A Decade of Reversal: An Analysis of 146 Contradicted Medical Practices

Vinay Prasad, MD; Andrae Vandross, MD; Caitlin Toomey, MD; Michael Cheung, MD; Jason Rho, MD; Steven Quinn, MD; Satish Jacob Chacko, MD; Durga Borkar, MD; Victor Gall, MD; Senthil Selvaraj, MD; Nancy Ho, MD; and Adam Cifu, MD

Abstract

Objective: To identify medical practices that offer no net benefits.

Methods: We reviewed all original articles published in 10 years (2001-2010) in one high-impact journal. Articles were classified on the basis of whether they addressed a medical practice, whether they tested a new or existing therapy, and whether results were positive or negative. Articles were then classified as 1 of 4 types: replacement, when a new practice surpasses standard of care; back to the drawing board, when a new practice is no better than current practice; reaffirmation, when an existing practice is found to be better than a lesser standard; and reversal, when an existing practice is found to be no better than a lesser therapy. This study was conducted from August 1, 2011, through October 31, 2012.

Results: We reviewed 2044 original articles, 1344 of which concerned a medical practice. Of these, 981 articles (73.0%) examined a new medical practice, whereas 363 (27.0%) tested an established practice. A total of 947 studies (70.5%) had positive findings, whereas 397 (29.5%) reached a negative conclusion. A total of 756 articles addressing a medical practice constituted replacement, 165 were back to the drawing board, 146 were medical reversals, 138 were reaffirmations, and 139 were inconclusive. Of the 363 articles testing standard of care, 146 (40.2%) reversed that practice, whereas 138 (38.0%) reaffirmed it.

Conclusion: The reversal of established medical practice is common and occurs across all classes of medical practice. This investigation sheds light on low-value practices and patterns of medical research.

Published by Elsevier Inc on behalf of Mayo Foundation for Medical Education and Research


We expect that new medical practices gain popularity over older standards of care on the basis of robust evidence indicating clinical superiority or noninferiority with alternative benefits (eg, easier administration and fewer adverse effects). The history of medicine, however, reveals numerous exceptions to this rule. Stenting for stable coronary artery disease was a multibillion dollar a year industry when it was found to be no better than medical management for most patients with stable coronary artery disease.1 Hormone therapy for postmenopausal women intended to improve cardiovascular outcomes was found to be worse than no intervention.2 and the routine use of the pulmonary artery catheter in patients in shock was found to be inferior to less invasive management strategies.3 Previously, we have called this phenomenon (when a medical practice is found to be inferior to some lesser or prior standard of care) a medical reversal.4-6 Medical reversals occur when new studies—better powered, controlled, or designed than their predecessors—contradict current practice.4 In a prior investigation of 1 year of publications in a high-impact journal, we found that of 35 studies testing standard of care, 16 (46%) constituted medical reversals.4 Another review of 45 highly cited studies that claimed some therapeutic benefit found that 7 (16%) were contradicted by subsequent research.7

Identifying medical practices that do not work is necessary. The continued use of such practices wastes resources, jeopardizes patient health, and undermines trust in medicine. Interest in this topic has grown in recent years. The American Board of Internal Medicine launched the Choosing Wisely campaign,8 a call on professional societies to identify the top 5 diagnostic or therapeutic practices in their field that should not be offered.8 In England, the National Institute for Health and Clinical Excellence has tried to “disinvest” from low-value practices, identifying more than 800 such practices in the past decade.10 Other researchers have found that scanning a range of existing health care databases can easily
generate more than 150 low-value practices. Medical journals have specifically focused on instances in which more health care is not necessarily better. The Archives of Internal Medicine created a new feature series in 2010 entitled “Less is More.”

Given ongoing and vigorous efforts to identify medical practices that offer little benefit and minimal empirical studies documenting the rate at which current practices are contradicted, we performed a review of 10 years of original publications in one high-impact journal.

METHODS
We used methods similar to our prior survey of 1 year of publications in a high-impact journal. We reviewed all articles under the heading “Original Articles” in the New England Journal of Medicine from 2001 to 2010. These years were the last complete 10 years when we began our investigation. Our choice of journal was made on the basis of the 5-year Hirsch index for medical journals. Two reviewers (C.T., A.V., M.C., J.R., S.Q., S.J.C., D.B., V.G., or S.S.) and V.P. independently extracted information for each calendar year. This study was conducted from August 1, 2011, through October 31, 2012.

