Original Article

A medical burden of proof: Towards a new ethic

Vinay Prasad[a] and Adam Cifu[b,*]

[a]Department of Medicine, Northwestern Memorial Hospital, Northwestern University, 333 E Ontario Suite 901B, Chicago, Illinois 60611, USA.

[b]Department of Medicine, University of Chicago Medical Center, University of Chicago, 5841 S. Maryland Avenue, Chicago, Illinois 60637, USA.

*Corresponding author.

Abstract In current medical practice, innovators, drug and device companies often debut new interventions before they have shown benefit in robust clinical trials. Practitioners readily use these new therapies, in many cases because the practice is financially rewarding and the intervention makes sense within practitioners' scientific worldview. Oftentimes, years after a practice was introduced, the medical community puts it to the test in large, well done randomized trials. Empirical evidence suggests that when this happens, nearly half of those practices are contradicted. We call this phenomenon 'medical reversal'. What are the implications of reversal on our current system of hasty adoption and widespread use of new therapies? Here, we outline the concept of burden of proof in medicine. In the era of evidence-based medicine, who has the burden of proof to show that a therapy works? Currently in clinical practice, innovators and manufacturers are not carrying the burden. Instead, third parties and brave researchers are often required to challenge medical standards years after their introduction. Here, we argue that such a system is untenable. The burden of proof to show that an intervention works must be held by those who develop a new therapy, and by practitioners who profit from the therapy before it is introduced. Here, we promote this as a new physician ethic.

BioSocieties advance online publication, 6 February 2012; doi:10.1057/biosoc.2011.25

Keywords: medical reversal; contradicted findings; evidence-based medicine; burden of proof; COURAGE trial; sham surgery

Introduction

When a doctor or drug company develops a new medical practice, what obligation do they have to show their product works before it is used in a widespread fashion? And, relatedly, for the countless medical practices already widely used, what obligation do practitioners have to ensure the practices are working? Despite a complex set of regulatory institutions, the answer to these questions is, essentially, none. This is troublesome as doctors are often biased towards an intervention. Most medical practices confer financial rewards to administering practitioners. But, what makes current standards most disconcerting is that

many interventions are shown, in retrospect, not to work – thus some practices are implemented in error, harming patients during the years they are in favor.

Although we are in the third decade of 'evidence-based medicine' (Evidence-based Medicine Working Group, 1992) (EBM), we continue to introduce practices (procedures, devices, medications, screening and diagnostic tests) well before we have data to support their use. The drug atenolol debuted in 1976, quickly becoming the first line therapy for hypertension. Early evidence showed that it safely and effectively lowered blood pressure better than placebo. Better blood pressure logically meant longer survival – based upon years of basic science and epidemiologic data. In 2002, a curious finding was made. Losartan, an antihypertensive from another class of drugs, significantly outperformed atenolol for those things that mattered – cardiovascular endpoints and mortality – in a large randomized trial (Dahlöf *et al*, 2002). Even stranger, both drugs had the same effect on 24-h blood pressure (Bang *et al*, 2007). Whether the results were owing to a benefit of losartan or a weak effect of atenolol was debated (Aronow, 2003). In 2004, a meta-analysis to resolve the dispute showed that, despite better blood pressure, treatment with atenolol carried mortality equivalent to placebo (Carlberg *et al*, 2004). Atenolol has subsequently fallen out of favor, and is no longer a first line therapy for hypertension.

The case of atenolol is deeply problematic. Hundreds of thousands of patients were treated with a medication, presumably to lengthen their life, which in retrospect did no such thing. Ideally, such 'reversals' should be rare in the age of EBM, where hard outcomes, death and disability, are examined, rather than silent surrogates, such as hypertension. Unfortunately, reversals abound. In many cases, both the introduction and refutation of a therapy all happened after the emergence of EBM.

In this article, we will explore the ethical obligations of physicians towards new and existing medical practices. We will argue that a novel ethic is required to ensure that the interest of patients remains at the center of patient care. A phrase, borrowed from the legal world, summarizes it: *semper necessitas probandi incumbit ei qui agit*, 'The necessity of proof always lies with he who lays charges'. Although referring to the legal concept of burden of proof, this principle is particularly apt when considering medical interventions. Diagnostic techniques, devices and medications often confer significant financial rewards to the practitioners administering them. The burden of proof, that is, the responsibility to show that the intervention actually works, must lie with those who charge for it. We will argue that should become a fundamental medical ethic.

Our ethic may seem self-evident, but among doctors, it is violated frequently. Often, medical specialties perform lucrative procedures, lacking reliable justification or ongoing attempts at remedy. In this article, we will highlight examples from cardiology, orthopedics and radiation oncology, and argue for an ethic of burden of proof, detailing precisely what must constitute 'proof' in an era of rapidly advancing technology and EBM.

The alternative argument

The primary alternative to this ethic is that of the negative formulation: the burden of proof – to show that an intervention *doesn't* work – must be met by opponents before a practice is discontinued. In the next few paragraphs, we will demonstrate why the alternative view

is mistaken. A positive formulation is a direct consequence of *primum non nocere or* 'first, do no harm', and must be applied to nearly every medical encounter.

Primum non nocere has been called 'the most fundamental principle of medicine', applicable to all situations involving patients (DeAngelis and Fontanarosa, 2010). The principle suggests that doctors should not perform (let alone profit from) services where the net result to the patient is harmful. Neutral interventions (those that neither help nor hurt) marketed under false promises are also prohibited, as they waste the patient's (and society's) finite financial and social resources.

A negative formulation of the burden of proof ethic violates *primum non nocere* not only because many interventions are shown, in retrospect, not to work, but also because of how the medical community handles those refutations. We call the phenomenon reversal, and note that it is ubiquitous in modern medicine.

