Medicine Is Not a Steak

Brian D. Sites, MD, MS,† and Thomas Dodds, MD*‡*

(*Reg Anesth Pain Med* 2015;40: 399–400)

You have probably heard this story before. The United States has the most expensive health care system in the world and the quality of the product has been called into question. In 2011, according to the Centers for Disease Control and Prevention, the United States spent 17.9% of its entire gross domestic product on health care.[1] This amounted to 2.7 trillion dollars or $8680 per person. The Dartmouth Atlas, studying the Medicare population, has drawn national attention to the repeated demonstration that populations living in regions with higher levels of spending and greater utilization of health care services do not experience better health care outcomes or better quality of care.[2]

These characteristics of high costs and questionable quality have prompted many health systems and executive leaders to turn toward industry models for guidance. One such example is the utilization of the *Lean Six Sigma* techniques. This business management strategy addresses processes and waste issues while focusing on variation and innovation to facilitate operational perfection. In a highly referenced essay entitled "Big Med," physician-author Gawande[3] argues in *The New Yorker* that the popular restaurant chain *Cheesecake Factory* could serve as the ideal model for improving health care delivery. For the *Cheesecake Factory*, Gawande outlines the elaborate production line including defined teams, rules, ingredients, systems, technologies, and protocols that result in the predictable production of high-quality gourmet food. He concludes:

We've let health-care systems provide us with the equivalent of greasy-spoon fare at four-star prices, and the results have been ruinous. The Cheesecake Factory model represents our best prospect for change.

Following the Cheesecake Factory model, many health care systems are emphasizing the need to standardize services and interventions.[4,5] Just as in the production of a succulent steak, standardization of process is thought to be a key technique to optimize value (outcomes/costs) for a health care product or service. We agree that there is substantial opportunity to improve health care value through the standardization of the specific microsystems that support comprehensive health care delivery. In fact, medical services and processes need to be standardized to be effectively studied through quality improvement methodologies such as statistical process control charts.

From the Departments of *Anesthesiology, †Orthopedic Surgery, and ‡Pain Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH.
Accepted for publication May 6, 2015.
Address correspondence to: Brian D. Sites, MD, MS, Dartmouth-Hitchcock Medical Center, One Medical Center Dr, Lebanon, NH 03756 (e-mail: brian.d.sites@hitchcock.org).
The authors declare no conflict of interest.
The Department of Anesthesiology at Dartmouth Medical School provided protected research time for Brian Sites and Thomas Dodds to support this project.

ISSN: 1098-7339
DOI: 10.1097/AAP.0000000000000291

However, it is unclear if borrowing industrial design models of production line techniques is a universally effective strategy for improving value. This uncertainty around value is heightened when the standardization involves major therapeutic decisions and interventions. The biggest threat to the wisdom of standardizing health care delivery in the name of value is the actual quality of the "evidence-based" information driving the policies and procedures. Standardization of health care services assumes that the processes and interventions (ie, key ingredients of the steak) that are being standardized are safe and do, in fact, lead to the best outcome for a given expenditure.

Prasad and colleagues examined 10 years of published research in the *New England Journal of Medicine*. Of the 363 articles testing standard of care, 146 (40.2%) reversed that practice, whereas 138 (38.0%) reaffirmed it.[6] This article was provocative and stimulated discussion both in the medical community and popular press, as there was significant morbidity and costs associated with the wrong therapies (eg, strict glycemic control in persons with diabetes) masquerading as gold standards. The authors did not commit to exactly why gold standard therapies are so often wrong but alluded to issues related to the excessive use of observational data (such as expert opinion) to drive their formation. Although not mentioned in the Prasad article, there is the ever-present dark plume of suspicion hanging over the plethora of pharmaceutical and industry-sponsored research. Fraud as well as research errors have become a serious concern resulting in a crisis of retracted publications that expose thousands of patients to wrong treatments each year.[7] Finally, there are legitimate concerns regarding the generalizability of randomized controlled trial findings into "real-world" practices within complex health care systems.[8] To add more complexity around what constitutes best practice are the plethora of recommendations from the Joint Commission, which often forces health systems to invest large amounts of human and monetary capital into areas that are unlikely to improve the value of health care services.

In the Prasad article, the 95% confidence interval around the 40% incorrect rate was 35% to 45%. This generates the question that if we standardize the *wrong* therapy at a rate between 35% and 45%, is it conceivable that actually promoting variation in treatment would provide better outcomes? From an industry perspective, such variation in care would likely be labeled as "unwarranted variation" and considered evidence for a dysfunctional delivery process. However, based on the Prasad success rate and assuming a dichotomous therapy and outcome, an individualized practice would only need to perform 11% better than a coin toss (61%) to outperform standardization of "best practice."

The difficult reality is that the "science" of health care delivery is in its infancy. There is so much that we do not know and what we do know may not even be translatable between complex health care systems. With respect to the aforementioned value equation (outcomes/costs) which is driving health care reform, administrators, physicians, and insurance companies cannot even agree upon appropriate numerator definitions (let alone how to measure them). This leads us to conclude that health care would be so much easier if all we had to do is make a good steak.

REFERENCES

1. Health United States, 2013. Available at: http://www.cdc.gov/nchs/fastats/health-expenditures.htm. Accessed March 18, 2015.
2. The Dartmouth Atlas of Health Care, 2015. Available at: http://www.dartmouthatlas.org. Accessed March 18, 2015.
3. Gawande A. Big Med. *The New Yorker*. Aug 13, 2012.
4. Berwick DM, Hackbarth AD. Eliminating waste in US health care. *JAMA*. 2012;307:1513–1516.
5. Feng Q, Manuel CM. Under the knife: a national survey of six sigma programs in US healthcare organizations. *Int J Health Care Qual Assur*. 2008;21:535–547.
6. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc*. 2013;88:790–798.
7. Steen RG. Retractions in the medical literature: how many patients are put at risk by flawed research? *J Med Ethics*. 2011;37:688–692.
8. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005;365:82–93.