On the basis of published abstracts, articles were classified as to whether they addressed a clinical practice. Articles addressing a medical practice were defined as any investigation that assesses a screening, stratifying, or diagnostic test, a medication, a procedure or surgery, or any change in health care provision systems. Many research articles concern the novel molecular basis of disease or novel insights in pathophysiology. These articles were excluded. When practice information could not be ascertained by abstract alone, full articles were read.

Two reviewers (C.T., A.V., M.C., J.R., S.Q., S.J.C., D.B., V.G., or S.S.) and V.P. read articles addressing a medical practice in full. On the basis of the abstract, introduction, and discussion, articles were classified as to whether the practice in question was new or existing. Methods were classified as one of the following: randomized controlled trial, prospective controlled (but nonrandomized) intervention study, observational study (prospective or retrospective), case-control study, or other methods. End points for articles were classified into those that reached positive conclusions and those that found negative or no difference in end points. Lastly, articles were given 1 of 4 designations. Replacement was defined as a new practice surpassing an older standard of care. Back to the drawing board was defined as a new practice failing to surpass an older standard. Reversal was designated when a current medical practice was found to be inferior to a lesser or prior standard. Reaffirmation was defined as an existing medical practice being found to be superior to a lesser or prior standard. Finally, articles in which no firm conclusion could be reached were termed inconclusive. The designation of an article was also performed in duplicate. When there were differences in opinion between the 2 reviewers, adjudication first involved discussion between the 2 readers to see whether agreement could be reached. If disagreement persisted, a third reviewer (A.C.) adjudicated the discrepancy. Less than 3% of articles required discussion, and less than 1% required adjudication. A table detailing each medical reversal was constructed (Supplemental Appendix; available online at http://www.mayoclinicproceedings.org), and the third reviewer (A.C.) reviewed all reversals.

Data are summarized using counts and percentages. A linear regression was performed to determine the relationship between percentage of reversals and time, and the Pearson $\chi^2$ test was used when appropriate. Analyses were conducted using Stata statistical software, version 12 (StataCorp LP).

RESULTS
From 2001 through 2010, 2044 original articles appeared in one high-impact journal. Most articles (1344 [65.8%]) addressed a medical practice. A total of 981 studies (73.0%) examined a new medical practice, whereas 363 (27.0%) addressed an existing practice. During these 10 years, there were 911 (67.7%) randomized controlled trials, 220 (16.4%) prospective controlled but nonrandomized studies, 117 (8.7%) observational studies, 43 (3.2%) case-control studies, and 53 (3.9%) studies using other methods.

Concerning the study results, 947 (70.5%) reached positive conclusions, whereas 397 (29.5%) reached negative conclusions or found no difference between comparators. As such, 756 articles (56.3%) found a new practice
surpassing current standard of care (replacement), 165 (12.3%) found a new practice failing to improve on the current practice (back to the drawing board), 146 (10.9%) were reversals, and 138 (10.3%) upheld standard of care over a lesser or prior standard (reaffirmation). A total of 139 (10.3%) were deemed inconclusive. Figure 1 shows a breakdown of articles. The single most common study type was a randomized trial examining a new practice and finding benefit for that practice; 530 (39.4%) of all 1345 articles were classified as such.

Of the 363 articles that tested an existing medical practice, 146 (40.2%) found it ineffective compared with a previous standard or its omission (reversals), whereas 138 (38.0%) upheld the practice, and 79 (21.7%) were inconclusive. Table 1 and Figure 2 provide, for articles testing existing standard of care, a breakdown of reversal, reaffirmation, and inconclusive articles by year. Of the 146 reversal articles, most were randomized controlled trials (111 [76.0%]); 13 (8.9%) were prospective, nonrandomized studies; 20 (13.7%) were retrospective studies; 1 was a case-control study; and 1 used an alternative study design.

Articles that tested new practices were more likely to find them beneficial than articles that tested existing ones (77.1% vs 38.0%; P<.001). Conversely, articles that tested existing standards were more likely to find those practices ineffective than articles testing new practices (40.2% vs 17.0%; P<.001).

Several of the reversal articles concerned the same topic. Four articles called into question the drug aprotinin,14-17 which was widely used in cardiac surgery but found to increase mortality. Three articles addressed use of a primary rhythm control strategy for patients with atrial fibrillation.18-20 Three articles in a single

![Figure 1. A breakdown of articles concerning a medical practice.](image-url)
issue found increased risks of cardiovascular events from using the cyclooxygenase 2 inhibitors, including rofecoxib. Three articles provided extended follow-up for a trial of children randomly assigned to early myringotomy with the insertion of tympanostomy tubes or a delayed procedure. Although the procedure was the most common operation performed on children beyond the newborn period and bolstered by expert guidelines, no difference was found in an early vs delayed strategy on outcomes at 3, 6, or 9 to 11 years of age.