Reversal

In recent years, we have seen many examples of medical reversal. Among medications, ezetimibe (Taylor *et al*, 2009), atenolol (Carlberg *et al*, 2004) and hormone replacement therapy (Writing Group for the Women's Health Initiative Investigators, 2002). All fell out of favor after trials showed they were either ineffective or harmful. Among surgeries and medical procedures, vertebroplasty was shown to be no better than placebo in 2009, despite years of widespread use. The COURAGE trial showed that for many patients with atherosclerotic disease of the coronary arteries, percutaneous coronary intervention (PCI) and stenting did not improve survival compared with medications alone (Boden *et al*, 2007). At the time of COURAGE, PCI was common in this group. More than half a million stents placed in the United States each year were contradicted by COURAGE (Rosamond *et al*, 2007). Our review of 1 year of high impact literature, found that 13 per cent of trials, concerning a medical practice, were medical reversals (Prasad *et al*, 2011).

It is worth examining one of these cases in detail. Vertebroplasty is a medical procedure where cement is injected into fractured spinal bone, in theory, restoring the original shape, diminishing pain and stabilizing the fragments. Interventional neuroradiologists pioneered its use in the United States in the 1990s (Kolata, 2005), well into the era of EBM. Patients, with spine fractures, who received vertebroplasty had remarkable improvements in pain and disability. By the late 1990s, case series were published, and technical details were shared (Jensen *et al*, 1997). A proper control group was never included in these early trials. Nonetheless, proponents of vertebroplasty lobbied Medicare to fund the procedure, and in 2001, their request was granted (Gray *et al*, 2007). In that year, more than 14 000 vertebroplasties were performed in the United States, and by 2004, that number was 27 000 (Kolata, 2005). Vertebroplasty quickly became a multi-million dollar a year industry (Elshaug and Garber, 2011). In 2009, the procedure was shown to be no better than a sham procedure in two *New England Journal of Medicine* articles (Buchbinder *et al*, 2009; Kallmes *et al*, 2009). In the sham procedure, patients were given conscious sedation, physical and verbal cues that mimicked the procedure, and were allowed to smell the medical cement. Both groups, those who underwent vertebroplasty and sham vertebroplasty, had identical and dramatic responses. Vertebroplasty had become a medical reversal.

When reversal occurs, the discredited practice is seldom immediately abandoned. Tatsioni *et al* (2007) studied the claim that beta-carotene could prevent cancer. By the mid-1990s, three randomized trials debunked the notion. They note that nearly a decade passed before counterarguments were uncommon (Tatsioni *et al*, 2007). Similarly, the use of vertebroplasty continues.

Reversal makes a negative formulation of the burden of proof ethic untenable. First, there is the harm to patients who have already undergone the intervention. Physicians unknowingly committed malfeasance in the care of these individuals (as they received treatment with only risk and no benefit). Second, there are the enduring harms of reversal. The discredited practice is not immediately abandoned, and future physicians err with future patients. Finally, reversal undermines trust in the medical system. The recent reversal of mammography screening guidelines (Woolf, 2010) and the subsequent firestorm hint at the magnitude this harm.

In addition to inertia, there is another way reversal is drawn out. The extent of the incorrect practice is rarely revealed all at once. The manner in which reversal happens is often analogous to 'whittling'. Proper studies typically begin on the fringes, among the extended indications of a practice, and, only incrementally undermine confidence in vast arenas of medicine. The case of COURAGE further highlights this point.

Figure 1 depicts a hypothetical axis (top arrow) of coronary artery disease, and worsening myocardial ischemia. The lower right arrow marks the realm of territory where routine PCI and stents, if needed, have proven mortality benefit. The lower left arrow shows the territory claimed by COURAGE, where routine PCI does not offer benefit. Patients excluded from COURAGE have yet to be critically examined, and future trials may further whittle the PCI paradigm.

The disputed territory between the two lower arrows is comprised by the patients excluded from the COURAGE trial and extends to patients with non-ST segment elevation myocardial infarctions (NSTEMI) with low risk scores (thrombolysis in myocardial

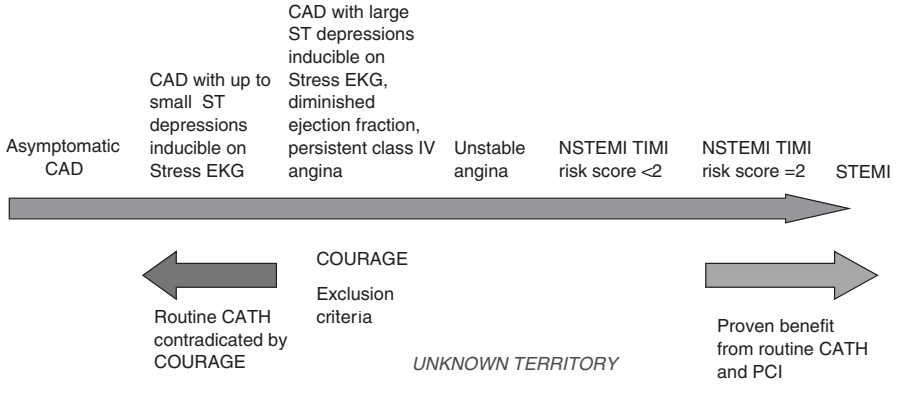

Figure 1: An axis of coronary artery disease, and worsening myocardial ischemia, and the breadth of disputed territory.
Notes: Moving from right to left along the top arrow represents greater degrees of myocardial ischemia. Abbreviations used: CAD – coronary artery disease, CATH – cardiac catheterization, NSTEMI – non-ST segment elevation myocardial infarction, TIMI – Thrombolysis in myocardial infarction, STEMI – ST segment elevation myocardial infarction, PCI – percutaneous coronary intervention.

infarction, 'TIMI' scores). Here PCI remains untested. The unknown territory encompasses a large range of clinical encounters.

