

Identifying Medical Reversals: An Introduction to a New Area of Study

Hannah Bolder *

January, 2019

Abstract

The value of medical research lies in its ability to improve patient outcomes; however, there is often a disconnect between the state of medical research and clinical practice. There is a growing body of research about technology diffusion in healthcare and adoption of new treatments; however discontinuing medical practices that are found to be ineffective or harmful (“medical reversals”) remains understudied. This white paper explains the concept of medical reversal and explores the frequency of reversals. Next, two case studies are introduced to demonstrate instances of medical reversal. Section 2 documents terminology used in this new literature, and section 3 presents resources to help find medical reversals and identify potential new ones. Section 4 discusses how the use of surrogate outcomes and subgroup effects complicates the analysis of the medical literature necessary to identify medical reversals. Section 5 concludes.

The research in this paper was supported by the National Institute on Aging under Award Number R24AG058049 to the National Bureau of Economic Research. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Aging or NBER.

*hbolder@umich.edu, PhD Student Candidate at the University of Michigan

1 Introducing Medical Reversal

1.1 What is a medical reversal?

A “medical reversal” occurs when new evidence (usually from a high-quality RCT) reveals that an existing clinical practice is ineffective or even harmful. Reversed practices often initially rose to popularity due to a compelling theoretical mechanism underlying the treatment as well as some promising preliminary evidence. Initial evidence could be as sparse as a few case studies, or could include preliminary observational studies. Often early studies present a procedure’s effect on surrogate markers, such as the presence or lack of extra heartbeats (see case study of flecainide’s ability to suppress PVCs below) rather than on endpoints such as mortality.

De-adopting reversed practices is not merely a matter of cost-effectiveness. Whether a procedure is cost effective involves weighing the magnitude and frequency of potential health benefits in a population with treatment costs. Reversed practices are almost by definition not cost effective, because they either lack health benefits or are harmful. Similarly, reversed practices are not merely “low value” practices - they lack value or even generate harm in some cases.

There are many reasons why the de-adoption process could differ from diffusion of new treatments. In the case of medical reversals, when evidence is found against a practice, it is often the case that the better alternative is not a new innovation, but to return to a prior treatment or to instead do nothing. Therefore the financial incentives and psychological barriers to de-adoption are different.

1.2 How frequent are medical reversals?

According to Chandra (2015), the “grey zones” in medicine are numerous and large. Grey zones represent areas of medicine in which treatment benefits are uncertain or controversial. Treatment decisions in these grey zones depend on physician preferences, patient preferences, and local norms or customs. Although pharmaceuticals require large-scale RCTs and FDA approval, clinical practices that lie in the grey area can gain popularity even absent solid evidence. Out of a group of 3,000 practices, BMJ Clinical Evidence found that 50% were of unknown effectiveness. (BMJ Clinical Evidence n.d) Furthermore, out of the 363 articles published in the New England Journal of Medicine between 2001 and 2010 that tested clinical practices, 146 (40.2%) constituted medical reversals, according to Prasad et al (Prasad et al 2013, 790). That is, the affected practices were deemed ineffective or even harmful when compared with either the prior standard of care or with doing nothing. Due to various methodological considerations, it is difficult to precisely estimate the frequency of medical reversals. However a back-of-the-envelope calculation using the aforementioned statistics would indicate that $(.5)(.4)(100)=20\%$ of current clinical practice will ultimately be categorized as medical reversals. The extent that medical practice relies on evidence varies by specialty, however in many specialties the grey area is quite large. According to Prasad et al (2012), “it is possible that some entire medical subspecialties are based on little evidence.” Other sources estimate that perhaps less than half of clinical practices are based on evidence, in which case medical reversals would be even more prevalent.

Another estimate of the frequency of medical reversal comes from a research letter published in the Archives of Internal Medicine by Prasad, Gall and Cifu (2011). They review all articles published in the New England Journal of Medicine in 2009

and determine that out of 124 articles that evaluated clinical practices, 89 (72%) studied new practices and 35 (28%) evaluated pre-existing practices. 12 (10%) of the 124 studies found that a new practice was no more effective than an existing one, and 16 (13%) constituted medical reversals.