Three articles further contradicted routine hormone therapy in postmenopausal women. Two articles contradicted routine use of the pulmonary artery catheter, and 2 articles found worse outcomes with recommended glycemic targets (as opposed to more permissive standards) for patients with diabetes. The benefit of stenting in patients with stable coronary artery disease was undermined by the Occluded Artery Trial, Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation trial, and a follow-up quality-of-life study from the Occluded Artery Trial. Two studies suggested that although ezetimibe improves low-density lipoprotein values, it does not improve carotid artery intima media thickness.

Arthroscopic surgery of the knee for osteoarthritis was called into question by 2 studies 5 years apart, whereas vertebroplasty for osteoporotic fracture was contradicted by 2 paired articles. Adjusting for the fact that several reversals concerned the same practice, 128 medical practices were contradicted during these 10 years.

Eight of the reversals we identified overlapped with an Australian study of 156 low-value practices (Supplemental Figure; available online at http://www.mayoclinicproceedings.org). These reversals include arthroscopic surgery for knee osteoarthritis, vertebroplasty for osteoporotic fractures, endovascular repair of infrarenal abdominal aortic aneurysms, stenting in patients with stable coronary artery disease, amnioinfusion for women with meconium staining, C-reactive protein testing, screening men with the prostate specific antigen test, and routine revascularization or stress testing before surgery. Thus, we provide at least 138 unique low-value practices.

Table 2 lists the 10 selected reversals in the decade and how each article contradicted current standard of care. The Supplemental Appendix details all 146 reversals. Figure 2 shows the percentage of articles that tested standard of care and, of those, the percentage of reversals and reaffirmations. The percentage of reversals among articles that tested standard of care were constant during the decade (P=.51).

DISCUSSION

Our review of 10 years of publications in a high-impact journal involved examining 2044 articles in duplicate to identify 146 medical reversals. Reversals included medications, procedures, diagnostic tests, screening tests, and even monitoring and treatment guiding devices. We were unable to identify any class of medical practice that did not have some reversal of standard of care (Supplemental Appendix).

The bispectral index monitor (BIS) illustrates many of the principles of medical reversal. Although rare, anesthesia awareness (or intraoperative awareness) is debilitating and is associated with posttraumatic stress disorder and anxiety.

The BIS monitor was developed to ensure that patients were receiving adequate anesthesia by using a single electroencephalographic lead to calculate a
TABLE 2. Key Reversals, 2001-2010