The burden of proof comes into question in the unknown territory. When a cardiologist opts to place a stent in the coronary arteries of a patient between the arrows, who has the obligation to show that such stenting is worthwhile? The specialist who stands to reap financial benefit? Or is the burden on some third party to show it doesn't work? In the case of the latter, even if it happens, must 10 years elapse before clinical practice is changed? Only one of these makes sense in light of the considerations we have highlighted.

Who has the burden of proof?

As shorthand, throughout this article, we suggest that the specialist performing a given procedure has an obligation to show its benefit. However, this is a simplification. Certainly, individual practitioners do have some obligation to practice EBM. But the obligation to produce this evidence cannot rest with practitioners alone, who are often neither trialists nor academics. The obligation to test medical practices is shared between innovators (those who pioneer and publicize novel techniques – such as the interventional radiologists who led the vertebroplasty field), regulatory agencies charged with protecting the public interest, and finally specialty organizations (not omitting primary care) that further the narrative of a practice's benefit, often through guideline statements. Despite this diffuse obligation, there remains some individual obligation to provide services that have been validated by the evidence, adopt novel practices cautiously (particularly when a financial bias may be present), and to abandon those practices that have been refuted.

Why reversal exists

The origins of reversal are diverse, and several biases contribute to this phenomenon. The first is our faith in science. Physicians have tremendous confidence in the basic science models, and indeed much of modern medicine is indebted to science. Unfortunately, scientific models are not, and may never be, complete and thus cannot alone be used to justify a therapy.

As we've seen, our confidence in science is often irrationally strong. When a plausible physiologic mechanism is contradicted by a randomized trial, physicians still remain reluctant to abandon the belief (Tatsioni *et al*, 2007). The source of this faith is rooted in a medical curriculum where basic science is taught for the first 2 years, and presented as a foundation for clinical medicine (Prasad, 2010). Such a curriculum may no longer be appropriate in the modern age (Prasad, 2010)

Next, there is an action bias among physicians. Just as goalkeepers routinely dive to one side during penalty kicks – despite evidence that suggests remaining in the center is the strategic position (Bar-Eli *et al*, 2007) – doctors often prefer giving an intervention a try. Patients share in the action bias. A host of psychological factors encourage sick patients to seek unproven remedies (Miller *et al*, 1998).

Unfortunately, the biomedical industry preys upon these biases. The industry often advertises a new therapy by convincing physicians of the strength of the basic science behind it (Elliott, 2010), as with SSRIs like Effexor (Ioannidis, 2008), and the HPV vaccine

Gardasil. In addition, industry-funded clinical trials are more likely to reach encouraging clinical conclusions (Kjaergard and Als-Nielsen, 2002; Baker *et al*, 2003; Cunningham *et al*, 2007) and favorable cost effectiveness analyses than government or university funded ones (Friedberg *et al*, 1999; Bell *et al*, 2006).

At the level of individual practitioners, financial interests continue to bias conduct. Studies suggest that among surgeons, changing payment scheme from capitated to fee-for-service may increase surgery rates up to 78 per cent (Shafrin, 2010). Generous reimbursement of chemotherapy biases oncologists to select more costly drugs, of which they receive a percentage mark up (Jacobson *et al*, 2006). Private urologists are more likely to prescribe unproven androgen deprivation therapy for local prostate cancer than academic urologists, a drug for which Medicare generously reimburses (Shahinian *et al*, 2007). Given the overwhelming evidence that financial conflicts taint practitioners, it seems reasonable to ask those same practitioners for some basic standards of evidence.

A new problem

You may reasonably wonder what circumstances in modern medicine require a new ethic of proof. Historically, proof has not been an issue. Once, all medical interventions were implemented before proper efficacy studies. Until the 1940s, the medical professional lacked what has become one of the key tools of modern outcomes research: the randomized trial (Meldrum, 2000). But, more to topic: the ability of the randomized trial to shake our confidence in pathophysiologic, or mechanistic, approaches to medical problems was not demonstrated until the early 1990s. Thus, for most of the twentieth century, burden of proof was a moot point. Pathophysiologic rationale, or a reasonable, mechanistic story for why something might work, was sufficient to proceed.

In the modern world, however, that paradigm is under fire. Therapies based on a pathophysiologic mechanism sometimes work as intended, but often, they don't. Each time there is a reversal, the argument that a reasonable story is enough to support a new therapy is weakened. Proper studies are required to have confidence that medical interventions offer more than snake oil.

What is required to show 'something "works" '?

Only a proper study can show that a practice works. There are two key parts to a proper study. A proper study must examine the right endpoint, and be performed with the right control. Doctors and patients should be randomized, and unaware whether the intervention was received.

Let's start with endpoints. Many endpoints have demonstrated that they can no longer be trusted. Surrogate endpoints – a biochemical, physiologic or other substitute marker – are inadequate. Both atenolol and losartan have the same effect on blood pressure (Bang *et al*, 2007), but in only one case does that translate to a reduction in mortality (Dahlöf *et al*, 2002). Fibrates and statins both can lower cholesterol, but only one confers mortality benefit (Studer *et al*, 2005).

Composite endpoints have come under question (Tomlinson and Detsky, 2010). In the Synergy between PCI with Taxus and Cardiac Surgery (SYNTAX) trial, which compared PCI with coronary artery bypass surgery (CABG), a composite primary endpoint of death, stroke, myocardial infarction and revascularization was created (Serruys *et al*, 2009). The trial shows composite endpoints at their worst: an amalgam of entities of uneven importance (death and revascularization), where the item of lesser importance (revascularization) drives the overall outcome. The trial then concludes quite misleadingly, 'CABG, as compared with PCI, resulted in lower rates of the combined end point of major adverse cardiac or cerebrovascular events at 1 year'. Included in the article, a plot of death from any cause, stroke or myocardial infarction showed no difference between PCI and CABG. These three outcomes are the only ones that are real, reliable and which matter.

