A Novel Approach to Synthesize the Evidence on Analgesic Adjuvants for Postoperative Pain

##### Authors

Michael H. Andreae, M.D., Department of Anesthesiology & Perioperative Medicine, H187, Penn State Health Milton S. Hershey Medical Center, Penn State College of Medicine, Hershey. PA

Nathan L Pace, MD., MStat., Department of Anesthesiology, University of Utah, Salt Lake City, UT

##### Corresponding author:

Michael Andreae, MD, Address: Department of Anesthesiology, Department of Anesthesiology, Penn State Health Milton S. Hershey Medical Center, 500 University Drive, Hershey. PA 17033, Tel: +1 (717) 531 6140, Email: [mandreae@pennstatehealth.psu.edu](mailto:mandreae@pennstatehealth.psu.edu)

##### Acknowledgement

We would like to acknowledge Doleman et al for providing the data to generate the figures.

##### Conflicts of Interest: None

**Number of words in this Editorial (1598), Figures (2)**

##### Keywords:

analgesic adjuvants, postoperative pain, meta-analysis, Bayesian statistics

## Warning: package 'pacman' was built under R version 3.3.3

# Editorial Text

Based on the best available evidence, which analgesic adjuvant is best for a particular patient to optimize postoperative pain control? For example, how much opioid sparing will the administration of intraoperative intravenous acetaminophen afford for a typical bariatric patient at Penn State Hershey or University of Utah Medical Centers? With patients and surgical procedures across the United States as diverse as species in the rain forest, this can be a difficult question. Often randomized controlled trials (RCT) yield varying estimates of the clinical effectiveness of analgesic adjuvants. Contradictory results of RCTs, seemingly at random, leave clinicians baffled. Doleman et al in the study “Baseline morphine consumption…” propose an ingenious new solution to solve this riddle1. They synthesized the evidence for morphine dose reduction with adjuvants by controlling for baseline risk (morphine consumption) across surgical procedures and patient populations. They postulate that with their novel approach, local audit data could be used to predict the expected average reduction in morphine consumption for any analgesic adjuvant. In the same breath, they modify and perhaps toss out the established paradigm of procedure-specific pain control.

We should expect the results of RCTs to vary, even if they investigate the same intervention in a similar population. Each RCT recruits patients by a convenience sample from a local subpopulation; it is not a random sample of the entire population who might receive the adjuvant. Furthermore, by pure chance alone, each RCT has a chance to over- or to under-estimate the effect of an intervention. Also, larger and smaller studies will lead to more or less precise estimates of the intervention effect. Meta-analysis pools available RCTs to synthesize the evidence for a more precise and robust effect estimate. This may reduce uncertainty in the face of seemingly contradictory results. However, if the results and studies are too heterogeneous, evidence synthesis may be inappropriate. Such an approach is frequently critiqued as mixing apples and oranges. Excessive between-study heterogeneity in meta-analysis raises concerns that the included studies are clinically and methodologically too different, making pooling all identified RCTs unreasonable. In the face of substantial between study differences in results, we should explore its possible causes2.

One approach to explain why RCTs yield contradictory results, is to group studies by surgical intervention. Stratification by surgical intervention makes clinical sense because between-study heterogeneity may be smaller within each stratum. This led to the current paradigm of procedure-specific pain control3: it is expected that different surgical procedures cause different amounts of pain. Populations undergoing different interventions for different diseases may vary in how they respond to pain, in their comorbidities, pharmacokinetics etc. Clearly, a population of elderly men following thoracotomy for lung cancer will differ from a population of young women after cesarean section. It follows that postoperative pain control should be tailored to the specific surgical intervention and the particular population. While this seems intuitive, it drastically reduces the number of available studies for clinical decision making on for a particular patient population undergoing a particular procedure. For example on thyroidectomy, Doleman et al.1 found only one single RCT investigating the effect of intravenous acetaminophen. Pooling only studies on a particular procedure limits the number of studies available and hence poses a significant challenge to any evidence-based approach to procedure-specific postoperative pain management.