<table>
<thead>
<tr>
<th>Reference, year</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antimicrobial treatment in diabetic women with asymptomatic bacteriuria</strong> (Harding et al. 48, 2002)</td>
<td>In contrast to European societies, several groups in the United States recommended screening and treating for asymptomatic bacteriuria in women with diabetes. This randomized trial found that although this practice leads to more antibiotic use, it did not reduce complications or improve the time to symptomatic infection.</td>
</tr>
<tr>
<td><strong>Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem-cell transplantation in high-risk breast cancer</strong> (Tallman et al. 19, 2003)</td>
<td>Multiple studies have claimed that high-dose chemotherapy with stem cell transplantation improves disease-free survival at 3 years to 65%-70%, an improvement of 20%-30% beyond standard adjuvant chemotherapy. High-dose chemotherapy and autologous stem cell transplantation became a common, costly, and controversial practice for more than a decade. This trial randomized patients with primary breast cancer with involvement of at least 10 ipsilateral axillary lymph nodes to standard adjuvant chemotherapy vs adjuvant chemotherapy followed by high-dose chemotherapy and stem cell transplant. The study arm was found to reduce risk of relapse, but no improvement in survival was found.</td>
</tr>
<tr>
<td><strong>Control of exposure to mite allergen and allergen-impermeable bed covers for adults with asthma</strong> (Woodcock et al. 54, 2003)</td>
<td>The cost of impermeable bed covers is in the millions of dollars annually, whereas the cost of all preventive interventions for asthma and allergic rhinitis is in the billions. US and European guidelines recommend these covers be used among many patients with asthma. This double-blind, randomized, placebo-controlled trial of &gt;1100 patients found no benefit on any clinical or physiologic outcome for this practice.</td>
</tr>
<tr>
<td><strong>Methylprednisolone, valacyclovir, or the combination for vestibular neuritis</strong> (Strupp et al. 58, 2004)</td>
<td>The cause of vestibular neuritis is presumed to be a viral infection, and yet it is unknown whether corticosteroids, an antiviral medication, or a combination of both have any benefit in treating this disease. At the time of this publication, physicians prescribed either or both. A prospective, randomized, double-blind, 2-by-2 factorial trial was performed assessing whether placebo, methylprednisolone, valacyclovir, or a combination of the 2 would improve symptoms. Only the corticosteroids, and not the antiviral, improved the recovery of patients with vestibular neuritis.</td>
</tr>
<tr>
<td><strong>Mild intraoperative hypothermia during surgery for intracranial aneurysm</strong> (Todd et al. 60, 2005)</td>
<td>Hypothermia was found to be helpful as a neurosurgical adjunct in 1955, especially for ischemic and traumatic insults. At the time of this publication, the practice was used in nearly 50% of aneurysm surgeries. This large randomized study, the Intraoperative Hypothermia for Aneurysm Surgery Trial (IHAST), found no improvement in neurologic outcomes with hypothermia, while noting an increase in bacterial infections with the intervention.</td>
</tr>
<tr>
<td><strong>Optimal medical therapy with or without PCI for stable coronary disease</strong> (Boden et al. 59, 2007)</td>
<td>Although treatment guidelines recommended an initial approach of intensive medical therapy, reduction of risk factors, and lifestyle modification (optimal medical therapy) for patients with stable coronary artery disease, percutaneous coronary intervention (PCI) was still a common initial treatment strategy for patients with stable coronary artery disease at the time this study was performed. The authors found that PCI added to optimal medical therapy did not reduce the risk of death, myocardial infarction, or other major cardiovascular events.</td>
</tr>
<tr>
<td><strong>In vitro fertilization with preimplantation genetic screening</strong> (Mastenbroek et al. 64, 2007)</td>
<td>Because low pregnancy rates in women of advanced maternal age undergoing in vitro fertilization (IVF) may result from chromosomal abnormalities, the use of preimplantation genetic screening had become increasingly more common at the time of this study. However, this multicenter, double-blind randomized controlled trial comparing IVF with and without preimplantation genetic screening found that screening significantly reduced rates of ongoing pregnancies and live births after IVF in women of advanced maternal age.</td>
</tr>
<tr>
<td><strong>Effects of intensive glucose lowering in type 2 diabetes</strong> (Action to Control Cardiovascular Risk in Diabetes Study Group et al. 68, 2008)</td>
<td>A target hemoglobin A1c of 7.0% or less as recommended for most patients with diabetes. The Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial found that target of &lt;7.0% sustained for 3.5 years increased mortality and did not significantly reduce major cardiovascular events compared with a more permissive goal.</td>
</tr>
<tr>
<td><strong>Revascularization versus medical therapy for renal-artery stenosis</strong> (ASTRAL Investigators et al. 70, 2009)</td>
<td>Renal artery stenosis is associated with hypertension and kidney disease, but it is unclear if the relationship is causal. Despite this uncertainty, data from studies in the United States indicate that revascularization is performed in 16% of patients with newly diagnosed atherosclerotic renovascular disease and hypertension. This large randomized trial of revascularization with medical management vs medical management alone found substantial risks but no evidence of benefit from revascularization in this population.</td>
</tr>
<tr>
<td><strong>Gentamicin-collagen sponge for infection prophylaxis in colorectal surgery</strong> (Bennett-Guerrero et al. 72, 2010)</td>
<td>The gentamicin-collagen sponge has been approved for use in numerous countries and used in millions of patients worldwide since 1985. A single-center, randomized trial found a 70% decrease in surgical site infection with implantation of the sponge. However, this large, multicenter, phase 3 trial found that the gentamicin-collagen sponge paradoxically resulted in significantly more surgical site infections, was associated with more visits to the emergency department or surgical office, and more frequently precipitated subsequent hospitalization for the infection.</td>
</tr>
</tbody>
</table>
Evidence, completed a review of 3000 med-off between bene
consequences of reversal. When medical
Previously, we have considered the causes and
most comprehensive study of medical reversal.
be found to have no net bene
high percentage of all practices may ultimately
complements these data and suggests that a
of unknown effectiveness. Our investigation
and the medical system are immense. As
common end points, such as mortality.
and the medical monitor was designed to prevent a rare event
monitoring strategy found no benefit for the device on anesthesia awareness. Many reversals have
similar narratives. Although there is a weak evi
dence base for some practice, it gains acceptance largely through vocal support from prominent advocates and faith that the mechanism of action is sound. Later, future trials undermine the therapy, but removing the contradicted practice often proves challenging. Although the BIS monitor was designed to prevent a rare event (anesthesia awareness), many reversals concern common end points, such as mortality.

