis (e.g., random effects, forest plots, publication bias), but who are relatively inexperienced in the actual practicalities of

conducting one. The specific aim is to illustrate the type of common practical problems faced so that anyone intending to conduct a meta-analysis can be aware of them and be armed with some ways to approach them.

Section Summary

- Hypnosis is often suggested as a means for reducing pain, but there is little established consensus on its effectiveness.
- We chose to conduct a systematic review and meta-analysis using all of the available evidence to address this question.

Research Design

The basic idea behind a meta-analysis is pretty simple. Identify the studies relevant to your research question and take the (weighted) average of the individual effect sizes to get a reliable overall estimate (in this case, of the average pain relief resulting from hypnosis). In reality, however, there are numerous design challenges that must be confronted to get to this point, and consequently numerous decisions to make along the way. As such, it is important for the authors of any meta-analysis to prepare a protocol that describes how they will conduct their review—for example, What interventions will be included? What primary outcomes will be looked at? How will relevant articles be identified? What type of analysis will be performed?

We designed our research protocol following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol guidelines or PRISMA-P (Moher et al., 2015). This is a checklist of 17 items intended to "facilitate the preparation and reporting of a robust protocol for the systematic review" (p. 1, Moher et al., 2015).

Intervention: What Did We Consider to Be (and Not to Be) a Hypnosis Intervention?

A definition of the intervention was especially pertinent in this field where there is no universally agreed consensus on what constitutes hypnosis. Based on existing key articles, we considered a hypnotic induction to be "a suggestion offered to another person to alter perceptual experience and voluntary action that typically involves relaxation, focused attention and imagery." We excluded hypnotic states induced by pharmacological agents (e.g., ketamine) or where there was co-administration with another intervention.

Control Condition: What Are We Judging the Effectiveness of Hypnosis Against?

Another intervention? A placebo? Nothing? As we were primarily interested in whether and to what extent hypnosis had painkilling effects, given that there appeared to be no consensus on this issue, we used a placebo/no intervention comparator. If there had been persuasive preexisting evidence that hypnosis was effective, we may instead have chosen another intervention or a usual care group (where the patient receives the care usually administered in everyday practice) as a comparator to examine which was superior.

Outcome: What Assessments Should We Include as Acceptable Measures of Pain?

We included pain ratings (e.g., 0–10) as a key outcome as they offer a clinically meaningful measure of pain, with various mapping studies showing specific score ranges can be roughly mapped onto verbal descriptors of intensity, for example, 0 to 4 = *mild pain* (Boonstra et al., 2016). We chose not to include physiological measures (such as heart rate)

as these currently do not provide accurate quantitative assessments of the magnitude of pain experienced.

What Clinical Characteristics Might Influence Treatment Effectiveness?

Logically, people high in hypnotic suggestibility should engage better with hypnosis, and thus hypnosis interventions may be more effective in these individuals. However, we were also aware that studies were likely to use a variety of different scales to measure suggestibility, and thus to include them all in the same analysis we had to have a plan for mapping them onto a single common scale. If suggestibility (rather than pain) had been an outcome variable, we could have solved this problem by computing the mean difference between the two groups for each study and then dividing by that study's standard deviation—that is, creating a standardized mean difference. However, we wanted to use suggestibility as a moderator/predictor variable. Unfortunately, when working with only a single (rather than a difference) score for each study for use in moderation analysis, a standardized mean difference cannot be computed. Fortunately, a review of the literature indicated that most scales provided cutoff scores based on normative data that can be used to classify suggestibility as low, medium, or high, and this provided the basis for a common metric (it is also typical for hypnosis studies to report the sample range). For example, despite differences across the Carleton, Stanford, and Harvard scales, score ranges of 0 to 2, 0 to 3 and 0 to 4, respectively, could be used to make a common classification of low hypnotic suggestibility.