One might criticize either of these estimates because the New England Journal of Medicine is not representative of the average medical journal. Thus it may be more likely to publish avant-garde and provocative findings if the methodology is sound. In addition, as Prasad and colleagues point out, it is possible that some of the findings in the “reversal” studies will not hold up to further scrutiny or be replicable. However it is certainly plausible that the New England Journal of Medicine would suffer less from these troubles given its high standards for publication.

Given the large number of practices and people affected by medical reversal and the financial implications of wasting valuable resources on ineffective procedures, it is crucial to understand the incentives that hinder the de-adoption or de-implementation process. First, in order to gain context, two case studies of reversed practices are presented: a medical device that was found to be no more effective than the pre-existing practices, and a treatment that was found to increase mortality.

1.3 Case Study - Bispectral Index Monitor (BIS)

The bispectral index monitor was developed to prevent a phenomenon called “intra-operative recall” or “anesthesia awareness.” In this rare but devastating condition, patients experience varying degrees of awareness during surgery. Patients can hear conversations in the operating room, feel surgeons’ hands inside of them, and even sometimes feel pain during the procedure. Unfortunately, they have no means of notifying physicians that they are in fact aware. According to Sullivan (2016) anes-

thetia awareness only occurs in about .1-.2% of patients, but due to the high volume of surgeries in the US, this amounts to 20,000-40,000 instances per year. Many of these patients experience PTSD after anesthesia awareness, and there has even been a report of a suicide (Leslie et al 2010 and The Associated Press 2007).

The bispectral index monitor was developed by Nassib Chamoun of Aspect Medical Systems to monitor EEG activity and produce a measure that would indicate patients' state of awareness during operations. The index ranges from 0 to 100, where higher values indicate more awareness. The target range for a surgical operation is generally 40-60.

Two early studies produced promising results concerning the efficacy of BIS monitoring. First, Puri and Murthy (2003) studied a group of patients undergoing coronary artery grafting and valve replacement under cardiopulmonary bypass. In this RCT, the group that had surgery with BIS monitoring had fewer instances of hypertension and tachycardia, however the authors lacked sufficient power to detect differences in anesthesia awareness. (There were only 30 participants enrolled in the study.) Myles et al (2004) enrolled 2463 participants in the double-blind multi-center RCT commonly referred to as the B-Aware trial (funded by Aspect Medical Systems). They found that the use of BIS by anesthesiologists decreased the risk of intraoperative awareness by 82%. Unfortunately the study does not describe the control arm in much detail, merely referring to it as "routine care" without BIS monitoring.

Some physicians began to question the efficacy of BIS as early as 2002. In 2002, Schneider et al published a study examining the ability of BIS to predict intraoperative awareness. They found that after intubation, 8 out of 20 patients exhibited signs of awareness (but no recall) despite similar BIS values of 50-60 among all 20 participants. They conclude that BIS is insufficient to distinguish patients that

show signs of intraoperative awareness, and “its value as a monitor for awareness and a measurement of the hypnotic component of anesthesia must be questioned” (Schneider et al 2002, 7).

In addition, Rampersad and Mulroy published a case report in 2005 concerning a patient who experienced intraoperative awareness during an open gastric bypass and cholecystectomy. A BIS was used during the surgery: the mean value was 44 and the highest value was 51 - well within the recommended range. When the patient awoke, he described “. . . vivid, painful recall of his surgery, with ‘unimaginable’ pain and the sensation that people were ‘tearing at me.’ He wished he were dead and tried to communicate his distress.” (Rampersad and Mulroy, 2005) They conclude, “In this case, not only did the BIS monitor not reliably predict the absence of awareness but its use may also have contributed to the occurrence of awareness. In the absence of this monitor, an end-tidal concentration of .45%-.8% sevoflurane, even with a working epidural, might have been regarded as an inadequate anesthetic and the level of sevoflurane would have been increased or benzodiazepines added to prevent recall (with additional vasopressor support).” (Rampersad and Mulroy 2005, 1364)