Recently, a project of BMJ, entitled Clinical Evidence, completed a review of 3000 medical practices. The project found that slightly more than a third of medical practices are effective or likely to be effective; 15% are harmful, unlikely to be beneficial, or a trade-off between benefits and harms; and 50% are of unknown effectiveness. Our investigation complements these data and suggests that a high percentage of all practices may ultimately be found to have no net benefits.

To our knowledge, this is the largest and most comprehensive study of medical reversal. Previously, we have considered the causes and consequences of reversal. When medical practices are instituted in error, most often on the basis of premature, inadequate, biased, and conflicted evidence, the costs to society and the medical system are immense. As such, we favor policies that minimize reversal. Nearly all such measures involve raising the bar for the approval of new therapies and asking for evidence before the widespread adoption of novel techniques. In all but the rarest cases, large, robust, pragmatic randomized trials measuring hard end points (with sham controls for studies of subjective end points) should be required before approval or acceptance. Our position is in contrast to efforts to lower standards for device and drug approval, which further erodes the value of the regulatory process.

One surprising type of reversal we observed was potentially beneficial therapies being with-held because of unfounded concerns about their potential to cause harm. Long-standing concerns that vaccinations precipitate flare of multiple sclerosis led many physicians to omit this intervention, but the concerns were largely undermined by the results of 2 studies in 2001. Concerns that oral contraceptives increase lupus flares created reluctance to prescribe this class of medications to women. This practice may contribute to a higher rate of elective abortions among patients with lupus. In 2005, 2 trials reported that oral contraceptives do not increase lupus flares. Although the American College of Obstetrics recommended that epidural anesthesia be delayed until cervical dilation has reached 4 cm —out of concern that earlier administration increases rates of cesarean section—randomized trials reported that this fear was unfounded. Warnings that turned out to be wrong represent a unique form of reversal and raise questions about other dubious restrictions taken at face value, for instance, that patients with Clostridium difficile infection should not be treated with antimotility agents for fear of increasing rates of toxic megacolon. Discerning readers may yet identify other novel patterns of contradiction.

The current study has several limitations. Our choice of journal was made on the basis of impact factor rankings; thus, we are unsure whether our results apply to all journals. As in any study of published research findings, one may wonder whether there exists a publication bias favoring certain studies, in this case, those that contradict standard of care. However, the testing of standard of care is rarely done and accordingly is in itself noteworthy. It seems unlikely that there exists a selection filter against reaffirmation articles.

Our classification scheme was based on prior work, but others may have alternative preferences for grouping medical articles. Whether a medical practice was considered new or existing was decided on the basis of the article’s abstract, introduction, and discussion.
We did not perform an independent search to verify that existing practices were indeed in use and new practices were not. As such, we may have made errors both of inclusion and exclusion. Some authors may have chosen to downplay a therapy’s real-world use, whereas others may have chosen to overemphasize it. An independent evaluation of practice patterns would have strengthened our investigation but would have been overly time-consuming because it would have required investigation of hundreds of topics, many of which are common medications that lack unique coding for their varying indications.

The reversals we have identified by no means represent the final word for any of these practices. Simply because newer, larger, better controlled or designed studies contradict standard of care does not necessarily mean that older practices are wrong and new ones are right. On average, however, better designed, controlled, and powered studies reach more valid conclusions. Nevertheless, the reversals we have identified at the very least call these practices into question. Some practices ought to be abandoned, whereas others warrant retesting in more powerful investigations. One of the greatest virtues of medical research is our continual quest to reassess it.

It is likely that others may feel differently about some of the reversals we have identified (Supplemental Appendix). Although we performed our analysis in duplicate, with little disagreement, others may nevertheless draw different conclusions. We interpreted articles in good faith, as the authors presented the results. In addition, the purpose of our investigation was to outline broad trends in medical practice and identify a large number of potential low-value practices. We do not seek to issue a final determination regarding any particular practice. Changing a dozen classifications would make little difference in the interpretation of our results.

CONCLUSION

We present 146 medical practices that were reversed in 10 years of publications in a high-profile journal. Our results may be of interest to practitioners and policymakers who seek to identify low-value practices and methodologists and scientists who are interested in the patterns of medical research.

ACKNOWLEDGMENTS

The views and opinions of Dr Prasad do not necessarily reflect those of the National Cancer Institute or National Institutes of Health.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental online material can be found online at http://www.mayoclinicproceedings.org.

Correspondence: Address to Vinay Prasad, MD, Medical Oncology Branch, National Cancer Institute, National Institutes of Health, 10 Center Dr 10/12N226, Bethesda, MD 20892 (vinayak.prasad@nih.gov).

REFERENCES