MAYO CLINIC

EDITORIAL

August 2013
Volume 88
Number 8

MAYO CLINIC
PROCEEDINGS

How Many Contemporary Medical Practices Are Worse Than Doing Nothing or Doing Less?

See also page 790

How many contemporary medical practices are not any better than or are worse than doing nothing or doing something else that is simpler or less expensive? This is an important question, given the negative repercussions for patients and the health care system of continuing to endorse futile, inefficient, expensive, or harmful interventions, tests, or management strategies. In this issue of *Mayo Clinic Proceedings*, Prasad et al[1] describe the frequency and spectrum of medical reversals determined from a review of all the articles published over a decade (2001-2010) in *New England Journal of Medicine* (*NEJM*). Their work extends a previous effort[2] that had focused on data from a single year and had suggested that almost half of the established medical practices that are tested are found to be no better than a less expensive, simpler, or easier therapy or approach. The results from the current larger sample of articles[1] are consistent with the earlier estimates: 27% of the original articles relevant to medical practices published in *NEJM* over this decade pertained to testing established practices. Among them, reversal and reaffirmation studies were approximately equally common (40.2% vs 38%). About two-thirds of the medical reversals were recommended on the basis of randomized trials. Even though no effort was made to evaluate systematically all evidence on the same topic (eg, meta-analyses including all studies published before and after the specific *NEJM* articles), the proportion of medical reversals seems alarmingly high. At a minimum, it poses major questions about the validity and clinical utility of a sizeable portion of everyday medical care.

Are these figures representative of the medical literature and evidence base at large? The sample assembled by Prasad et al is highly impressive, but it accounts for less than 1% of all randomized trials published in the same decade (an estimated >10,000 per year) and an even more infinitesimal portion of other types of study designs. If one could extrapolate from this sample by proportion, perhaps there have been several tens of thousands of medical reversal studies across all 23 million articles entered to date in PubMed. One has to be cautious with extrapolations, however. *New England Journal of Medicine* is clearly different from other journals in many ways besides having the highest impact factor among the list of 155 general and internal medicine journals.[3] It is widely read, and it has high visibility and impact both on the mass media and on medical practitioners. In this regard, the collection of 146 medical reversals reviewed by Prasad et al is a compendium of widely known, visible examples, and thus it can make excellent reading for medical practitioners and researchers, teachers, and trainees. At the same time, this characteristic is also a disadvantage: the articles published by *NEJM* are a highly selected sample, probably susceptible to publication and selective outcome reporting bias. There is substantial empirical evidence that the effect sizes of randomized trials published in *NEJM*, *Lancet*, or *JAMA* (the top 3 general and internal medicine journals in terms of impact factor[3]) are markedly inflated, in particular for small trials[4]; conversely, the effect sizes for large trials are similar to those seen in large trials on the same topic in other journals.[4] The interpretation of the results in *NEJM* is also likely to be more exaggerated compared with other journals because authors may feel pressured to claim that the results are impressive in order to get their work published in such a

competitive venue.[5] Finally, when the quantitative data on effect sizes are examined, studies published in *NEJM* and other major journals have higher informativity (information gain or change in entropy),[6] ie, their results do change previous evidence more than the change incurred by the results of studies published elsewhere.

On the basis of these considerations, the frequency of medical reversals published in *NEJM* may be somewhat higher than what might be seen in publications in other journals. However, there are also some other counterbalancing forces that could cause bias in the opposite direction. For example, evaluations published in *NEJM* are likely to focus on commonly used, established medical practices. Such commonly used practices are likely to have had at least some previous evidence generated in the past supporting their use. Conversely, established interventions that are more narrowly applied and specialized (eg, those for which randomized trials might be published in small-circulation, highly specialized journals) may have been originally endorsed with even more sparse and worse-quality evidence, or even no evidence at all.

Other empirical approaches may also offer some insight about how commonly useless or even harmful treatments are endorsed. The *Cochrane Database of Systematic Reviews* has assembled considerable current medical evidence from clinical trials on diverse interventions. An empirical evaluation of Cochrane reviews in 2004 showed that most (47.8%) concluded that there is insufficient evidence to endorse the examined interventions.[7] A repeated evaluation in 2011 showed that this trend has not changed, with the percentage of insufficient evidence remaining as high as 45%.[8] Often, non-Cochrane reviews tend to have more positive conclusions about the assessed interventions, but it is unclear whether this finding reflects genuine superiority of the assessed interventions or bias in the interpretation of the results.[9] Although a substantial proportion of interventions are clearly harmful or inferior to others, many are still being used because of reluctance or resistance to abandoning them.[10] Some are even widely used despite the poor evidence, as Prasad et al[1] eagerly highlight with several examples. Moreover, different medical specialties may vary in their lack of evidence—eg, primary care, surgery, and dermatology interventions more frequently lack evidence to support their use compared with internal medicine interventions.[11]

Most new interventions that are successfully introduced into medical care have small effects that translate to modest, incremental benefits.[12] Empirical evaluations have suggested that well-validated large benefits for measurable outcomes such as mortality are uncommon in medicine.[13] Under these circumstances, even subtle changes in the composition and spectrum of the treated population over time, emergence of previously unrecognized toxicities, or a relatively disadvantageous cost can easily tip the evidence balance against the use of these interventions. Moreover, the introduction of interventions with limited or no evidence of benefit continues at fast pace even in specialties that have a strong tradition of evidence-based methods. For example, in almost half (48%) of the recommendations in major cardiology guidelines, the level of evidence is grade C, ie, limited evidence and expert opinion have a highly influential presence.[14] Once we divert beyond traditional treatments (eg, drugs or devices) to diagnostic tools, prognostic markers, health systems, and other health care measures, randomized trials are a rarity.[15] For example, it has been estimated that, on average, there are only 37 publications per year of randomized trials assessing the effectiveness of diagnostic tests.[15] Some modern technologies (eg, "omics") promise to introduce new tools into medical management at such a high pace that many investigators are wary of even thinking about the possibility of randomized testing. Despite better laboratory science, fascinating technology, and theoretically mature designs after 65 years of randomized trials, ineffective, harmful, expensive medical practices are being introduced more frequently now than at any other time in the history of medicine. Under the current mode of evidence collection, most of these new practices may never be challenged.