This last issue in particular illustrates one of the main pitfalls of secondary data analysis—you have absolutely no control on what was done in the primary study or how it was done. As a consequence, there are numerous different and equally valid decisions that can be made in terms of potential courses of remedial action. When making these design decisions, it is important to make sure that the clinical research question is always the

primary driver (which is probably best exemplified above with reference to the selection of a control group). Pragmatic considerations of course are also important—it is no use deciding to accept only randomized controlled trials of hypnosis interventions that use virtual reality imagery if no such studies exist—but it is the consideration of the research question that should be paramount.

Section Summary

- It is essential to prepare a detailed protocol of how you will conduct your systematic review and meta-analysis.
- The PRISMA-P statement provides a checklist of important considerations that can help you prepare your protocol.
- Using secondary data means you have no control over how the studies you will use have been performed, and tackling the problems this invites is one of the biggest challenges in a systematic review.

Research Practicalities

Although the protocol documents key research design issues, there are also a number of practical challenges involved in conducting a systematic review and meta-analysis. We faced several such challenges, and below we describe two that will be familiar to anyone who has previously undertaken a meta-analysis.

First, the process was extremely time-consuming. We were pleased that we had 85 studies, which is higher than average, as this ensures more robust conclusions can be drawn. The downside of course is that this means more time is spent extracting data. Inevitably, and as always, the project took longer to complete than we anticipated. We partially offset the extra time needed by recruiting additional researchers, although this did tend to cause extra complications in managing the timing of everyone's contributions to make the project work

efficiently. For example, if I had put aside time to do my analysis in March and data extraction had not finished until May, then it was likely to be some time before I was free to do the analysis—and this is exactly what happened.

Second, we could not get hold of all the data we needed to compute effect sizes using standard formulae. This was the case for around 10% of eligible studies and this is fairly typical for the reviews I have been involved in. Sometimes data were missing due simply to not being able to retrieve the article at all. Sometimes we could retrieve the article, but the means, standard deviations, and sample sizes we needed to compute the mean difference effect size (and its variance) were not directly reported and were often only in graph form. There were also a few instances where we had concerns over the reliability of the data, such as discrepancies across tabular and graphical presentations of the same data. The most obvious solution to these problems is to contact the study authors to obtain the articles or to provide clarification. Such attempts met with limited success, however. Many studies were published a long time ago, with a few articles 50 years old in a time when computers ran off the taps and email would have been regarded as some sort of witchcraft (OK perhaps I'm exaggerating this somewhat). Other study authors seemed to have moved to other institutions, had retired, or quite possibly deceased. Of the 20 authors we did try to contact, only six replied. How we tackled some of the issues this caused are described in the Practical Lessons Learned section.

Section Summary

- A systematic review and meta-analysis can be extremely time-consuming and you must be prepared for this
- Missing data is one of the most common problems faced when conducting a meta-analysis.

Method in Action

The process of constructing a research protocol helped us to anticipate many of the problems that confronted us, and so when they arose, we had a ready-made plan for dealing with them. For example, we had already considered how we would deal with the fact that different studies would be likely to use different instruments to quantify hypnotic suggestibility. As such, we were able to simply implement what we had already carefully decided on without needing to delay the project while we worked on a solution or worry that after spending months extracting data there was no resolution to the problem. Also, by carefully considering beforehand exactly what we would accept as a "hypnotic intervention" helped us to decide whether to include a study in some of the more ambiguous cases. For example, some studies used hypnotic-like visual imagery but without the hypnotic induction procedure that typically involves deep relaxation and focused attention, and thus we could confidently exclude these studies.

You cannot anticipate all of the challenges, however, and several aspects of the review process did not work so well. One such issue was the difficulty of deciding whether the control group used in two specific studies was consistent with the protocol definition of a "no active intervention" control group. One of these studies asked participants in the control condition to read, whereas another asked them to relax. Although this hardly seems to fit the definition of active interventions, reading may provide a source of cognitive distraction that reduces pain and relaxation could even be an important component of the analgesic effect of hypnosis. In the end, we opted to include these two studies as our consensus was that if any painkilling effects of these control procedures were present, they were likely to be minimal.