Finally in 2008 and 2011, Avidan and colleagues published two RCTs that called into question the value of BIS for preventing intraoperative awareness and recall. The study population in both trials consisted of individuals with high-risk of intraoperative awareness undergoing general anesthesia using isoflurane, sevoflurane or desflurane. In both studies the control arm received ETAC-guided anesthesia. In the 2008 study, referred to as the B-Unaware trial, the anesthesiologists of patients in both the control and treatment arms could view ETAC concentrations, but only anesthesiologists of patients in the treatment arm could view the patients’ BIS values. This was also true in the 2011 study, termed the BAG-RECALL study. In both studies, patients were interviewed to assess intraoperative awareness and recall, and

expert reviewers determined presence or lack of awareness for study purposes. Neither interviewers nor expert reviewers knew patients' treatment assignments. The B-Unaware trial included only one center with 2000 participants, whereas the BAG-RECALL study was a multicenter trial with 6041 participants. Neither study found evidence that the BIS was superior to ETAC concentration monitoring as a method to prevent intraoperative awareness. Notably, the 2011 study also found some cases of intraoperative awareness in patients with BIS values below 60.

After the early promising results of BIS monitoring, many hospitals purchased these devices. According to an article in the Atlantic Monthly, by July 2007 about half of US operating rooms contained a BIS monitor (Lang, 2013). Despite the results from the B-Unaware trial and the BAG-RECALL trial, Gelfand et al (2017) find that about 53% of patients were subject to BIS monitoring during their surgeries at Brigham and Women's Hospital between January 2013 and October 2014. Clearly, despite the dubious efficacy of BIS monitoring in preventing intraoperative awareness, their use remains common, at least in one major US hospital. ¹

1.4 Case Study - Flecainide to prevent extra heartbeats after heart attacks

According to Prasad and Cifu (2015), flecainide was often prescribed after myocardial infarction (heart attacks) during the late 1980s and early 1990s in order to prevent premature ventricular contractions (PVCs). PVCs are essentially extra heartbeats

¹Although there is some evidence that BIS monitoring may decrease time to eye-opening, time to response to command, time to extubation, and time spent in the PACU (see Punjasawadwong et al 2014), most of these effects amount to a 2-3 minute reduction (with the exception of PACU time, which was reduced by 6-7 minutes). Although these effects are statistically significant, one could question whether they are likely to be clinically significant. In neither the B-Unaware nor the BAG-RECALL trial was the ETAC concentration control arm associated with increased mortality or increased use of anesthesia.

that originate in one of the ventricles. When the ventricle contracts prematurely, the heart’s electrical system “resets.” This causes the heartbeat to pause for a few seconds and patients report that it seems like their heart skips a beat. PVCs sometimes occur after heart attacks, which is associated with an increased risk of sudden death. During the 1980s, physicians believed that flecainide was an effective treatment for post-heart attack patients at risk of PVCs because flecainide suppresses extra heartbeats. As Prasad and Cifu (2015) explain, the logic was “ironclad.” In 1991 however, the CAST trial was published in the *New England Journal of Medicine*. It demonstrated that although flecainide suppresses extra heartbeats, it also increases mortality rates. This effect persists whether one analyzes the effect of flecainide on death caused by arrhythmia, death caused by any cardiac event, or death for all causes (Echt et al, 1991). According to Andrikopoulos et al (2015), due to the results of the CAST trial, flecainide is not advisable for patients with structural heart disease or coronary artery disease. It is interesting to note that according to Prasad and Cifu (2015) many cardiologists were so confident of that flecainide saved lives that they refused to allow their patients to participate in the CAST trial.

2 Terminology

Medical reversal and de-adoption of reversed practices is a relatively new area of study. In Niven et al’s (2015) review of articles that discuss de-adoption of low value practices, the majority were written after 2010. Since this literature is new and the field has not converged on a uniform set of terms to describe relevant concepts, there are no MeSH terms for medical reversal or de-adoption, which makes it more difficult to identify relevant articles in National Library of Medicine databases such as PubMed (Niven et al, 2015). According to Sutton et al (2018) finding relevant

articles is “labor intensive.”

