

## Title

### Authors

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

Nathan L Pace, MD., MStat., Department of Anesthesiology, Univeristy of Utha, 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

### Acknowledgement

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

### Conflicts of Interest: None

### Keywords:

analgesic adjuvants, postoperative pain, evidence synthesis, meta-regression, Bayesian statistics

## 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 trial yield varying estimates of the clinical effectiveness of analgesic adjuvants. RCTs results, seemingly at random, leave clinicians baffled. Doleman et al in the study “Baseline morphine consumption...” propose an ingenious new solution to solve this riddle (Doleman 2017). They synthesized the evidence for morphine dose reduction with adjuvants by controlling for baseline risk (morphine consumption) across surgical procedures and patient populations. With their new approach, local audit data could be used to predict the expected average local 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.

Some variability in outcome estimates from RCTs investigating the same intervention in the similar population is to be expected. Each RCT obtains patients by a convenience sample; it is not a random sample of the entire population who might receive the adjuvant. 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 effect of interventions. Meta-analysis pools available RCTs to synthesize the evidence and to increase the precision of effect estimates. This may reduce the uncertainty in the face of seemingly contradictory results. However, if the heterogeneity of the RCTs is very large (too inconsistent and disparate), then these summary statistics may be inappropriate and are frequently described 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.

To explain why RCTs yield contradictory results, we can 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 control (Kehlet 2007): 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 to base clinical decision on for a particular patient population undergoing a particular procedure. For example, on thyroidectomy Doleman et al. found only one single RCT investigating the effect of intravenous acetaminophen. This poses a significant challenge to any evidence based approach to procedure specific postoperative pain management.

For these studies, the effect estimate is the mean difference in morphine dose between those receiving and those not receiving the adjuvant. The mean dose of those not receiving the adjuvant can be considered to be the baseline risk for pain. Doleman et al. used Bayesian hierarchical meta-regression modeling adapted to requirements of such evidence synthesis (Sutton citations). They demonstrated that surgical procedure is not the main predictor of between-study variability in effect estimates for analgesic adjuvants to control postoperative pain, but baseline risk is. Restated, 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.

Doleman et al. found 25 RCTs investigating intravenous acetaminophen for postoperative pain control. In Figure 1, a classical forest plot is shown, ordered by surgical interventions. The 95% confidence intervals provided by Doleman et al. 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 different effect sizes; the results are inconsistent, varying widely even within the same surgery.

In Figure 2, a schematic of the acetaminophen subplot of Figure 2 is rendered in the statistical software R (R Core Team 2013) with data provided by the authors. Each study is represented by an 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 some such studies are far apart on the y-axis. By contrast, studies line up neatly around the regression line, morphine reduction regressed on baseline risk. 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.

Doleman et al. proposes that this is true for all analgesic adjuvants across all populations, for all surgical interventions, as shown in the other subplots of Doleman's Figure 2. This contradicts the dominant paradigm of procedure specific postoperative pain control. What is more, Doleman et al also proposes that clinicians can use local audit data, collected on local surgical populations, to estimate the average reduction of morphine consumption for each adjuvant. A great feature to have for a new unifying theory on how to synthesize the evidence for postoperative pain control.

Doleman et al. fit both a classical model and a Bayesian model that differs from the classical approach (AKA frequentist statistics). The Bayesian approach incorporates prior information and combines it with the newly observed data, much like a physician would in clinical practice. Both frequentist and Bayesian statistical models are based on subjective assumptions; the model choice itself is often more important for correct inferences. In both types of modeling, the analyst should explore the sensitivity of results and inferences to assumptions and model choices. In an available supplement, Doleman et al provided software code, data and model details. They used very weak prior assumptions and in their sensitivity analysis found their results to be robust. They made their Bayesian approach transparent, sound and trustworthy.

Based on their previous work (Achana 2013), they 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 characteristics (Achana 2013). Their conjecture is supported by their meta analyses. This now suggests a set of testable hypotheses and predictions. This makes 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's predictions hold. Meta-analysts can replicate the approach by Doleman et al 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 go on to use meta-regression to discuss rather refined and specific questions, for example if acetaminophen is more effective if administered intravenously rather than orally, a question of great financial and clinical importance for which there are simply too few studies if we insisted to only combine studies comparing the effects in the same population/surgical intervention. They provide a league table of efficacy of analgesic adjuvants.