The data collected by Prasad et al[1] offer some hints about how this dreadful scenario might be aborted. The 146 medical reversals that they have assembled are, in a sense, examples of success stories that can inspire the astute clinician and clinical investigator to challenge the

Mayo Clin Proc. ■ August 2013;88(8):779-781 ■ http://dx.doi.org/10.1016/j.mayocp.2013.05.010
www.mayoclinicproceedings.org

status quo and realize that doing less is more.[16] It is not with irony that I call these disasters "success stories." If we can learn from them, these seemingly disappointing results may be extremely helpful in curtailing harms to patients and cost to the health care system. Although it is important to promote effective practices ("positive success stories"), it is also important to promote and disseminate knowledge about ineffective practices that should be reversed and abandoned. Also, research is needed to find the most efficient ways of applying the knowledge learned from these "negative" studies. Does it suffice to compile lists of practices that should be abandoned?[10] What types of educational approaches and reinforcement could enhance their abandonment? What are the obstacles (commercial, professional, system inertia, or other) that hinder this disimplementation step and how can they be best overcome? Are there some incentives that we can offer to practitioners and health systems to apply this "negative" knowledge toward simplifying and streamlining their practices?

Some of the messaging may require inclusion in guidelines, given the widespread attention that these documents gain, particularly when issued by authoritative individuals or groups, and their capacity to affect clinical practice. Should we require generally higher levels of evidence before practice guidelines are recommended? Moreover, if and when practice guidelines are discredited or overturned by additional information, should notification of practitioners and the public not be undertaken with the same, if not more, vigor as when the practices were first recommended?

Finally, are there incentives and anything else we can do to promote testing of seemingly established practices and identification of more practices that need to be abandoned? Obviously, such an undertaking will require commitment to a rigorous clinical research agenda in a time of restricted budgets. However, it is clear that carefully designed trials on expensive practices may have a very favorable value of information, and they would be excellent investments toward curtailing the irrational cost of ineffective health care.

John P. A. Ioannidis, MD, DSc
Stanford Prevention Research Center
Department of Medicine
Department of Health Research and Policy
Stanford University School of Medicine

Department of Statistics
Stanford University School of Humanities and Sciences
Stanford, CA

Correspondence: Address to John P. A. Ioannidis, MD, DSc, Stanford Prevention Research Center, Medical School Office Bldg Rm X306, 1265 Welch Rd, Stanford, CA 94305 (jioannid@stanford.edu).

REFERENCES

1. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc*. 2013;88(8):790-798.
2. Prasad V, Gall V, Cifu A. The frequency of medical reversal. *Arch Intern Med*. 2011;171:1675-1676.
3. ISI Web of Science. Journal Citation Reports. Accessed May 9, 2013.
4. Siontis KC, Evangelou E, Ioannidis JP. Magnitude of effects in clinical trials published in high-impact general medical journals. *Int J Epidemiol*. 2011;40(5):1280-1291.
5. Young NS, Ioannidis JP, Al-Ubaydli O. Why current publication practices may distort science. *PLoS Med*. 2008;5(10):e201.
6. Evangelou E, Siontis KC, Pfeiffer T, Ioannidis JP. Perceived information gain from randomized trials correlates with publication in high-impact factor journals. *J Clin Epidemiol*. 2012;65(12):1274-1281.
7. El Dib RP, Atallah AN, Andriolo RB. Mapping the Cochrane evidence for decision making in health care. *J Eval Clin Pract*. 2007;13(4):689-692.
8. Villas Boas PJ, Spagnuolo RS, Kamegasawa A, et al. Systematic reviews showed insufficient evidence for clinical practice in 2004: what about in 2011? the next appeal for the evidence-based medicine age [published online ahead of print July 3, 2012]. *J Eval Clin Pract*. http://dx.doi.org/10.1111/j.1365-2753.2012.01877.x.
9. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D. Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. *J Clin Epidemiol*. 2009;62(4):380-386.e1.
10. Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. *JAMA*. 2012;307(1):37-38.
11. Matzen P. How evidence-based is medicine? A systematic literature review [in Danish]. *Ugeskr Laeger*. 2003;165(14):1431-1435.
12. Djulbegovic B, Kumar A, Glasziou PP, et al. New treatments compared to established treatments in randomized trials. *Cochrane Database Syst Rev*. 2012;10:MR000024.
13. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012;308(16):1676-1684.
14. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines [published correction appears in *JAMA*. 2009;301(15):1544]. *JAMA*. 2009;301(8):831-841.
15. Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol*. 2012;65(3):282-287.
16. Grady D, Redberg RF. Less is more: how less health care can result in better health. *Arch Intern Med*. 2010;170(9):749-750.

MAYO CLINIC

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 ■ Mayo Clin Proc. 2013;88(8):790-798

For editorial comment, see page 779

From the National Cancer Institute, National Institutes of Health, Bethesda, MD (V.P.); Department of Medicine, Yale University, New Haven, CT (A.V.); Department of Medicine (C.T., J.R., S.J.C.) and Feinberg School of Medicine (D.B., S.S.), Northwestern University, Chicago, IL; Department of Medicine, Lankenau Medical Center, Philadelphia, PA (M.C.); Department of Medicine (S.Q.) and Department of Surgery (V.G.), George Washington University, Washington, DC; Department of Medicine, University of Maryland, Baltimore, MD (N.H.); and Department of Medicine, University of Chicago, Chicago, IL (A.C.).