In their review of articles, Niven et al (2015) found 43 unique terms to describe the phenomenon of de-adoption. The most common term was disinvest, which was present in about 40% of the articles. De-adopt was only used in 3% of the articles; however they advocate for the use of this term (or de-implement - used in 4% of articles) because they more accurately describe the process of discontinuing the use of a procedure, rather than disinvest, which should be used to refer to the specific aspect of discontinuation that involves funding. Other notable terms include exnovation and evidence reversal. Bekelis et al (2017) define exnovation as “scaling back but not necessarily abandoning a practice,” and Sutton et al (2018) advocate for the use of the broader term “evidence reversal” to describe the phenomenon in which evidence is later contradicted in fields outside of clinical medicine. They argue that the study of “evidence stability” should not be confined to clinical medicine. They also propose that the term evidence reversal should be used in cases when the evidence base has shifted but clinical practice has not responded. That is, they believe that a medical reversal implies that the practices of clinical medicine have responded to the new evidence, whereas evidence reversal carries no such implication. ²

The following is a list of terms that will be useful to search for relevant research and describe concepts related to medical reversal. This list is based in part on the findings of Niven et al (2015) and Sutton et al (2018): medical reversal, evidence reversal, contradicted practice, disinvest, de-adopt or dis-adopt, de-implement, exnovation, abandon, decommission, de-list, defund, and undiffusion.

²Given that the literature continues to use medical reversal to describe both situations where de-adoption has and hasn't occurred, this paper will continue to use the term medical reversal in both cases.

3 Where to find examples of medical reversal

Of course new medical reversals may be identified by reading medical journals and doing a comprehensive literature review on a given medical practice. However this can be time consuming to do for every possible candidate treatment. It is worth discussing methods that may allow for more efficient identification of potential new medical reversals by narrowing the scope of the search. Procedures, devices or drugs that may be candidates for medical reversal can be found in guidelines (discussed in Section 3.1-3.3).

Although there are lists of low-value care, there is no database or organization that produces lists of practices that have already been classified as medical reversals. Studies about medical reversal or de-adoption may be found by searching databases of academic papers such as Google Scholar or PubMed or databases of clinical trials such as Cochrane CENTRAL using the various terms mentioned in Section 2 above. Other sources that may be helpful will be discussed in Section 3.4.

3.1 Guidelines

Cochrane systematic reviews, US Preventive Service Task Force recommendations and other guidelines may be useful to help identify potential reversals. A wide variety of other organizations produce guidelines (for example specialty societies, insurers, hospitals, etc) and they vary in their development methodology, motivations, and reliance on evidence. Crucial for the purpose of identifying new medical reversals, many guidelines are not updated very frequently, so there could be long lags between published study findings and a change in guideline recommendations. Given the large variation in guideline quality, researchers must interpret their recommendations with caution; however evidence-based guidelines from reputable organizations should still

be useful to researchers interested in de-adoption. See Catillon (2017) for a more complete discussion of systematic reviews, and guideline quality and development, and as well as repositories.

3.2 The future of the National Guideline Clearinghouse

Unfortunately funding for the AHRQ's National Guideline Clearinghouse ended on July 16, 2018 and their repository is no longer available. As of November 2018, the ECRI Institute will now host the National Guideline Clearinghouse guidelines through the ECRI Guidelines Trust. The ECRI Institute developed and maintained the National Guideline Clearinghouse website for 20 years, according to a recent news release from ECRI (ECRI Institute, 2018). The ECRI Guidelines Trust is available at: <https://guidelines.ecri.org/>

Other alternative repositories include: Guideline Central and Guidelines International Network (mentioned as well by Catillon (2017)).

3.3 Choosing Wisely

Choosing Wisely guidelines may be especially useful to identify practices that should be de-adopted; however Choosing Wisely focuses on low-value practices, therefore not all practices mentioned in Choosing Wisely publications constitute medical reversals. Many of these procedures are valuable in some contexts but are unnecessary and overused in others; for example X-rays, CT scans and MRIs are low-value in the context of lower back pain but they are all very useful in other contexts. Nevertheless, some medical reversals begin when evidence emerges that they may be less effective than previously thought for some subgroups. Once an RCT investigates the practice for larger populations the practice may be reversed entirely. This is not likely to be

the case for X-rays, CT scans or MRIs of course, but it may be the case for some treatment protocols, cancer screening tests or drugs.

