1

Baseline Observations

Mark Pauly

INTRODUCTION

Over the last 40 years, spending on both hospital and physician services in the United States has inexorably increased, often faster than gross domestic product (GDP) or any other aggregate measure. In contrast to industries such as computer software, hospitality, sports and recreation – where spending has also grown faster than the economy – health care spending growth is not thought to be matched by increased customer or patient satisfaction or improved outcomes. For some groups, especially those that are socially disadvantaged or lower income, measures of health have remained stubbornly lower relative to the rest of the population. Despite continuous criticism of the status quo and calls for transformation, little has changed. Why has this sector of the economy uniquely resisted changes in products, productivity, and services aimed at improving consumer satisfaction or reducing spending growth?

It is not from a lack of discussion about the need for change, or a shortage of proposals with promising ideas for change. Armies of innovation promoters, often speaking a special language of their own that is foreign to the language of health care, have descended upon the industry. They bring shiny new technologies and services that lack evidence, resulting in little improvement and much misdirection. While these promoters have succeeded in raising the alarm that major changes are coming, so far the transformation they promise has not worked to produce detectable improvements.

The claim and substance of this book is, fundamentally:

Health care in the United States needs help, and evidence can provide that help.

Some of the problems in health care are the result of investments in unsubstantiated – and often costly – innovations. At the same time this practice is rampant, investment in and implementation of evidence-based, effective innovations in the delivery of hospital and medical services appears to have lagged. This situation must change, and this book is a call to action to provide the evidence needed to support that change. Where there is underuse and misuse of evidence, our goal is to find out why that happens and propose ways to change things.

THE PROBLEM

In contrast to our past, and to the rest of the world, US health outcomes are now more resistant to improvement despite health care spending outpacing all measures of national income growth. This unchanging pattern is especially unsettling in light of the striking advancements and discoveries in human biology and clinical medicine. While it is true that in recent decades we have seen cures for diseases that were previously considered incurable, the combination of social and environmental problems that have worsened health and our inability to improve the efficiency of care delivery is distressing. In health care, in both the economic and public health spheres, things were not going well even before the novel coronavirus appeared, and there is little reason to believe that they will improve even after the pandemic is controlled. Attempts to improve the fundamentals of equitable and efficient care delivery through evidence-based care have faltered. There is reason to believe that effective, evidence-based interventions to lower spending growth and improve outcomes exist – and these should guide health care management strategies that, in turn, should promote strategies that prevent or minimize known potential negative effects on equity or cost. At a minimum, we need to see what interventions can do more harm than good. We also need to see if the problem of absence of evidence for the effectiveness of innovations is getting worse. Particularly regarding innovations in health care and insurance management, has the use of evidence eroded in favor of fad, fashion, or a misdirected need to do something, even if it is unproven? This book will provide the answers to such questions.

Thankfully, the management story is not all gloom and doom. There have been bright spots of innovations in care delivery that lower cost and improve outcomes, or at least accomplish one of those tasks without affecting the other. There is considerable interest in implementing programs that generate such innovations. Not only can we make better use of

new scientific discoveries with better management, but we can also discover innovative ways to manage the technology we currently have that go beyond the buzzwords of "leadership" and "cultural change" to produce improvements.

However, frustrations still persist. Many of the demonstrated improvements have languished without adoption and dissemination. Instead of investing in programs with strong evidence of effectiveness, managers instead opt for unproven programs backed only by plausible stories, often because they require less investment or are easier to deliver, but sometimes even when they have substantial, known upfront costs. Through both trying to spend less than what is known to be needed and spending more on efforts not known to be effective, resources are wasted. Moreover, the benefits of past programs that worked when piloted on a smaller scale do not yet add up to noteworthy changes in aggregate measures of spending growth and quality improvement. Too often, they get overtaken by the tide of more spending with less improvement, and so poor overall performance rolls on.