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.[9] 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

Mayo Clin Proc. ■ August 2013;88(8):790-798 ■ http://dx.doi.org/10.1016/j.mayocp.2013.05.012
www.mayoclinicproceedings.org ■ Published by Elsevier Inc on behalf of Mayo Foundation for Medical Education and Research

generate more than 150 low-value practices.[11] 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."[12]

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.[4] 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.[13] 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 χ^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

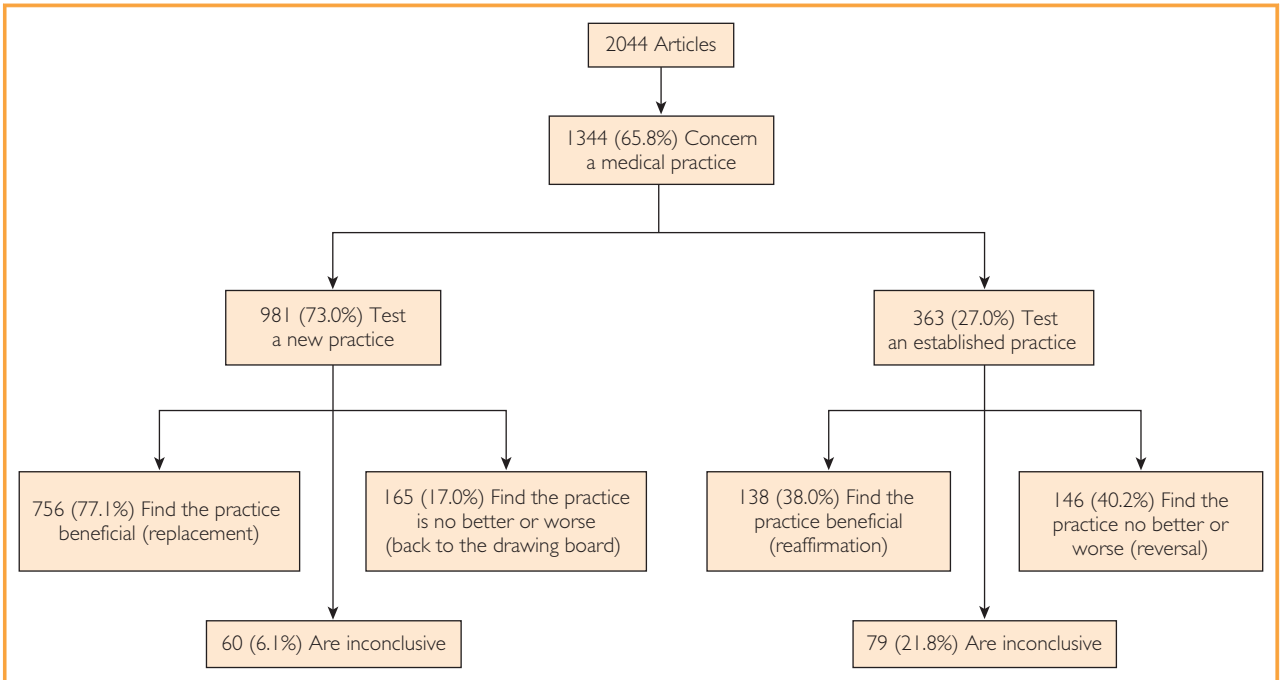

FIGURE 1. A breakdown of articles concerning a medical practice.

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.

TABLE 1. Number (Percentage) of Reversal, Reaffirmation, and Inconclusive Articles by Year

Year	Reversal	Reaffirmation	Inconclusive
2001 (n=48)	14 (29.2)	20	14
2002 (n=26)	12 (46.2)	9	5
2003 (n=31)	12 (38.7)	12	7
2004 (n=33)	12 (36.4)	15	6
2005 (n=41)	19 (46.3)	14	8
2006 (n=20)	12 (60.0)	5	3
2007 (n=54)	18 (33.3)	17	19
2008 (n=32)	15 (46.9)	13	4
2009 (n=35)	16 (45.7)	16	3
2010 (n=43)	16 (37.2)	17	10
Total (N=363)	146 (40.2)	138 (38.0)	79 (21.7)

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 (27.3%) 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

Mayo Clin Proc. ■ August 2013;88(8):790-798 ■ http://dx.doi.org/10.1016/j.mayocp.2013.05.012
www.mayoclinicproceedings.org

issue found increased risks of cardiovascular events from using the cyclooxygenase 2 inhibitors, including rofecoxib.[21-23] 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[24] and bolstered by expert guidelines,[25] no difference was found in an early vs delayed strategy on outcomes at 3,[24] 6,[26] or 9 to 11 years of age.[27]

Three articles further contradicted routine hormone therapy in postmenopausal women.[28-30] Two articles contradicted routine use of the pulmonary artery catheter,[3,31] and 2 articles found worse outcomes with recommended glycemic targets (as opposed to more permissive standards) for patients with diabetes.[32,33] The benefit of stenting in patients with stable coronary artery disease was undermined by the Occluded Artery Trial,[34] Clinical Outcomes Utilizing Revascularization and Aggressive Drug Evaluation[35] trial, and a follow-up quality-of-life study from the Occluded Artery Trial.[36] Two studies suggested that although ezetimibe improves low-density lipoprotein values, it does not improve carotid artery intima media thickness.[37,38] Arthroscopic surgery of the knee for osteoarthritis was called into question by 2 studies 5 years apart,[39,40] whereas vertebroplasty for osteoporotic fracture was contradicted by 2 paired articles.[41,42] 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[11] (Supplemental Figure; available online at http://www.mayoclinicproceedings.org). These reversals include arthroscopic surgery for knee osteoarthritis,[40] vertebroplasty for osteoporotic fractures,[17] endovascular repair of infrarenal abdominal aortic aneurysms,[43] stenting in patients with stable coronary artery disease,[1] amnioinfusion for women with meconium staining,[44] C-reactive protein testing,[45] screening men with the prostate specific antigen test,[46] and routine revascularization or stress testing before surgery.[47] Thus, we provide at least 138 unique low-value practices.