A second approach to explain differences in results between RCTs is to control for baseline risk4. For studies investigating adjuvants for improved postoperative pain control, investigators frequently use the mean difference in morphine dose between those receiving and those not receiving the adjuvant to estimate the effect. The mean dose in the control group, (those *not* receiving the adjuvant), can be considered as the baseline risk for pain. Doleman et al.1 tried to control for baseline risk in a meta-regression, using Bayesian hierarchical modeling adapted to requirements of evidence synthesis5. In their models, surgical procedure was no longer the main predictor of between-study variability, but baseline risk was. Restated according to Doleman, the benefit of the analgesic adjuvant is mainly driven by baseline morphine consumption in the control group and not by the surgical procedure. Two figures using acetaminophen data as an example will illustrate their findings.

In Figure 1, we present a classical forest plot to investigate the effect of acetaminophen on postoperative morphine consumption, ordering the 25 RCTs Doleman et al.1 identified by surgical interventions.[[1]](#footnote-28) The 25 RCTs have different effect sizes; the results seem inconsistent, sometimes varying widely even within the same surgery.

In Figure 2, we rendered a schematic of the acetaminophen subplot of Doleman's Figure 21 to illustrate their meta-regression approach.[[2]](#footnote-29) Each study is represented by a colored icon. If surgery were the best way to explain why studies yield different results, then studies investigating the same surgical procedure should have similar results. Hence the same color icons should be grouped together (along the same value on the y-axis indicating the estimated effect). But studies in the same surgical group appear far apart on the y-axis. By contrast, studies with similar baseline risk (close on the x-axis, i.e. baseline morphine consumption in the control group) tend to have similar effects, (they are close on the y-axis, i.e. mean morphine-equivalent reduction), illustrated in a neat alignment along the regression line in Figure 2. Doleman's conclusion is that a meta-regression controlling for baseline morphine consumption is superior to stratification by surgery in explaining between study variance in results.

However, one would expect a larger *absolute* difference between experimental group and control group if the *absolute* mean in the control group is higher. What may matter more instead is the *relative* reduction in morphine consumption. We may hence consider the *ratio* of the mean morphine consumption in the experimental versus in the control group. While at higher baseline risk an unchanged relative reduction would always correspond to a larger absolute reduction in the outcome, this should not be misinterpreted as baseline risk explaining effect differences4. A third approach to control for baseline risk in meta-analysis of continuous outcomes is hence to pool the *ratio of (geometric) means* (RoM)9. The geometric mean is the nth root of the product of n values. Figure 3 offers graphical intuition about the *ratio of (geometric) means* (RoM) and its use to explore between-study heterogeneity in meta-regression for evidence synthesis.

In Figure 3, we present a L'Abbe plot2 to investigate the heterogeneity between studies with a meta-regression, using the RoM as a *relative* outcome measure4. The RoM involves a *logarithmic* transformation, as we plotted the means in the experimental versus the control group for each study on the *logarithmic scale*. We shape and color coded studies by surgery group and sized them by the weight studies were given in the meta-regression. Studies below the dashed diagonal demonstrated benefit. However, studies are sprayed around the black regression line, which has almost a slope of one. In conclusion, and in contrast to Doleman et alt1, this meta-regression explains only 2% of the between study variability, (and agrees with Doleman's own results when they performed meta-regression using RoM, [results not published]). Also, Friedrich et alt9 showed that meta-regression using the *relative* RoM could not explain between study variability (by reducing heterogeneity). The conclusion from the l'Abbe plot and our meta-regression is that baseline risk does not explain between-study differences in mean morphine reductions, when outcomes are expressed as a RoM.

However, if we follow Doleman et alt, and consider the ratio measures as a flawed measure of analgesic effect, the results by Doleman et alt. still hold promise. Doleman et al.1 argue that baseline risk explains effect differences for all analgesic adjuvants across all populations, for all surgical interventions, as shown in the other subplots of Doleman’s Figure 21. Their conclusions contradict the dominant paradigm of procedure-specific postoperative pain control. What is more, Doleman et al.1 also propose that clinicians can use local audit data, collected on local surgical populations, to estimate the average reduction of morphine consumption for each adjuvant. This is a great feature to have for a new unifying theory on how to synthesize the evidence for postoperative pain control. Indeed, *absolute* reduction of opioid consumption may matter more than relative *reduction*. For example, the same *relative* ratio of means of 0.5 could reflect an *absolute* reduction in morphine consumption by 50mg (if 100mg consumed on average) or 5mg (if 10mg consumed on average). Especially in the face of the current obesity epidemic, adverse events may depend more on *absolute* dose10, as patients with obstructive sleep apnea are highly sensitive to opioids.