3.4 Books and academic reviews

For researchers interested in lists of practices, devices or procedures that have already been classified by others as medical reversals (rather than identifying new reversals) there are several articles that may be useful. Many of the reversals listed in these articles would serve as a great starting point for de-adoption research. The supplemental appendix from Prasad et al (2013) includes a list of 146 medical reversals from published articles in the New England Journal of Medicine from 2001-2010. Prasad and Cifu's book *Ending Medical Reversal: Improving Outcomes, Saving Lives* also includes a list of medical reversals and describes the evolution of medical practices, providing useful background and context. For a list of drugs that were approved based on initially promising evidence on surrogate outcomes and later demonstrated to be harmful, see Svensson, Menkes, and Lexchin (2013). In addition, the supplemental materials from Niven et al (2015) contain a list of articles that identify low-value practices and/or analyze the de-adoption process.

4 Interpreting randomized control trial evidence from the medical literature

As in any scientific field, knowledge in medicine does not progress linearly. Medical reversals are a more extreme example of this non-linearity. However the term "medical reversal" belies the complicated analysis and interpretation of the medical literature that is needed to justify its use. Studies differ in their methods. When

these studies have mixed or contradictory findings, it necessary to weigh their relative strengths and weaknesses to form a conclusion about the state of the evidence for or against the medical practice. Section 4.1 briefly discusses standard internal and external validity concerns of RCTs. Sections 4.2 and 4.3 discuss in more detail two issues that arise frequently in the medical literature: surrogate endpoints and subgroup effects.

4.1 Standard internal and external validity concerns in RCTs

RCTs in medicine have many strengths, but like those in other fields they may be subject to internal validity problems such as attrition, contamination of the control group due to cross-over, or design problems such as an improperly defined control group. For example, in an RCT that attempts to estimate the effect of a surgical procedure on pain alleviation, the proper control should usually be a sham surgery – not an oral medication. If the control is an oral pain medication, patients and physicians will not be blind to treatment assignment which can introduce biases and may confound the true treatment effect with a placebo effect. In addition, RCTs in medicine may also suffer from external validity constraints because the population that is recruited to participate in the RCT may differ systematically from the population outside the trial that would use the treatment. These population differences could be unintentional, or in some cases an intentional part of the study design in order to save time and money and increase convenience.

In addition to these standard concerns, there are some issues that warrant a more in-depth discussion because of the context in which they occur in the medical field.

4.2 Surrogate outcomes / end points

In order to determine whether a treatment is beneficial to patients, it is important to collect data on patient-centric outcomes, such as mortality, morbidity, quality of life, level of pain, etc. However, it is often difficult to collect such data in an RCT. If RCT patients do not have a terminal illness, investigators might need to wait years or even decades before enough mortality data on these patients becomes available to be able to draw conclusions. Even for other patient health outcomes such as number of strokes or heart attacks, it may be faster and less expensive to use alternative “surrogate” measures rather than direct patient outcomes. Surrogate outcomes or end points are measures that are usually derived from laboratory values, radiological tests, or some other biological measure of disease progression (for example, tumor size). Although they are thought to be correlated with patient welfare, they are not measures that the patient can directly perceive.

Problems arise when surrogate outcomes are used in a context in which their role in the pathophysiological process of a disease is unknown or not well understood. A surrogate may be a symptom of the underlying condition but not the true cause of disease progression. In addition, problems also arise if a treatment affects key health outcomes through channels not mediated by the surrogate outcome. A treatment may have positive effects on the surrogate outcome but come with side effects or unintended consequences that overshadow any potential benefits. Recall from section 1.4 that initial promising evidence for the use of flecainide came from its ability to reduce PVCs, or extra heartbeats. PVCs acted as a surrogate outcome for the true outcome of interest – mortality. Flecainide had a positive effect on the surrogate outcome but increased mortality. It is possible that while flecainide reduced mortality due to PVCs, it also caused other biological changes that increased mortality so much