Disease-specific death is also likely to be inadequate. Often cancer prevention trials aim to show the intervention prevents death from that particular cancer. For example, two paired studies regarding prostate cancer aim to show that the prostate specific antigen (PSA) test reduces prostate cancer deaths. One showed a 20 per cent reduction, though overall survival was unchanged (Schröder *et al*, 2009). What accounts for this phenomenon? Possibly, the trial was statistically underpowered, but another study suggests that prostate cancer screening and diagnosis may have implications for cardiovascular events and even suicide (Fall *et al*, 2009). Few would have believed this, and many would overlook this fact by focusing on disease-specific mortality.

Breaking even, that is, balancing lives saved from prostate cancer with deaths incurred through heart attacks, and suicide (at tremendous cost), does not justify a medical practice. Thus, we are left only with endpoints that are important to patients themselves such as mortality and disability. Any study of a new practice should aim for this mark (Figure 2).

	Right Control	Wrong Control	No Control
Right Endpoint	Vertebroplasty in the treatment of Osteoporotic fractures, COURAGE	Orthopedic surgery for pain	IL-2 in the treatment of Metastatic Melanoma
Wrong Endpoint	Early studies of Atenolol vs. placebo as a therapy for hypertension	SYNTAX (DES vs CABG) Composite endpoint, and should have included a control of medical management	
Misunderstood endpoint	The PSA test in the screening of prostate cancer	Radiation therapy in the treatment of breast cancer	Many preclinical studies

Figure 2: Common errors with endpoints and controls.

Let's address the related issue of control. A trial justifying an intervention must have a control arm of 'the closest thing to it'. For surgery, the 'closest thing to it' is a sham procedure that continues up until the step of interest. Nowhere is this more important than with subjective outcomes like pain relief. Over 60 years, we have witnessed notable reversals where a sham-controlled trial demonstrated that a commonly performed procedure had no real benefits. Famous examples include: sham ligation of the internal mammary artery for angina (Cobb *et al*, 1959), arthroscopic surgery for knee osteoarthritis (Moseley *et al*, 2002), and recently, vertebroplasty for osteoporotic fractures (Buchbinder *et al*, 2009; Kallmes *et al*, 2009). In all three of these cases, preliminary (non-controlled or improperly controlled) studies were highly promising, and it required a sham study to illuminate the true null effect. In medicine, no field is dominated by as many hitherto untested procedures as orthopedic surgery.

Orthopedics and sham surgery

The cost of replacing a knee or hip is between US$30 000 and $40 000. In 2007, the total amount spent in America for hip and knee replacements were $19 billion and $26 billion, respectively (Alderman, 2010). Spinal surgery for lower back pain is equally costly, and rapidly growing. Orthopedics as a specialty remains among the most lucrative, largely because of high billing procedures like joint replacement and back surgery.

The majority of such surgeries are performed because of the debilitating pain of osteoarthritis. Specific indications for performing surgery are vague, though they are generally reserved for patients who have failed other forms of medical and physical therapies. In the case of back pain, even when surgery has been decided upon, consensus for the particular operation required is wholly lacking (Deyo *et al*, 2010). Individual surgeons have wide discretion.

As we've seen, caution is warranted when pain is the outcome of interest. The placebo response can be overwhelming (Turner *et al*, 1994). And yet, for knee and hip replacements, there are absolutely no proper studies, or ongoing attempts at remedy. Moreover, there is little financial incentive for orthopedists to question their practice, asking the question: do these surgeries improve pain beyond simply the belief that they should? Do they work better than a sham procedure? Billions of dollars of revenue, and vast arenas of scientific belief are in jeopardy if they don't.

The ethics of clinical trials is tied to this criticism, and is currently at a crossroad. Debate has centered on the question: when are randomized controlled trials appropriate? Discussants have the task of balancing the interests of trial participants with society's interest in generating knowledge for drug and device approval and coverage decisions. One high profile publication (Miller and Joffe, 2011) has argued that conventional ethical concepts (such as equipoise), which have regulated trials for decades, may no longer be relevant to deal with the complexities of modern medicine. Here, we advance a similar position, arguing in favor of sham studies. Ethical opposition to sham studies is typically shortsighted, and confined to the risks and benefits of the participants of the trial. The concept of reversal again provides clarity.

In his critique of sham surgery of the spine, Peter Angelos, an endocrine surgeon and ethicist, highlights the problems (Angelos, 2007): a sham procedure would deprive patients

the chance for non-operative management (physical therapy and exercise counseling), and is therefore unethical; the risks of a sham procedure include general anesthesia, postoperative infection and pain, which are great and therefore unethical: there is no way to achieve adequate blinding (which is not so much an ethical objection, as a practical one).

What Angelos doesn't note is that *performing* the surgery deprives patients of non-operative management. Performing the surgery includes all the risk of incision, general anesthesia and beyond. The risk of dural perforation and CSF leak, one of the most morbid complications, is present only in the surgery group, where the lamina is breached. The flaw underlying Angelos's arguments is a preconceived bias that the intervention likely works. If we lacked that bias, we might say it is unethical to do the surgery.

Other authors have sided with Angelos regarding the ethics of sham surgery. Ruth Macklin (1999) provided a famous critique in the *New England Journal of Medicine*. Her article extensively detailed the risks involved to the participants in a 1999 trial of Parkinson's therapy; however, she ignored the risks to patients outside of the trial. What if a practice gains widespread use, lacking evidence it works? The concept of reversal changes the way we must think about sham surgery. Figure 3 takes a pragmatic, utilitarian view.