Besides a classical model, Doleman et al.1 fit a Bayesian model that differs from the classical approach (also known as frequentist statistics). The Bayesian model incorporates prior information and combines it with the newly observed data, much like a physician would in clinical practice. All statistical models (classical and Bayesian) are based on subjective assumptions; the model choice itself is often more important for correct inferences. In classical and Bayesian statistical modeling, investigators should explore the sensitivity of results and inferences to their assumptions and model choices. In their online supplement, Doleman et al.1 provided software code, data and model details. They used very so-called "weak priors", (assumptions expected to impact results only minimally), and in their sensitivity analysis found their results to be robust. This made their Bayesian approach more transparent, sound and trustworthy.

Based on their previous work11, Doleman et alt. speculated that baseline risk might better explain variability. Baseline risk for pain may modify the treatment effect of analgesic adjuvants as a proxy for unmeasured patient-level characteristics11. Their conjecture is supported by their meta-analyses. This now suggests a set of testable hypotheses and predictions, making their proposed novel and unifying theory attractive. Progress in science and medicine is made by proposing models and hypotheses and then attempting to falsify these. Indeed, investigators can use local audit data to predict the reduction in morphine requirements and then prospectively observe the actual reduction achieved to investigate how well the Doleman model predictions hold. Meta-analysts can replicate the approach by Doleman et al.1 and for example test if it can be extended to evidence synthesis of regional anesthesia to explain and/or reduce the between study variability of the effects of regional anesthesia versus conventional multimodal pain control for postoperative pain control. If Doleman’s hypothesis holds, it would greatly simplify evidence synthesis and provide clinicians with much stronger and much more granular evidence on what analgesic to use for which patient. In fact, in the second part of their paper, Doleman et al.1 go on to use meta-regression to discuss rather refined and specific questions. For example, the question, if acetaminophen is more effective if administered intravenously rather than orally, is of great financial and clinical importance. But there are simply too few studies to address this question, if we insisted on combining only studies comparing the effects in the same population/surgical intervention. Doleman et al.1 provide a league table of efficacy of analgesic adjuvants.

Doleman et al.1 acknowledge the many limitations of their work. Their meta-regression, like any meta-analysis was an observational study, even if it had an a priori defined analysis plan. The confidence of their predictions and inferences will be weaker for ranges of baseline risk (baseline morphine consumption in the control group) where they have less studies for the particular analgesic adjuvant at hand. We will list these limitations using acetaminophen as an example:

* The range of surgical procedures included in the acetaminophen trials was limited, (see Doleman’s Table 11).
* The patient mix in the acetaminophen studies are predominantly female.
* The standard morphine dose (50 mg) at which dose reduction was estimated is much higher than the mean dose (28 mg) in the included acetaminophen studies. Inferences and predictions in the higher dose ranges therefore rely on fewer studies and/or on problematic extrapolation.
* For most adjuvants studied, there was little benefit (almost equal benefit) for studies reporting low baseline consumption, which is evidenced in Figure 2 of Doleman’s article1.
* Finally, the estimates in their regression equation apply to aggregate groups not individual patients.

We and the reader may remain somewhat skeptical and cautious about the approach and inferences suggested by Doleman et al.1 Provocative as the novel hypothesis by Doleman et al.1 may be, it would afford great utility. It also exemplifies the benefit and potential of Bayesian meta-regression to support clinical decision making.

# Figures

## Figure 1

### Figure 1 Caption

Estimates for the mean reduction of morphine equivalent afforded by acetaminophen are shown in a classical forest plot with studies grouped by surgical interventions. The mean reduction of morphine equivalents and their 95% confidence intervals were provided by Doleman et al.. The later were used to estimate the standard error (Higgins 2011). The figure was produced with the Cochrane Collaboration software RevMan (RevMan 2014). The 25 RCTs have inconsistent effects for the drug under investigation (acetaminophen), varying widely even within the same surgery.