Table 2[48-73] 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.[74] The BIS monitor was developed to ensure that patients were receiving adequate anesthesia by using a single electroencephalographic lead to calculate a

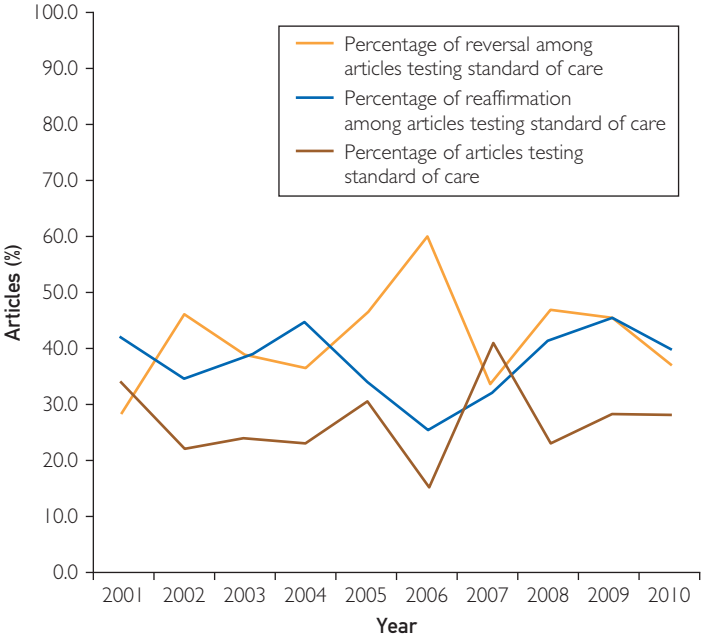

FIGURE 2. Percentage of reversal, reaffirmation, and all articles testing standard of care.

TABLE 2. Key Reversals, 2001-2010

Reference, year	Description
Antimicrobial treatment in diabetic women with asymptomatic bacteriuria (Harding et al,[48] 2002)	In contrast to European societies, several groups[49,50] 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
Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem-cell transplantation in high-risk breast cancer (Tallman et al,[51] 2003)	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.[52,53] 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
Control of exposure to mite allergen and allergen-impermeable bed covers for adults with asthma (Woodcock et al,[54] 2003)	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.[55] US[56] and European[57] guidelines recommend these covers be used among many patients with asthma. This double-blind, randomized, placebo-controlled trial of >1100 patients found no benefit on any clinical or physiologic outcome for this practice
Methylprednisolone, valacyclovir, or the combination for vestibular neuritis (Strupp et al,[58] 2004)	The cause of vestibular neuritis is presumed to be a viral infection,[59] 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
Mild intraoperative hypothermia during surgery for intracranial aneurysm (Todd et al,[60] 2005)	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.[61] 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
Optimal medical therapy with or without PCI for stable coronary disease (Boden et al,[35] 2007)	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.[62,63] The authors found that PCI added to optimal medical therapy did not reduce the risk of death, myocardial infarction, or other major cardiovascular events
In vitro fertilization with preimplantation genetic screening (Mastenbroek et al,[64] 2007)	Because low pregnancy rates in women of advanced maternal age undergoing in vitro fertilization (IVF) may result from chromosomal abnormalities, the use of preimplanation genetic screening had become increasingly more common at the time of this study.[65-67] 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
Effects of intensive glucose lowering in type 2 diabetes (Action to Control Cardiovascular Risk in Diabetes Study Group et al,[68] 2008)	A target hemoglobin A_{1c} of 7.0% or less as recommended for most patients with diabetes.[69] The Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial found that target of <7.0% sustained for 3.5 years increased mortality and did not significantly reduce major cardiovascular events compared with a more permissive goal
Revascularization versus medical therapy for renal-artery stenosis (ASTRAL Investigators et al,[70] 2009)	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.[71] 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
Gentamicin-collagen sponge for infection prophylaxis in colorectal surgery (Bennett-Guerrero et al,[72] 2010)	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.[73] 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

Mayo Clin Proc. ■ August 2013;88(8):790-798 ■ http://dx.doi.org/10.1016/j.mayocp.2013.05.012
www.mayoclinicproceedings.org

dimensionless measure of consciousness. In theory, anesthesia could be titrated to the BIS reading. In 1997, the US Food and Drug Administration approved the device. Only 2 trials existed before the reversal study. One, an industry-sponsored trial, did not use a standardized protocol for the comparator arm and found the device reduced awareness.[75] The other was underpowered to make any conclusions.[76] Nevertheless, the monitor's use increased. By July 2007, half of all operating rooms in the United States had a BIS monitor.[77] Then in 2008, a large, randomized trial comparing the BIS monitor with a standardized sedation monitoring strategy found no benefit for the device on anesthesia awareness.[78] Many reversals have similar narratives.[4] Although there is a weak evidence 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.[79,80] 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,[81] 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.[4-6,82] When medical practices are instituted in error, most often on the basis of premature, inadequate, biased, and conflicted evidence,[4] the costs to society and the medical system are immense.[5] As such, we favor policies that minimize reversal. Nearly all such measures involve raising the bar for the approval of new therapies[6,83,84] and asking for evidence before the widespread adoption of novel techniques. In all but the rarest cases,[82] 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,[85] which further erodes the value of the regulatory process.