A second issue was related to the way in which outcomes were assessed. Some studies compared pain ratings for hypnosis versus control, but used a tolerance paradigm to do so. This means that the participant can terminate the stimulus at any time when they can no

longer tolerate it and the length of time they have kept their hand in, for example, a vessel of ice cold water is used as a pain outcome. It is, of course, completely reasonable and indeed ethically necessary to allow participants to withdraw from a pain stimulus at any time, but the use of both tolerance and pain ratings in the same study can invite problems. Specifically, if one group had a shorter pain exposure (tolerance) than time, this, rather than the intervention itself, may have affected their pain rating. It is precisely this type of issue that is difficult to anticipate unless you are familiar with experimental pain testing, but such issues should not be ignored. We decided to retain these studies as they all reported longer pain exposure times for hypnosis, and therefore if hypnosis *still* resulted in lower pain ratings, this would provide good evidence that hypnosis was effective *despite* the greater exposure (if the hypnosis group reported lower pain ratings but were exposed to the pain stimuli for less time than the control group, this would have been trickier to deal with).

A third issue concerns the categorization of hypnotic suggestibility. Although we had a plan for converting the scores from different suggestibility scales used by the various studies to a common metric of low, medium, and high suggestibility, we had several difficulties in implementing this. Some studies reported score ranges that did not fit neatly into the classification system. For example, some studies reported their participant scores ranged from 0 to 4 on the Stanford scale (where 0–3 is the range that represents low suggestibility on this instrument). How exactly should we classify this—"low with a bit of medium suggestibility"? Similarly, a few studies did not report the range, but instead reported the mean suggestibility score, so we could not be sure that the range of scores did not span multiple classifications. What about these studies? We could have left them out, but this would reduce the power of analysis and leave a remaining set of studies that could be potentially unrepresentative. Ultimately, we decided to classify the suggestibility of the samples in these studies to their closest approximation, so when a range of 0 to 4 is the

classification guideline for low suggestibility, for example, score ranges of 0 to 5 or a mean score of 3 was classified as low.

The key lesson to be learned from these examples is that while a clearly defined protocol will provide a clear framework that helps you make many decisions, there will always be aspects that do not fit neatly with the protocol, and they can be difficult to anticipate in advance. In these instances, some sort of decision still has to be made, and in the next section, some guidance is provided on strategies for approaching this type of ad hoc decision-making.

Section Summary

- A detailed protocol can help you deal with many challenging problems that arise during the review process.
- Even with a well-constructed protocol, not all problems can be anticipated in advance, and ad hoc decisions will often have to be made during the process itself.

Practical Lessons Learned

In addition to the task of learning the underlying principles of meta-analysis, it is evident from the previous sections that there are also a number of practical challenges that must be confronted. During the course of our hypnosis and pain meta-analysis, and several other reviews, we have adopted some general strategies for dealing with these types of challenges, and we describe these below.

Prepare a Protocol

It should be clear by now to prepare a well thought-out and detailed research protocol that describes how you intend to conduct the review is critical to the success of the project. This leads to a better review and helps you identify what the important issues are before you begin. It also promotes transparency in research practice, and you can also register your protocol for free in a repository where it can be freely accessed by anyone who wants to read

it. An example of such a repository for research in health-related outcomes is PROSPERO, which is a searchable database that provides the current status of a review (provided this status is kept up-to-date by the review authors). Registration helps avoid the study being unnecessarily duplicated by others, although it can also be disheartening to find someone else has registered your research idea (nevertheless, this would still seem preferable to spending months conducting your study only to find the same review published by others just as you complete your review). Registration also reduces the potential for bias by allowing anyone to compare the methods described in the final review with those described in the protocol. If substantial protocol variations are present but with no apparent justification, it can raise suspicions about the integrity of the review. In particular, it can prompt concerns that numerous analytic data fishing expeditions may have been attempted until a significant effect was found and that this could be the reason for the protocol change.