Regarding the ethics of sham surgery in clinical research, there are two possibilities. We can perform a sham-controlled RCT or we can skip it (the first branch point in Figure 3). In addition there are two empirical truths: either the surgery works or it doesn't (the second branch). The concept of reversal suggests there is a massive harm that these ethicists have ignored. If we don't perform the study, and the treatment does not work, every patient who later undergoes the surgery is harmed (far right column of Figure 3). In the case of vertebroplasty, 63 and 40 control patients were subjected to the risks of sedation and sham treatment in the two *New England Journal of Medicine* studies. In the preceding decade, at

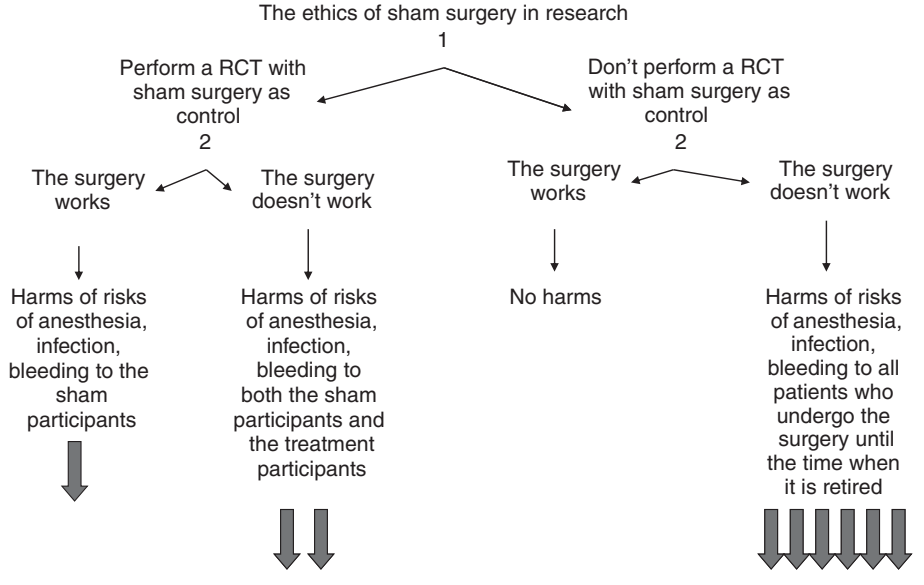

Figure 3: A pragmatic view of the ethics of sham research.

least 100 times more people underwent the procedure each year, paid for by Medicare alone (Gray *et al*, 2007). These patients were subjected to the same risks of sedation, and the final risk of the injection of medical cement, for which we no longer have long-term data. If, as Macklin (1999) states, the foremost ethical principle is that, 'the risk of harm to subjects must be minimized in the conduct of research', then it is unethical not to conduct the trial; in this case by about a hundred fold per annum.

In short, we are arguing beyond the ethics of sham surgery. It is unethical not to do sham procedures, and continue to bill for lucrative interventions, if we are unsure they work. In some cases (not all) we will undergo reversal, whose harms are far greater, and delivered to many more than the risks of anesthesia or incision in any clinical trial. Orthopedists must be held to an ethic of burden of proof.

The case within radiation oncology

The cost of radiation therapy (RT) in the treatment of breast cancer is routinely in the tens of thousands of dollars per patient (Smith, 2004). Proponents of RT emphasize that radiation is cheaper than other modalities of treatment, such as surgery and drugs (Lievens and Van den Bogaert, 2005). Because of the prevalence of breast cancer, whether or not RT is routinely used has large financial implications.

The burden of proof principle we are advocating would suggest that radiation oncologists have an ethical obligation to show that RT works. Works – defined as improving the overall survival of women with breast cancer, or improving their quality of life without diminishing survival. Unfortunately, they have failed to do this. RT is an example of a misunderstood endpoint.

Two of the largest randomized trials on RT with respect to breast cancer were sponsored by the National Surgical Adjuvant Breast and Bowel Project (NSABP). NSABP B-04 and NSABP B-06 showed similar overall survival rates among patients who received RT as those who did not, despite the fact that there was substantial improvement in local control with RT. Both trials included over a thousand participants, and were well done. Both serve as examples of a surrogate (local control) endpoint failing to properly correlate with the outcome of interest (dying from breast cancer).

Many other randomized trials have failed to show mortality benefit with the inclusion of RT. To our knowledge only one randomized trial in the breast cancer literature claims overall mortality benefit from RT. That trial is by Overgaard *et al* (1997), and in it patients received sub-par surgical resection of lymph nodes, inadequately dosed chemotherapy, and had cancer recurrence rates and deaths far greater than comparable randomized trials. Moreover the trial looked at a particular group of women (pre-menopausal, with high risk cancers), and thus the generalizability of the findings – even if we get past the formidable criticism – is suspect.

In addition, one well-done meta-analysis argues against any overall mortality benefit from RT, suggesting that deaths saved from breast cancer may be lost through increased non-cancer death, including heart disease and lung cancer (Early Breast Cancer Trialists' Collaborative Group, 2000). Proponents of RT, often cite a different meta-analysis, however. This one examined whether local failure is a predictor of survival, and lumped together

trials of RT with those of more aggressive surgery. Unfortunately, such a design inherently invites confounding results, as it considers two very different things together – surgery and radiation (Clarke *et al*, 2005).

If practitioners of RT were forthcoming, they would inform patients that RT represents a trade off. On the one hand, you are more likely to accelerate coronary artery disease, or die of lung cancer, but on the other hand, you are less likely to have a relapse of your breast cancer. Patients then could make the individualized decision whether to undergo, or to omit radiation. Unfortunately, too often, such decisions are couched under different terms. RT represents 'doing everything possible' to kill cancer. Such a statement makes tacit claims about overall mortality, which the data do not support.