One surprising type of reversal we observed was potentially beneficial therapies being withheld 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.[86,87] 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.[88] In 2005, 2 trials reported that oral contraceptives do not increase lupus flares.[89,90] Although the American College of Obstetrics recommended that epidural anesthesia be delayed until cervical dilation has reached 4 cm[91]—out of concern that earlier administration increases rates of cesarean section—randomized trials reported that this fear was unfounded.[92] 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.[93] 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[5] 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,[4] 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.[94] 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

1. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med*. 2007;356(15):1503-1516.
2. Rossouw JE, Anderson GL, Prentice RL, et al; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321-333.
3. Sandham JD, Hull RD, Brant RF, et al; Canadian Critical Care Clinical Trials Group. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *N Engl J Med*. 2003;348(1):5-14.
4. Prasad V, Gall V, Cifu A. The frequency of medical reversal. *Arch Intern Med*. 2011;171(18):1675.
5. Prasad V, Cifu A, Ioannidis JPA. Reversals of established medical practices: evidence to abandon ship. *JAMA*. 2012;307(1):37-38.
6. Prasad V, Cifu A. Medical reversal: why we must raise the bar before adopting new technologies. *Yale J Biol Med*. 2011;84(4):471-478.
7. Ioannidis JA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294(2):218-228.
8. Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA*. 2012;307(17):1801-1802.
9. Brody H. Medicine's ethical responsibility for health care reform—the top five list. *N Engl J Med*. 2010;362(4):283-285.
10. Garner S, Littlejohns P. Disinvestment from low value clinical interventions: NICEly done? *BMJ*. 2011;343:d4519.
11. Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. *Med J Aust*. 2012;197(10):556-560.
12. Grady D, Redberg RF. Less is more: how less health care can result in better health. *Arch Intern Med*. 2010;170(9):749-750.
13. Top Publications Health & Medical Sciences. http://scholar.google.com/citations?view_op=top_venues&hl=en&vq=med. Accessed June 30, 2011.
14. Mangano DT, Tudor IC, Dietzel C. The risk associated with aprotinin in cardiac surgery. *N Engl J Med*. 2006;354(4):353-365.
15. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med*. 2008;358(8):771-783.
16. Shaw AD, Stafford-Smith M, White WD, et al. The effect of aprotinin on outcome after coronary-artery bypass grafting. *N Engl J Med*. 2008;358(8):784-793.
17. Fergusson DA, Hébert PC, Mazer CD, et al. A comparison of aprotinin and lysine analogues in high-risk cardiac surgery. *N Engl J Med*. 2008;358(22):2319-2331.
18. Roy D, Talajic M, Nattel S, et al; Atrial Fibrillation and Congestive Heart Failure Investigators. Rhythm control versus rate