Decision-Making

The examples given in the previous section illustrate something which is extremely common when conducting a review—it is straightforward in principle, but rarely so in practice. Even with a really well thought-out and detailed protocol, there will inevitably be unanticipated challenges, and decisions have to be made on how to handle these. It is important to realize that there is often no gold standard solution out there, and so there is no reason to stress if you cannot identify one—such decisions will always possess an element of subjectivity.

In the process of carrying out the hypnosis and pain meta-analysis, and other reviews, I have found a three-step procedure to be useful when making these types of decisions. *Step 1*: Make sure that you justify the decision you have made. State why it is a sensible, logical and ideally the best way to handle the problem. *Step 2*: Be completely transparent in exactly how you have implemented your decision in practice. This ensures that even if someone disagrees

with you, there is complete transparency in what you have done. *Step 3*: Conduct sensitivity analysis to examine the impact of your decision. This means to perform the meta-analysis on the data both before *and* after your decision has been implemented. If there is little substantive difference in effect size, then there is no compelling reason to suspect your decision-making has had a significant impact on the overall findings. In our review, relaxing the criteria for classifying low, medium, and high suggestibility had a negligible impact on effect size estimates, but it allowed us to increase the number of studies we could include in the analysis from 40 to 67 and thus increased the precision of the estimates. If of course your decision *does* have a substantive impact on the effect size, then you can speculate upon why this might be, or even reconsider your decision.

Work in a Research Team

The familiar idiom of "a problem shared is a problem halved" can be applied to the review process! Meta-analytic reviews typically take a long time, and sharing the load with a (carefully selected) research team means the project is completed more quickly, more efficiently, and is ultimately just more enjoyable.

Our research team did not initially include a subject specialist in hypnosis. Although I am very familiar with the experimental pain literature and have a decent level of experience with pain management strategies in general, I am not an expert in hypnosis and I struggled with making informed decisions in this area. As a result, I contacted Devin Terhune, who works at a London University just up the road from mine and asked him whether he would like to be involved. This proved to be incredibly useful, not only for helping to make informed decisions but also for having another person to discuss decisions made in other areas which was very reassuring (he also introduced me to a café with excellent coffee and chewy ginger biscuits). Everybody benefits from this arrangement—your project is better and more enjoyable, and your collaborators benefit (hopefully) from a publication.

It is, nevertheless, important to be careful in how you select your research team. Try to include those with skills that are likely to be useful to the review, especially when these are skills you may not necessarily have. Try also to invite those who you can trust to contribute diligently to the screening and data extraction process (the most credible reviews have two independent screeners and data extractors). The team would ideally also include someone with prior experience of conducting a review and/or meta-analysis, a subject expert, someone with appropriate statistical skills, and other competent individuals to help with the screening and data extraction. It is usually possible to find people who can assist with screening/extraction, and most people know someone who is good at stats. Finding a subject expert or someone with previous meta-analysis experience may be less easy, and this may be a good time to speak to your mentor or an academic colleague for advice. What is important when involving others is to make it clear in advance exactly what you are expecting them to do.

Missing Effect Size Data

When the data you need to calculate effect size are not directly available, we found there are often other ways to compute the effect size. As discussed earlier, it was not uncommon for us to find that the means and standard deviations we needed to calculate effect size were often reported in graphical format. There are a number of software applications that can extract these values from a bar chart for you, and we found the online program WebPlotDigitizer (https://automeris.io/WebPlotDigitizer) to be by far the best software for doing this and is also completely free; just make sure that the bar chart is displaying standard deviations and not standard errors (see Nagele, 2003). In addition, even if the values you need to compute effect size using standard formulae are not available, it is often possible to compute the same effect sizes with test statistics such as *p* and *t* using alternative formulae. A

classic source of these alternative formulae is provided by Lipsey and Wilson (2001) and several online calculators (see the "Web Resources" section).