## Figure 2

### Figure 2 Caption

In Figure 2, we reproduce the acetaminophen subplot of Figure 2 of Doleman et al. Each study is represented by a dot. We coded and colored the studies by surgical procedure, according to the adjacent legend. If surgery were the best way to explain why studies yield different results, then studies investigating the same surgical procedure should have similar effects. Hence the same shape (and color) dots should be clumped together, (and the green-square-box cholecystectomy studies somewhat are). But most studies of the same color are far apart, for example orthopedic (violet-star) or hysterectomy (blue-cross-in-box) study dots are spread out over the entire range of the plot. By contrast, most studies line up neatly around the blue regression line, when we regress the mean reduction in morphine consumption (y-axis) against the baseline risk (baseline morphine consumption in the control group) in the x-axis. The conclusion is that a meta-regression controlling for baseline morphine consumption in the control group is superior to stratification by surgery in explaining between study variance in results.

## Figure 3

### Figure 3 Caption

L'Abbe plot investigates the heterogeneity between studies using the ratio of (geometric) means (R0M), as a *relative* measure, plotting the mean morphine consumption in the experimental versus the control group for each study, on the logarithmic scale, (in mg morphine equivalent). Studies are shape (and color) coded by surgery group and sized by the weight studies were given in the meta-regression. Studies below the dashed diagonal demonstrated benefit. Studies are sprayed around the black regression line, seemingly at random. The regression line has a slope of about one, reflecting that this meta-regression explains only 2% of the between study variability. Hence baseline risk does not explain between-study differences in mean morphine reductions, when outcomes are expressed as a RoM.

# References

1. Doleman B, Sutton A, Sherwin M, Lund J, J W. Baseline morphine consumption may explain between-study heterogeneity in meta-analyses of adjuvant analgesics and improve precision and accuracy of effect estimates. Anesthesia and Analgesia 2017.

2. Song F. Exploring heterogeneity in meta-analysis. Journal of Clinical Epidemiology 1999;52:725–30. Available at: <http://www.sciencedirect.com/science/article/pii/S0895435699000669>.

3. Kehlet H, Wilkinson RC, Fischer HBJ, Camu F, Prospect Working Group. PROSPECT: evidence-based, procedure-specific postoperative pain management. Best practice & research Clinical anaesthesiology 2007;21:149–59. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17489225>.

4. Houwelingen HC van, Arends LR, Stijnen T. Advanced methods in meta-analysis: Multivariate approach and meta-regression. Statistics in Medicine 2002;21:589–624. Available at: <http://dx.doi.org/10.1002/sim.1040>.

5. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. Statistical methods in medical research 2001;10:277–303. Available at: <http://journals.sagepub.com/doi/10.1177/096228020101000404 http://www.ncbi.nlm.nih.gov/pubmed/11491414>.

6. Higgins JPT, Deek JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions 5.1.0 [updated march 2011]. Cochrane Collaboration, 2011:Available from www.cochrane–Available from handbook.org.

7. Review manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

8. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013. Available at: <http://www.R-project.org>.

9. Friedrich JO, Adhikari NK, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. Journal of Clinical Epidemiology 2011;64:556–64. Available at: <http://www.sciencedirect.com/science/article/pii/S0895435611000035>.

10. Zhao SZ, Chung F, Hanna DB, Raymundo AL, Cheung RY, Chen C. Dose-response relationship between opioid use and adverse effects after ambulatory surgery. Journal of Pain and Symptom Management 2004;28:35–46. Available at: <http://www.sciencedirect.com/science/article/pii/S0885392404001083>.

11. Achana FA, Cooper NJ, Dias S, Lu G, Rice SJC, Kendrick D, Sutton AJ. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. Statistics in Medicine 2013;32:752–71. Available at: <http://doi.wiley.com/10.1002/sim.5539>.

1. The 95% confidence intervals provided by Doleman et al.1 were used to estimate the standard error6. Figure 1 was generated with the Cochrane Collaboration software RevMan7. [↑](#footnote-ref-28)
2. Figure 2 and 3 were generated by a R8 statistical software package with data provided by the authors1, The data and code are available online. [↑](#footnote-ref-29)