control for atrial fibrillation and heart failure. *N Engl J Med*. 2008;358(25):2667-2677.
19. Van Gelder IC, Hagens VE, Bosker HA, et al; Rate Control versus Electrical Cardioversion for Persistent Atrial Fibrillation Study Group. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med*. 2002;347(23):1834-1840.
20. Wyse DG, Waldo AL, DiMarco JP, et al; The Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) Investigators. A comparison of rate control and rhythm control in patients with atrial fibrillation. *N Engl J Med*. 2002;347(23):1825-1833.
21. Solomon SD, McMurray JJV, Pfeffer MA, et al; Adenoma Prevention with Celecoxib (APC) Study Investigators. Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. *N Engl J Med*. 2005;352(11):1071-1080.
22. Nussmeier NA, Whelton AA, Brown MT, et al. Complications of the COX-2 inhibitors parecoxib and valdecoxib after cardiac surgery. *N Engl J Med*. 2005;352(11):1081-1091.
23. Bresalier RS, Sandler RS, Quan H, et al; Adenomatous Polyp Prevention on Vioxx (APPROVe) Trial Investigators. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med*. 2005;352(11):1092-1102.
24. Paradise JL, Feldman HM, Campbell TF, et al. Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years. *N Engl J Med*. 2001;344(16):1179-1187.
25. *American Academy of Otolaryngology—Head and Neck Surgery 2000 Clinical Indicators Compendium*. Alexandria, VA: American Academy of Otolaryngology—Head and Neck Surgery; 2000:10.
26. Paradise JL, Campbell TF, Dollaghan CA, et al. Developmental outcomes after early or delayed insertion of tympanostomy tubes. *N Engl J Med*. 2005;353(6):576-586.
27. Paradise JL, Feldman HM, Campbell TF, et al. Tympanostomy tubes and developmental outcomes at 9 to 11 years of age. *N Engl J Med*. 2007;356(3):248-261.
28. Hays J, Ockene JK, Brunner RL, et al; Women's Health Initiative Investigators. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med*. 2003;348(19):1839-1854.
29. Hodis HN, Mack WJ, Azen SP, et al; Women's Estrogen-Progestin Lipid-Lowering Hormone Atherosclerosis Regression Trial Research Group. Hormone therapy and the progression of coronary-artery atherosclerosis in postmenopausal women. *N Engl J Med*. 2003;349(6):535-545.
30. Manson JE, Hsia J, Johnson KC, et al; Women's Health Initiative Investigators. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med*. 2003;349(6):523-534.
31. National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network, Wheeler AP, Bernard GR, Thompson BT, et al. Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N Engl J Med*. 2006;354(21):2213-2224.
32. Gerstein HC, Miller ME, Byington RP, et al; Action to Control Cardiovascular Risk in Diabetes Study Group. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559.
33. Patel A, MacMahon S, Chalmers J, et al; ADVANCE Collaborative Group. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2008;358(24):2560-2572.
34. Hochman JS, Lamas GA, Buller CE, et al; Occluded Artery Trial Investigators. Coronary intervention for persistent occlusion after myocardial infarction. *N Engl J Med*. 2006;355(23):2395-2407.
35. Boden WE, O'Rourke RA, Teo KK, et al; COURAGE Trial Research Group. Optimal medical therapy with or without PCI for stable coronary disease. *N Engl J Med*. 2007;356(15):1503-1516.
36. Mark DB, Pan W, Clapp-Channing NE, et al; Occluded Artery Trial Investigators. Quality of life after late invasive therapy for occluded arteries. *N Engl J Med*. 2009;360(8):774-783.
37. Kastelein JJ, Akdim F, Stroes ES, et al; ENHANCE Investigators. Simvastatin with or without ezetimibe in familial hypercholesterolemia. *N Engl J Med*. 2008;358(14):1431-1443.
38. Taylor AJ, Villines TC, Stanek EJ, et al. Extended-release niacin or ezetimibe and carotid intima-media thickness. *N Engl J Med*. 2009;361(22):2113-2122.
39. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002;347(2):81-88.
40. Kirkley A, Birmingham TB, Litchfield RB, et al. A randomized trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2008;359(11):1097-1107.
41. Kallmes DF, Comstock BA, Heagerty PJ, et al. A randomized trial of vertebroplasty for osteoporotic spinal fractures. *N Engl J Med*. 2009;361(6):569-579.
42. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med*. 2009;361(6):557-568.
43. De Bruin JL, Baas AF, Buth J, et al; DREAM Study Group. Long-term outcome of open or endovascular repair of abdominal aortic aneurysm. *N Engl J Med*. 2010;362(20):1881-1889.
44. Fraser WD, Hofmeyr J, Lede R, et al; Amnioinfusion Trial Group. Amnioinfusion for the prevention of the meconium aspiration syndrome. *N Engl J Med*. 2005;353(9):909-917.
45. Danesh J, Wheeler JG, Hirschfield GM, et al. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med*. 2004;350(14):1387-1397.
46. Andriole GL, Crawford ED, Grubb RL, et al; PLCO Project Team. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360(13):1310-1319.
47. McFalls EO, Ward HB, Moritz TE, et al. Coronary-artery revascularization before elective major vascular surgery. *N Engl J Med*. 2004;351(27):2795-2804.
48. Harding GK, Zhanel GG, Nicolle LE, Cheang M; Manitoba Diabetes Urinary Tract Infection Study Group. Antimicrobial treatment in diabetic women with asymptomatic bacteriuria. *N Engl J Med*. 2002;347(20):1576-1583.
49. Patterson JE, Andriole VT. Bacterial urinary tract infections in diabetes. *Infect Dis Clin North Am*. 1997;11(3):735-750.
50. US Preventive Services Task Force. Screening for asymptomatic bacteriuria, hematuria and proteinuria. *Am Fam Phys*. 1990;42(2):389-395.
51. Tallman MS, Gray R, Robert NJ, et al. Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem-cell transplantation in high-risk breast cancer. *N Engl J Med*. 2003;349(1):17-26.
52. Somlo G, Doroshow JH, Forman SJ, et al. High-dose chemotherapy and stem-cell rescue in the treatment of high-risk breast cancer: prognostic indicators of progression-free and overall survival. *J Clin Oncol*. 1997;15(8):2882-2893.
53. Tomas JF, Perez-Carrion R, Escudero A, Lopez-Lorenzo JL, Lopez-Pascual J, Fernandez-Ranada JM. Results of a pilot study of 40 patients using high-dose therapy with hematopoietic rescue after standard-dose adjuvant therapy for high-risk breast cancer. *Bone Marrow Transplant*. 1997;19(4):331-336.
54. Woodcock A, Forster L, Matthews E, et al; Medical Research Council General Practice Research Framework. Control of exposure to mite allergen and allergen-impermeable bed covers for adults with asthma. *N Engl J Med*. 2003;349(3):225-236.
55. Terreehorst I, Hak E, Oosting AJ, et al. Evaluation of impermeable covers for bedding in patients with allergic rhinitis. *N Engl J Med*. 2003;349(3):237-246.
56. Ad Hoc Working Group on Environmental Allergens and Asthma. Position statement. Environmental allergen avoidance in allergic asthma. *J Allergy Clin Immunol*. 1999;103(2, pt 1):203-205.