that the net effect of flecainide on mortality was detrimental. Alternatively, it is possible that PVCs are caused by a third factor that flecainide does not address, and that preventing PVCs in the presence of this third factor is worse than doing nothing. Clearly, choosing an appropriate surrogate is difficult and requires a sophisticated understanding of the mechanisms through which the treatment works as well as the pathophysiological process. For a full discussion of various theoretical conditions that would cause a surrogate end point measure to fail, see Fleming and DeMets (1996). Fleming and DeMets (1996) also discuss examples of failed surrogate end points in cardiology and oncology. For a more recent discussion, see Fleming and Powers (2012). Ciani et al (2017) discusses how to properly select, validate and use surrogates while warning that their appropriateness can only be determined relative to a specific context. That is, a given surrogate may be appropriate to analyze the effectiveness of intervention 1 in disease context A but inappropriate to analyze the effectiveness of intervention 2 in disease context A, or to analyze the effect of intervention 1 in disease context B. Unfortunately, it is difficult to rigorously validate a surrogate endpoint, and many studies have failed to do so.

4.3 Subgroup effects / heterogeneous treatment effects

According to Assmann et al (2000), “Of all the various multiplicity problems in clinical trials, subgroup analysis remains the most overused and overinterpreted.” (Assmann et al 2000, page 1067). Assmann et al (2000) made this claim in the discussion of their findings after analyzing the methods of clinical trials published in four major medical journals in 1997. More recently, others have found that the methods used in subgroup analysis in the medical literature are often problematic. According to Sun et al (2010), some studies define subgroups based on characteristics

that could have been affected by the treatment, rather than based on baseline characteristics or other unaffected variables. Sun et al (2012) and Wallach et al (2017) find that in most studies in their sample where subgroup effects are claimed, stratified randomization was not performed nor was there any evidence of prespecification of subgroup tests (as opposed to running many post-hoc subgroup tests). In addition, interaction tests are not performed for most of the subgroup effects, and according to Wallach et al (2017) most subgroup claims made in the paper abstracts failed to find support in the paper's subsequent data analysis. Very few studies mentioned any adjustment due to multiple hypothesis testing. Wang et al (2007) finds similar problems after analyzing papers published in the New England Journal of Medicine between July 2005 and June 2006.

According to Sun et al (2014), “the challenge for readers of the medical literature is to distinguish credible from less than credible reports of subgroup effects. Clinicians cannot rely on study authors to do this for them.” (Sun et al 2014, 405). In all fields, researchers must exercise judgment to determine the current state of the literature. Inevitably, in the medical literature this involves weighing the relative contribution of several papers and determining which (if any) subgroup effects to believe. Sun et al (2010) and Wang et al (2007) suggest standards to evaluate the subgroup claims made in clinical trials. Aside from the metrics mentioned above, it is important to assess whether the claimed subgroup effect on the primary outcome persists when examining secondary study outcomes. Crucially, there should also be theoretical evidence to support the existence of subgroup effects. If no compelling biological mechanism can explain the subgroup effect, one may wonder if it was spurious. Since no formula can establish whether a subgroup effect seems credible and studies use inconsistent methods to investigate subgroup effects, the reader may need to conduct additional research in order to make an informed decision.

The validity of studies' surrogate outcome measures, the validity of claimed subgroup effects and the potential for unknown subgroup effects should all impact the decision of whether the term “medical reversal” applies.

5 Conclusion

This white paper reviews the definition of a medical reversal, a term first coined by Prasad and Cifu in 2011. It is important to note that medical reversal differs from low-value care, and unfortunately may be quite common in the evolution of medical knowledge. Inherent in the task of identifying medical reversals is a thorough examination of the medical literature, which can be complicated by internal and external validity concerns, especially those pertaining to subgroup effects and surrogate endpoints. Adopting practices into clinical medicine before a rigorous body of evidence has time to develop and support their efficacy may be ex-ante optimal in some cases, as long as patients and other stakeholders are informed of the risks. However once a clinical practice is contradicted by more rigorous evaluation, there is little evidence to guide effective methods of de-adoption. Medical reversal and de-adoption are new areas of study within the field of translational research; therefore further research is needed to promote efficient de-adoption of ineffective or harmful practices.