There is an additional worry here. Meta-analysis probably does not fall under our heading of a proper study. It is likely insufficient to show that a treatment works. There have been prodigious discrepancies between meta-analysis and subsequent large RCTs. Examples include the role of magnesium after myocardial infarction, which seemed promising in meta-analysis but was overturned by RCT and angiotensin-receptor blockers to prevent atrial fibrillation, which experienced the same fate (Hennekens and DeMets, 2009). One systematic comparison suggests that meta-analysis was contradicted by subsequent RCT about 35 per cent of the time (LeLorier *et al*, 1997).

Thus, despite having failed to show that RT works with a single proper study, or to uniformly counsel patients on the true risks and benefits, radiation oncologists continue to reap significant profits from the therapy. Billions of dollars are at stake when it comes to radiation treatment of common cancers. Radiation oncologists must subscribe to a basic ethic of burden of proof in order to promote the best interests of patients.

Where the burden of proof can be suspended

There are special cases where the ethic of burden of proof may be suspended. These occur in clinical situations that are unique, acute, dire or of overwhelming benefit. In these cases it is reasonable to continue to practice standard of care until evidence becomes available. We will use a real clinical encounter to expand on each of these.

In cases that are unique, practitioners are rightly exempt from burden of proof. There can be no prospective study when there is no one to study. For example, *a 20-year-old man with late stage AIDS (his CD4 count is less than 50) presents with a headache. Evaluation reveals histoplasmosis antigen and antibodies in his urine and CSF and blastomycosis antigen in his urine. An extensive literature search reveals only case reports of patients with similar presentations. The patient is treated with liposomal amphotericin B, itraconazole and HAART therapy.*

In situations that are acute (such as trauma), it is reasonable to suspend the ethic of burden of proof, until testable questions can be defined. For example *a 50-year-old woman presents after a motor vehicle accident. She is increasingly unresponsive, and imaging of the brain suggests imminent herniation. A neurosurgeon recommends intubation, hyperventilation, manitol, barbiturates and hemicranectomy.* Until it becomes feasible to perform proper studies, hemicranectomy would be exempt from a burden of proof ethic.

In cases that are dire, temporizing measures (even guesses) are reasonable. As an example, *a 77-year-old with metastatic esophageal cancer presents with marantic endocarditis with CNS emboli and neurologic impairment. How should he be treated?* Of course, if a clinical situation demonstrates itself as increasingly more common, that is, if we find ourselves encountering dozens of patients in a similar situation, then the ethic of burden of proof must be reinstated.

In some cases, the burden of proof may be suspended because a novel therapy revolutionizes the natural history of an illness. One notable example is the case of *imatinib in the treatment of CML. With imatinib, initial data was so promising that a trial was constructed that allowed crossover* (Guilhot *et al*, 2009). *As of the last update, 65 per cent of patients assigned to previous therapy had switched to imatinib for a multitude of reasons, and though it is certain imatinib improves overall survival, the trial cannot be said to have proven this because too many patients withdrew from the control arm.*

In terms of raw numbers, the exceptions to the burden of proof principle are very much that, exceptional cases. Where the burden of proof ethic must be enforced is for a wide range of encounters: situations that are common, costly, concern a subjective outcome, and particularly among patients who otherwise feel well. In other words, the vast majority of physicians have an obligation to show their therapy(ies) work(s).

The critics of EBM

Despite the longevity and dominance of the evidence-based movement, there remain critics. One famous satire notes the lack of randomized trials showing the effectiveness of parachutes (Smith and Pell, 2003). Our analysis of the burden of proof ethic is not entirely incompatible with critics of randomized trials, however. John Worrall, a philosopher of science, has argued that in certain cases the magnitude of benefit is so overwhelming, there is no need to randomize. The example Worrall favors is extracorporeal membranous oxygenation (ECMO) for neonates with persistent pulmonary hypertension (Worrall, 2007). ECMO transformed mortality in these patients from 80 per cent to 20 per cent, nearly overnight, according to Worrall, making a future RCT seem superfluous. Our position does not disagree with this view, however. We agree that in rare cases, a novel treatment may change the natural history of a disease, and randomized trials may not be feasible. The examples we favor are imatinib and some types of organ transplantation. However, we maintain that these are indeed exceptional cases. The vast majority of medical interventions likely do require evidence, as their benefit is not visible to the naked eye.

Ironically, another treatment Worrall offers as never been tested by randomized trial, but which 'no one seriously doubts', (Worrall, 2007) is appendectomy for acute appendicitis. In fact, appendectomy was recently subjected to a randomized trial against antibiotics for the treatment of acute appendicitis. While this study (Vons *et al*, 2011) did not show that antibiotics were as good as surgery (antibiotics were not non-inferior, strictly speaking), the study strongly suggests that there is a subset of patients with uncomplicated appendicitis that do not need surgery, and would be fine with antibiotics alone. Future trials are needed

to tease out this subgroup. There is a broad theoretical point here. Critics of EBM have a difficult task of showing where it does not apply. Each passing year has extended the scope of the discipline, forging bold directions in research, and even the most sacrosanct of medical practices can be questioned. Omitting a medical practice from rigorous study must continue to be the exception, and not the rule.