Doleman et al acknowledge the many limitations of their work. Their meta-regression, like any meta-analysis was an ex post 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 pars pro toto using acetaminophen as an example:

- The range of surgical procedures included in the acetaminophen trials was limited, (see Doleman's Table 1).
- The patient mix in the acetaminophen studies appears to be 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 article.
- Finally, the estimates in their regression equation apply to aggregate groups not individual patients.

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

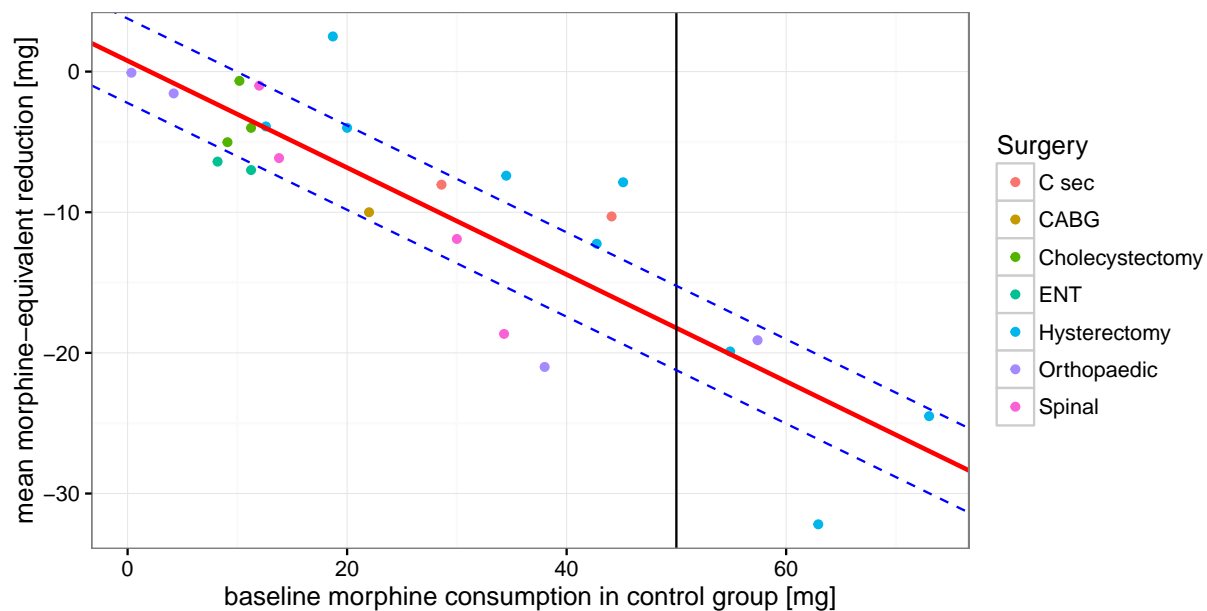
## Figures

### Figure 1

#### Figure 1 Caption

In Figure 1 estimates for the mean reduction of morphine equivalent afforded by acetaminophen are shown in a classical forest plot with studies ordered by surgical interventions. The mean reduction of morphine equivalents and their 95% confidence intervals 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 estimate variable and 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 colored the studies by surgical procedure, according to the adjacent colour 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 color dots should be clumped together, (and the green cholecystectomy studies somewhat are). But most studies of the same color are far apart, for example orthopaedic (blue) or hysterectomy (violet) study dots are spread out over the entire range of the plot. By contrast, all studies line up neatly around the red 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.

## References

Achana 2013

Achana FA, Cooper NJ, Dias S et al. Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Stat Med* 2013; 32: 752-71.

Higgins 2011

Higgins JPT, Green S (editors). Obtaining standard errors from confidence intervals and P values: absolute (difference) measures. In *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.handbook.cochrane.org](http://www.handbook.cochrane.org).

Doleman 2017

Doleman B, Sutton A, Sherwin M, Lund J, Williams J. 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

Kehlet 2007

Kehlet H, Wilkinson RC, Fischer HB, Camu F; Prospect Working Group.. PROSPECT: evidence-based, procedure-specific postoperative pain management. *Best Pract Res Clin Anaesthesiol*. 2007 Mar;21(1):149-59. Review. PubMed PMID: 17489225.

Sutton 2001

Sutton AJ1, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*. 2001 Aug;10(4):277-303.

R Core Team (2013).

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

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