6 References

- Andrikopoulos George, Sokratis Pastromas and Stylianos Tzeis. 2015. "Flecainide: Current status and perspectives in arrhythmia management." *World Journal of Cardiology* 7(2) February 26:76-85.
- Assmann Susan, Stuart Pocock, Laura Enos, and Linda Kasten. 2000. "Subgroup analysis and other (mis)uses of baseline data in clinical trials." *The Lancet* 355:1064-1069.
- Avidan Michael, Lini Zhang, Beth Burnside, Kevin Finkel, Adam Searlemen, Jacqueline Selvidge, Leif Saager, Michelle Turner, Srikar Rao, Michael Bottros, et al. 2008. "Anesthesia Awareness and the Bispectral Index." *New England Journal of Medicine* 358:1097-108.
- Avidan Michael, Eric Jacobsohn, David Glick, Beth Burnside, Lini Zhang, Alex Villafrañca, Leah Karl, Saima Kamal, Brian Torres, Michael O'Connor, et al. 2011. "Prevention of Intraoperative Awareness in a High-Risk Surgical Population." *New England Journal of Medicine* 365:591-600.
- Bekelis Kimon, Jonathan Skinner, Daniel Gottlieb and Philip Goodney. 2017. "De-adoption and exnovation in the use of carotid revascularization: retrospective cohort study." *British Medical Journal* 359(j4695) <http://dx.doi.org/10.1136/bmj.j4695>
- BMJ Clinical Evidence. n.d. "What Conclusions has Clinical Evidence Drawn about What Works, What Doesn't Based on Randomized Controlled Trial Evidence?" Accessed Jan 9, 2018. <http://clinicalevidence.bmj.com/x/set/static/cms/efficacy-categorisations.html>
- Catillon, Maryaline. "Medical Knowledge Synthesis: A Brief Overview." November 8, 2017. White Paper for NBER-IFS International Network of the Value of Medical Research. http://www.nber.org/aging/valmed/research_ifs.html
- Chandra, Amitabh, Dhruv Khullar and Thomas Lee. 2015 "Addressing the Challenge of Gray-Zone Medicine." *New England Journal of Medicine* 372(3): 203-205.
- Ciani Oriana, Marc Buyse, Michael Drummond, Guido Rasi, Everardo Saad, and Rod Taylor. 2017. "Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward." *Value in Health* 20: 487-495.
- Echt Debra, Philip Liebson, Brent Mitchell, Robert Peters, Dulce Obias-Manno, Allan Barker, Daniel Arensberg, Andrea Baker, Lawrence Friedman, Leon Greene, et al. 1991. "Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo: the Cardiac Arrhythmia Suppression Trial." *New England Journal of*

Medicine 324(12):781-8.

ECRI Institute. Nov 19, 2018. "ECRI Institute Opens Access to Clinical Practice Guidelines." Accessed January 3, 2019. <https://www.ecri.org/press/ecri-institute-opens-access-to-clinical-practice-guidelines/>

Fleming Thomas and David DeMets. 1996. "Surrogate End Points in Clinical Trials: Are We Being Misled?" *Annals of Internal Medicine* 125: 605-613.

Fleming Thomas and John Powers. 2012. "Biomarkers and Surrogate Endpoints in Clinical Trials." *Statistics in Medicine* 31(25):2973-2984.

Gelfand Melanie, Rodney Gabriel, Robert Gimlich and Sascha Beutler. 2017. "Practice Patterns in the intraoperative use of bispectral index monitoring." *Journal of Clinical Monitoring and Computing* 31:281-289.

Lang, Joshua. "Awakening." *The Atlantic*. Jan/Feb 2013. <https://www.theatlantic.com/magazine/archive/2013/01/awakening/309188/>

Leslie Kate, Matthew Chan, Paul Myles, Andrew Forbes, and Timothy McCulloch. 2010. "Posttraumatic Stress Disorder in Awake Patients from the B-Aware Trial." *Anesthesia & Analgesia* 10(3):823-828.

Myles Paul, Kate Leslie, John McNeil, Andrew Forbes, Matthew Chan. 2004. "Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial." *The Lancet*. 363: 1757-63.

Niven Daniel, Kelly Mrklas, Jessalyn Holodinsky, Sharon Straus, Brenda Hemmelgarn, Lianne Jeffs and Henry Stelfox. 2015. "Towards understanding the de-adoption of low-value clinical practices: a scoping review." *BioMed Central Medicine* 13:255.

Prasad Vinayak, Victor Gall, and Adam Cifu. 2011. "The Frequency of Medical Reversal." *Archives of Internal Medicine* 171(18):1675-1676.