57. Bousquet J, Van Cauwenberge P, Khaltaev N; Aria Workshop Group; World Health Organization. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol*. 2001;108(5, suppl): S147-S334.
58. Strupp M, Zingler VC, Arbusow V, et al. Methylprednisolone, valacyclovir, or the combination for vestibular neuritis. *N Engl J Med*. 2004;351(4):354-361.
59. Baloh RW. Clinical practice: vestibular neuritis. *N Engl J Med*. 2003;348(11):1027-1032.
60. Todd MM, Hindman BJ, Clarke WR, Torner JC; Intraoperative Hypothermia for Aneurysm Surgery Trial (IHAST) Investigators. Mild intraoperative hypothermia during surgery for intracranial aneurysm. *N Engl J Med*. 2005;352(2):135-145.
61. Pemberton PL, Dinsmore J. The use of hypothermia as a method of neuroprotection during neurosurgical procedures and after traumatic brain injury: a survey of clinical practice in Great Britain and Ireland. *Anaesthesia*. 2003;58(4):371-373.
62. Gibbons RJ, Abrams J, Chatterjee K, et al; American College of Cardiology; American Heart Association Task Force on practice guidelines (Committee on the Management of Patients With Chronic Stable Angina). ACC/AHA 2002 guideline update for the management of patients with chronic stable angina—summary article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on the Management of Patients With Chronic Stable Angina). *J Am Coll Cardiol*. 2003; 41(1):159-168.
63. Smith SC Jr, Feldman TE, Hirshfeld JW Jr, et al; American College of Cardiology/American Heart Association Task Force on Practice Guidelines; American College of Cardiology/American Heart Association/Society for Cardiovascular Angiography and Interventions Writing Committee to Update the 2001 Guidelines for Percutaneous Coronary Intervention. ACC/AHA/SCAI 2005 Guideline Update for Percutaneous Coronary Intervention—summary article: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update the 2001 Guidelines for Percutaneous Coronary Intervention). *Circulation*. 2006;113(1):156-175.
64. Mastenbroek S, Twisk M, van Echten-Arends J, et al. In vitro fertilization with preimplantation genetic screening. *N Engl J Med*. 2007;357(1):9-17.
65. Wilton L. Preimplantation genetic diagnosis for aneuploidy screening in early human embryos: a review. *Prenat Diagn*. 2002;22(6):512-518.
66. Sermon KD, Michiels A, Harton G, et al. ESHRE PGD Consortium data collection VI: cycles from January to December 2003 with pregnancy follow-up to October 2004. *Hum Reprod*. 2007;22(2):323-336.
67. Verlinsky Y, Cohen J, Munne S, et al. Over a decade of experience with preimplantation genetic diagnosis: a multicenter report. *Fertil Steril*. 2004;82(2):292-294.
68. Action to Control Cardiovascular Risk in Diabetes Study Group; Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559.
69. American Diabetes Association. Standards of medical care in diabetes—2007. *Diabetes Care*. 2007;30(suppl 1):S4-S41.
70. ASTRAL Investigators; Wheatley K, Ives N, Gray R, et al. Revascularization versus medical therapy for renal-artery stenosis. *N Engl J Med*. 2009;361(20):1953-1962.
71. Kalra PA, Guo H, Kausz AT, et al. Atherosclerotic renovascular disease in United States patients aged 67 years or older: risk factors, revascularization, and prognosis. *Kidney Int*. 2005; 68(1):293-301.
72. Bennett-Guerrero E, Pappas TN, Koltun WA, et al; SWIPE 2 Trial Group. Gentamicin-collagen sponge for infection prophylaxis in colorectal surgery. *N Engl J Med*. 2010;363(11):1038-1049.
73. Rutten HJ, Nijhuis PH. Prevention of wound infection in elective colorectal surgery by local application of a gentamicin-containing collagen sponge. *Eur J Surg Suppl*. 1997;578:31-35.
74. Lennmarken C, Bildfors K, Enlund G, Samuelsson P, Sandin R. Victims of awareness. *Acta Anaesthesiol Scand*. 2002;46(3): 229-231.
75. Myles PS, Leslie K, McNeil J, Forbes A, Chan MTV. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. *Lancet*. 2004; 363(9423):1757-1763.
76. Puri GD, Murthy SS. Bispectral index monitoring in patients undergoing cardiac surgery under cardiopulmonary bypass. *Eur J Anaesthesiol*. 2003;20(6):451-456.
77. Lang J. Awakening. *The Atlantic*. January/February 2013.
78. Avidan MS, Zhang L, Burnside BA, et al. Anesthesia awareness and the bispectral index. *N Engl J Med*. 2008;358(11):1097-1108.
79. Hall JB. Use of the pulmonary artery catheter in critically ill patients: was invention the mother of necessity? *JAMA*. 2000; 283(19):2577-2578.
80. Tatsioni A, Bonitsis NG, Ioannidis JP. Persistence of contradicted claims in the literature. *JAMA*. 2007;298(21):2517-2526.
81. What conclusions has Clinical Evidence drawn about what works, what doesn't based on randomised controlled trial evidence? http://clinicalevidence.bmj.com/x/set/static/cms/efficacy-categorisations.html. Accessed June 30, 2011.
82. Prasad V, Cifu A. A medical burden of proof: towards a new ethic. *Biosocieties*. 2012:772-787.
83. Prasad V, Rho J, Cifu A. The diagnosis and treatment of pulmonary embolism: a metaphor for medicine in the evidence-based medicine era. *Arch Intern Med*. 2012;172(12):955-958.
84. Prasad V, Rho J, Cifu A. The inferior vena cava filter: how could a medical device be so well accepted without any evidence of efficacy? *JAMA Intern Med*. 2013;173(7):493-495.
85. Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. *N Engl J Med*. 2013; 368(13):1169-1171.
86. Confavreux C, Suissa S, Saddier P, Bourdès V, Vukusic S; Vaccines in Multiple Sclerosis Study Group. Vaccinations and the risk of relapse in multiple sclerosis. *N Engl J Med*. 2001;344(5):319-326.
87. Ascherio A, Zhang SM, Hernán MA, et al. Hepatitis B vaccination and the risk of multiple sclerosis. *N Engl J Med*. 2001; 344(5):327-332.
88. Fine LG, Barnett EV, Danovitch GM, et al. Systemic lupus erythematosus in pregnancy. *Ann Intern Med*. 1981;94(5):667-677.
89. Sánchez-Guerrero J, Uribe AG, Jiménez-Santana L, et al. A trial of contraceptive methods in women with systemic lupus erythematosus. *N Engl J Med*. 2005;353(24):2539-2549.
90. Petri M, Kim MY, Kalunian KC, et al; OC-SELENA Trial. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med*. 2005;353(24):2550-2558.
91. Goetzl LM; ACOG Committee on Practice Bulletins-Obstetrics. ACOG Practice Bulletin. Clinical Management Guidelines for Obstetrician-Gynecologists Number 36, July 2002. Obstetric analgesia and anesthesia. *Obstet Gynecol*. 2002;100(1):177-191.
92. Wong CA, Scavone BM, Peaceman AM, et al. The risk of cesarean delivery with neuraxial analgesia given early versus late in labor. *N Engl J Med*. 2005;352(7):655-665.
93. Koo HL, Koo DC, Musher DM, DuPont HL. Antimotility agents for the treatment of *Clostridium difficile* diarrhea and colitis. *Clin Infect Dis*. 2009;48(5):598-605.
94. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.