Section Summary

- Decisions should be justified and clearly described, and sensitivity analysis should be conducted to examine the influence of the decisions on the results.
- Systematic reviews and meta-analyses benefit enormously from the use of a carefully selected research team.
- When conventional data used to calculate effect sizes are not available from articles, there are often other ways to compute these.

Conclusion

A systematic review is a highly effective method for tackling a clinical research question by synthesizing all of the available evidence, and a meta-analysis provides an excellent means of quantifying this evidence. However, the process of conducting a systematic review and meta-analysis presents significant challenges, and it is important to be as prepared as possible for these before undertaking the review. This typically involves the preparation of a protocol, assembling a suitable research team, putting aside a significant amount of time to complete the review, and being as rigorous and transparent as possible in your decision-making. Despite the challenges involved, a properly conducted systematic review and meta-analysis has the potential to be a highly influential piece of research that can represent an important evidence base for researchers, clinicians, and policymakers alike.

Classroom Discussion Questions

1. Choose a clinical research question you are interested in and provide a few sentences for each of Items 7, 8, 9, and 13 from the PRISMA-P checklist (see the Web Resources section).
2. Imagine you have identified 20 eligible studies for your systematic review, but only half of these report the means and standard deviations you need to calculate the effect sizes you want. Develop a hierarchical strategy that describes how you would tackle this issue by listing the courses of remedial action you would attempt in order of priority.
3. Choose any one of the three challenges described in the Method in Action section, suggest an alternative approach that could have been taken, and justify why this approach would have been equally reasonable (or better).

Multiple Choice Quiz Questions

Which of the following is *not* an advantage of using a pre-prepared protocol?

a. Potential problems can be identified before the review has begun
b. The review procedures can be more easily standardized across different members of the review team
c. Research bias is eliminated

Why can it be problematic when different studies included in a review use different measurement scales?

a. Scores cannot be easily combined when they use different measurement metrics
b. These studies are also likely to vary on other important dimensions
c. It could indicate that the quality of studies included in the review may be poor

What are the potential advantages of working in a research team?

a. The project is likely to be completed more quickly
b. The review is likely to be enhanced by having a range of researchers with different specialist skills
c. Both of the above

Imagine you stated in your protocol that you would only include studies that diagnosed depression using the DSM (*Diagnostic Statistical Manual*). What would *not* be an acceptable course of action for a study that used the ICD (International Classification of Diseases) as a classification tool?

a. To include the study in your review
b. To include the study but perform sensitivity analysis
c. To exclude the study from your review

Declaration of Conflicting Interests

The Author declares that there is no conflict of interest.

Further Reading

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Cooper, H., Hedges, L., & Valentine, J. (2009). *Handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Web Resources

Online effect size calculators:

- https://automeris.io/WebPlotDigitizer/
- https://www.meta-analysis.com/pages/tutorials.php
- https://www.psychometrica.de/effect_size.html
- http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php
- https://www.uccs.edu/lbecker/

The PRISMA-P checklist: http://prisma-statement.org/Extensions/Protocols.aspx

References

Boonstra, A. M., Stewart, R. E., Köke, A. J., Oosterwijk, R. F., Swaan, J. L., Schreurs, K. M. G., & Schiphorst Preuper, H. R. (2016). Cut-off points for mild, moderate, and severe pain on the numeric rating scale for pain in patients with chronic musculoskeletal pain: Variability and influence of sex and catastrophizing. *Frontiers in Psychology, 7*, Article 1466.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Card, N. A. (2012). *Applied meta-analysis for social science research*. Guilford Publications.

Cooper, H., Hedges, L., & Valentine, J. (2009). *Handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions*. Wiley.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. SAGE.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*, 1.

Nagele, P. (2003). Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *British Journal of Anaesthesia, 90*(4), 514–516.