Prasad Vinayak, Adam Cifu, and John Ioannidis. 2012. "Reversals of Established Medical Practices: Evidence to Abandon Ship." *Journal of the American Medical Association*. 307(1):37-38.

Prasad Vinay, Andrae Vandross, Caitlin Toomey, Michael Cheung, Jason Rho, Steven Quinn, Satish Chacko, Durga Borkar, Victor Gall, Senthil Selvaraj, et al. 2013. "A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices." *Mayo Clinic Proceedings* 88: 790-798

Prasad Vinayak and Adam Cifu. 2015. *Ending Medical Reversal: Improving Outcomes, Saving Lives*. Baltimore: Johns Hopkins University Press.

- Punjasawadwong Yodying, Aram Phongchiewboon and Nutchanart Bunchungmongkol. 2014. "Bispectral index for improving anaesthetic delivery and postoperative recovery." *Cochrane Database of Systematic Reviews*. Issue 6. Art. No.: CD003843.
- Puri Goverdhan and Sudha Murthy. 2003. "Bispectral index monitoring in patients undergoing cardiac surgery under cardiopulmonary bypass." *European Journal of Anaesthesiology*. 20:451-456.
- Rampersad Sally and Michael Mulroy. 2005. "A Case of Awareness Despite an "Adequate Depth of Anesthesia" as Indicated by a Bispectral Index Monitor." *Anesthesia and Analgesia* 100:1363-4.
- Schneider Gerhard, Klaus Wagner, Wolfram Reeker, Frank Hänel, Christian Werner, and Eberhard Kochs. 2002. "Bispectral Index (BIS) May Not Predict Awareness Reaction to Intubation in Surgical Patients." *Journal of Neurosurgical Anesthesiology*. 14(1):7-11
- Sullivan, Caitlin. 2016. "Awareness with Recall: A Systematic Review." *American Association of Nurse Anesthetists Journal*. 84(4):283-288.
- Sun Xin, Matthias Briel, Stephen Walter and Gordon Guyatt. 2010. "Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses." *British Medical Journal* 340(7751):850-854.
- Sun Xin, Matthias Briel, Jason Busse, John You, Elie Akl, Filip Mejza, Malgorzata Bala, Dirk Bassler, Dominik Mertz, Natalia Diaz-Granados, et al. 2012. "Credibility of claims of subgroup effects in randomised controlled trials: systematic review." *British Medical Journal* 344(e1553): doi: 10.1136/bmj.e1553 (Published March 15 2012).
- Sun Xin, John Ioannidis, Thomas Agoritsas, Ana Alba and Gordon Guyatt. 2014. "How to Use a Subgroup Analysis: Users' Guides to the Medical Literature." *Journal of the American Medical Association* 311(4):405-411. doi:10.1001/jama.2013.285063
- Sutton Desirée, Riaz Qureshi and Janet Martin. 2018. "Evidence reversal - when new evidence contradicts current claims: a systematic overview review of definitions and terms." *Journal of Clinical Epidemiology* 94: 76-84.
- Svensson Staffan, David Menkes, Joel Lexchin. 2013. "Surrogate outcomes in clinical trials: a cautionary tale." *Journal of the American Medical Association Internal Medicine* 173(8):611-612.
- The Associated Press. 2007. "Family sues after man gets wide-awake surgery." *NBC*

News, April 10, 2007.

http://www.nbcnews.com/id/18040894/ns/health-health_care/t/family-sues-after-man-gets-wide-awake-surgery/.XE-6Wc17nIU

Wallach Joshua, Patrick Sullivan, John Trepanowski, Kristin Sainani, Ewout Steyerberg, and John Ioannidis. 2017. "Evaluation of Evidence of Statistical Support and Corroboration of Subgroup Claims in Randomized Clinical Trials." *Journal of the American Medical Association Internal Medicine* 177(4):554-560.
doi:10.1001/jamainternmed.2016.9125

Wang Rui, Stephen Lagakos, James Ware, David Hunter, and Jeffrey Drazen. 2007. "Statistics in Medicine - Reporting of Subgroup Analyses in Clinical Trials." *New England Journal of Medicine* 357(21):2189-2194.