A new ethic

As with law, a medical burden of proof makes the assumption that encounters are, to some degree, of adversarial nature. Admitting this is the case in medicine is profoundly difficult. While we wish physicians always acted in the best interests of patients alone, there is evidence that this is not the case. Likely this is done, not because of malice or greed, but because our decision-making is flawed. Our enthusiasm for therapies is not rationally grounded, and these false beliefs are shared among patients and doctors. It is likely very hard for individual practitioners to impartially weigh the evidence for an intervention when their own finances are strongly tied to that decision. There is an additional concern with respect to academic physicians. As patterns of promotion and tenure are tied to publication records, inferior quality, though easy to perform studies, are preferentially done. Retrospective, sub-group analyses, mechanistic science, systematic reviews, meta-analyses and cost-effective studies are 'easy' papers compared to prospective, randomized, blinded clinical trials that offer a stronger truth claim. Thus, the nature of academic promotion may also be adversarial with the best interests of patients. Our insistence on 'proper studies' is a way around this. In order for medicine to continue to abide by the highest ethic, *primum non nocere*, now more than ever, we need its corollary: the ethic of burden of proof.

About the Authors

Vinay Prasad is a physician at Northwestern University and the Feinberg School of Medicine. He is a graduate of the University of Chicago, and Michigan State University. His research and writings have appeared in prominent medical journals including the *Journal of the American Medical Association* and the *Archives of Internal Medicine*. Together with Adam Cifu, he coined the term 'medical reversal'. This subject has received much press, and is featured in *Consumer Reports*. He has written on a variety of subjects, including professionalism, medical education, biomarkers, cardiac risk, humanism in medicine, language and medical reversal. His foremost interest is in health policy, and the adoption of medical technology.

Adam Cifu is an Associate Professor in General Internal Medicine at the University of Chicago, Pritzker School of Medicine where he is the director of the internal medicine clerkship and a Master in the Academy of Distinguished Medical Educators. His research interests include medical education, evidence based medicine and the adoption of medical technology. Together with Vinay Prasad, he coined the term 'medical reversal'. He is currently working on a mobile application that would assist students with medical

decision making. He is one of the authors of *Symptom to Diagnosis: An Evidence Based Guide*, a textbook that teaches clinical reasoning to medical students.

References

Alderman, L. (2010) Caring for hips and knees to avoid artificial joints. *The New York Times*, 23 April.

Angelos, P. (2007) Sham surgery in clinical trials. *Journal of the American Medical Association* 297(14): 1545–1546.

Aronow, W.S. (2003) Might losartan reduce sudden cardiac death in diabetic patients with hypertension? *The Lancet* 362(9384): 591–592.

Baker, C.B., Johnsrud, M.T., Crismon, M.L., Rosenheck, R.A. and Woods, S.W. (2003) Quantitative analysis of sponsorship bias in economic studies of antidepressants. *British Journal of Psychiatry* 183(6): 498–506.

Bang, L.E. *et al* (2007) Losartan versus atenolol on 24-hour ambulatory blood pressure: A LIFE substudy. *Blood Pressure* 16(6): 392–397.

Bar-Eli, M., Azar, O.H., Ritov, I., Keidar-Levin, Y. and Schein, G. (2007) Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of Economic Psychology* 28(5): 606–621.

Bell, C.M. *et al* (2006) Bias in published cost effectiveness studies: Systematic review. *British Medical Journal* 332(7543): 699–703.

Boden, W.E. *et al* (2007) COURAGE Trial Research Group: Optimal medical therapy with or without PCI for stable coronary disease. *New England Journal of Medicine* 356(15): 1503–1516.

Buchbinder, R. *et al* (2009) A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *New England Journal of Medicine* 361(6): 557–568.

Carlberg, B., Samuelsson, O. and Lindholm, L. (2004) Atenolol in hypertension: Is it a wise choice? *The Lancet* 364(9446): 1684–1689.

Clarke, M. *et al* (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. *The Lancet* 366(9503): 2087–2106.

Cobb, L.A., Thomas, G.I., Dillard, D.H., Merendino, K.A. and Bruce, R.A. (1959) An evaluation of internal-mammary-artery ligation by a double-blind technic. *New England Journal of Medicine* 260(22): 1115–1118.

Cunningham, M.R. *et al* (2007) Industry-funded positive studies not associated with better design or larger size. *Clinical Orthopaedics and Related Research* 457(51): 235–241.

Dahlöf, B., Devereux, R.B. and Kjeldsen, S.E. *et al* (2002) Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (LIFE): A randomised trial against atenolol. *The Lancet* 359(9311): 995–1003.

DeAngelis, C.D. and Fontanarosa, P.B. (2010) Ensuring integrity in industry-sponsored research primum non nocere, revisited. *Journal of the American Medical Association* 303(12): 1196–1198.

Deyo, R.A. *et al* (2010) Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *Journal of the American Medical Association* 303(13): 1259–1265.

Early Breast Cancer Trialists' Collaborative Group. (2000) Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: An overview of the randomised trials. *The Lancet* 355(9217): 1757–1770.

Elliott, C. (2010) *White Coat Black Hat: Adventures on the Dark Side of Medicine*. Boston, MA: Beacon Press.

Elshaug, A.G. and Garber, A.M. (2011) How CER could pay for itself: Insights from vertebral fracture treatments. *New England Journal of Medicine* 364(15): 1390–1393.

Evidence-based Medicine Working Group. (1992) Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association* 268(17): 2420–2425.

Fall, K. *et al* (2009) Immediate risk for cardiovascular events and suicide following a prostate cancer diagnosis: Prospective cohort study. *PLoS Medicine* 6(12): e1000197.

Friedberg, M., Saffran, B., Stinson, T.J., Nelson, W. and Bennett, C.L. (1999) Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *Journal of the American Medical Association* 282(15): 1453–1457.

Gray, D.T., Hollingworth, W., Onwudiwe, N., Deyo, R.A. and Jarvik, J.G. (2007) Thoracic and lumbar vertebroplasties performed in US Medicare enrollees, 2001–2005. *Journal of the American Medical Association* 298(15): 1760–1762.

Guilhot, F. *et al* (2009) High rates of durable response are achieved with imatinib after treatment with interferon α plus cytarabine: Results from the International Randomized Study of Interferon and STI571 (IRIS) trial. *Haematologica* 94(12): 1669–1675.

Hennekens, C.H. and DeMets, D. (2009) The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. *Journal of the American Medical Association* 302(21): 2361–2362.

Ioannidis, J.P.A. (2008) Effectiveness of antidepressants: An evidence myth constructed from a thousand randomized trials? *Philosophy, Ethics, and Humanities in Medicine* 3(1): 14.

Jacobson, M., O'Malley, A.J., Earle, C.C., Pakes, J., Gaccione, P. and Newhouse, J.P. (2006) Does reimbursement influence chemotherapy treatment for cancer patients? *Health Affairs* 25(2): 437–444.

Jensen, M.E., Evans, A.J., Mathis, J.M., Kallmes, D.F., Cloft, H.J. and Dion, J.E. (1997) Percutaneous polymethylmethacrylate vertebroplasty in the treatment of osteoporotic vertebral body compression fractures: Technical aspects. *American Journal of Neuroradiology* 18(10): 1897–1904.

Kallmes, D.F. *et al* (2009) A randomized trial of vertebroplasty for osteoporotic spinal fractures. *New England Journal of Medicine* 361(6): 569–579.

Kjaergard, L.L. and Als-Nielsen, B. (2002) Association between competing interests and authors' conclusions: Epidemiological study of randomised clinical trials published in the BMJ. *British Medical Journal* 325(7358): 249–253.

Kolata, G. (2005) Spinal cement draws patients and questions. *The New York Times*, 28 August.

LeLorier, J., Grégoire, G., Benhaddad, A., Lapierre, J. and Derderian, F.N. (1997) Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *New England Journal of Medicine* 337(8): 536–542.

Lievens, Y. and Van den Bogaert, W. (2005) Proton beam therapy: Too expensive to become true? *Radiotherapy and Oncology* 75(2): 131–133.

Macklin, R. (1999) The ethical problems with sham surgery in clinical research. *New England Journal of Medicine* 341(13): 992–996.

Meldrum, M. (2000) A brief history of the randomized controlled trial from oranges and lemons to the gold standard. *Hematology/Oncology Clinics of North America* 14(4): 745–760.

Miller, F.G. and Joffe, S. (2011) Equipoise and the dilemma of randomized clinical trials. *New England Journal of Medicine* 364(5): 476–480.

Miller, M., Boyer, M.J., Butow, P.N., Gattellari, M., Dunn, S.M. and Childs, A. (1998) The use of unproven methods of treatment by cancer patients: Frequency, expectations and cost. *Supportive Care in Cancer* 6(4): 337–347.

Moseley, J.B. *et al* (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine* 347(2): 81–88.

Overgaard, M. *et al* (1997) Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. *New England Journal of Medicine* 337(14): 949–955.

Prasad, V. (2010) Beyond storytelling in medicine: An encounter based curriculum. *Academic Medicine* 85(5): 794–798.

Prasad, V., Gall, V. and Cifu, A. (2011) The frequency of medical reversal. *Archives of Internal Medicine* 171(18): 1675–1676.

Rosamond, W. *et al* (2007) Heart disease and stroke statistics, 2007 update: A report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 115(5): e69–e171.

Schröder, F.H. *et al* (2009) Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine* 360(13): 1320–1328.

Serruys, P.W. *et al* (2009) Percutaneous coronary intervention versus coronary-artery bypass grafting for severe coronary artery disease. *New England Journal of Medicine* 360(10): 961–972.

Shafrin, J. (2010) Operating on commission: Analyzing how physician financial incentives affect surgery rates. *Health Economics* 19(5): 562–580.

Shahinian, V.B., Kuo, Y.F., Freeman, J.L. and Goodwin, J.S. (2007) Characteristics of urologists predict the use of androgen deprivation therapy for prostate cancer. *Journal of Clinical Oncology* 25(34): 5359–5365.

Smith, G.C.S. and Pell, J.P. (2003) Parachute use to prevent death and major trauma related to gravitational challenge: Systematic review of randomised controlled trials. *British Journal of Medicine* 327(7429): 1459–1461.

Smith, S. (2004) Older breast cancer patients are urged to avoid radiation. *Boston Globe*, 2 September.

Studer, M., Briel, M., Leimenstoll, B., Glass, T.R. and Bucher, H.C. (2005) Effect of different antilipidemic agents and diets on mortality: A systematic review. *Archive of Internal Medicine* 165(7): 725–730.

Tatsioni, A., Bonitsis, N.G. and Ioannidis, J.P.A. (2007) Persistence of contradicted claims in the literature. *Journal of the American Medical Association* 298(21): 2517–2526.

Taylor, A.J. *et al* (2009) Extended-release niacin or ezetimibe and carotid intima-media thickness. *New England Journal of Medicine* 361(22): 2113–2122.

Tomlinson, G. and Detsky, A.S. (2010) Composite end points in randomized trials: There is no free lunch. *Journal of the American Medical Association* 303(3): 267–268.

Turner, J.A., Deyo, R.A., Loeser, J.D., Von Korff, M. and Fordyce, W.E. (1994) The importance of placebo effects in pain treatment and research. *Journal of the American Medical Association* 271(20): 1609–1614.

Vons, C. *et al* (2011) Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: An open-label, non-inferiority, randomised controlled trial. *The Lancet* 377(9777): 1573–1579.

Writing Group for the Women's Health Initiative Investigators. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *Journal of the American Medical Association* 288(3): 321–333.

Woolf, S.H. (2010) The 2009 Breast Cancer Screening Recommendations of the US Preventive Services Task Force. *Journal of the American Medical Association* 303(2): 162–163.

Worrall, J. (2007) Why there's no cause to randomize. *British Journal for the Philosophy of Science* 58(3): 451–488.