THIS BOOK

Our goal with this book is to contribute to understanding and to potentially improve the use of evidence in management decisions about innovative interventions. Before we offer remedies, we first tackle the task of describing the use or nonuse of evidence in decision-making around innovation in health care management by everyone from top-level systems managers, insurers, line managers, and employers buying insurance on behalf of their workers to physicians and other health professionals acting collectively or individually. We then attempt to explain the current trends in evidence use – when evidence is sought, how it is responded to when it appears, and what happens if it is impossible to obtain. We describe how decisions come about with and without evidence and evaluate whether evidence is currently generated and used in the most efficient ways. We take a neutral position, noting that the usefulness of evidence is context dependent, and that at times it does not pay to incur the cost to obtain or consume evidence (though accurate information that falls in your lap by chance is often of value, too). Finally, we address the challenge of deciding whether and when better evidence should be sought and how that evidence should be used. Here again we are neutral and acknowledge that there are multiple ways to generate more useful evidence for managers and methods to ensure it is used properly. As noted above, we pay particular attention to

evidence (or its absence) for innovations in health care spanning clinical, managerial, and financial decisions.

We offer as much evidence as we can muster on the big strategy questions: How much gain is possible with better use of evidence in the most important places? Are there areas where gains have already been exhausted? Is the cost-quality relationship in health care as good as it is going to get? We try to be optimistic and offer some innovative ideas that come from our review, along with sage advice about what to do when the only thing you know is that no one knows.

RUNNING DOWN THE COMPETITION

Are there alternatives to the development and use of novel evidence that could lower spending growth and/or improve health care outcomes? A common answer is to change leadership or culture. There is evidence, mostly from the education setting where outcomes are measured by student test scores, that variation in school performance can be attributed to differences in their leadership and/or culture – the principal, the administration, and the morale among the teaching staff.

Indeed, there are similar opinions even in the literature on evidence-based health care that the most important key to progress and improvement is getting the right people to implement the best evidence-based new programs, though that is not always clear.[1] However, while we acknowledge the role of leadership and culture, we plan to resist the temptation to go for these nostrums for two reasons: inability to measure leadership/culture, and a shortage of good leaders. Leadership and culture are often hard to measure objectively, and so are inferred from observations after the fact. Entities with the best performance are almost always identified retrospectively, and the leaders who presided over their performance (and therefore must have led it) are named as heroes. They are given credit for wise leadership and for building a relationship with and among their workers that is happy, satisfying, and productive. A prospective study identifying organizations with good leadership beforehand and then seeing which kinds do better is rare.

The problem is that the production process for leadership and culture is not known and surely constrained. Many people talk about how to develop someone into a leader of the workers who will implement an organization's

[1] Kovner, A. R., and D'Aunno, T. A. (2017). *Evidence-Based Management in Healthcare: Principles, Cases, and Perspectives*. Chicago: Health Administration Press.

culture, and yet we are always confronted by a shortage of good leaders. Further, changing culture at all levels in an organization is hard. Often there are limited numbers of potential workers who will fit an organization's culture, and an even smaller number of gifted leaders – it is hard to mold people to an organizational culture when they have divergent or different cultural preconditioning.

The upshot is that this book will focus more on health interventions with documented effectiveness for improving outcomes and the implementation strategies that were necessary for the adoption of evidence. We will talk about the role of leadership and culture in the context of evidence-based programs that change leadership and culture for the better. Prospective efforts to produce great leaders or otherwise alter things for the better are on the same footing with organizational changes in systems or behavioral changes in insurance buyers, patients, and physicians – all of which are also hard to change. These efforts are worth a look – but a skeptical one. What is crucial, however, is sound decision-making by leaders based on evidence (whenever possible) and clear concepts of costs and benefits (and knowledge of who will bear the costs and receive the benefits). Recognizing that there will always be "unknown unknowns," applying the existing evidence base can reduce uncertainty and de-risk new concepts, but judgment will always be needed. In addition, leaders in innovative implementation must recognize the required culture change that starts with clear bidirectional communication that will facilitate buy-in from frontline staff who will be responsible for change, as well as the population of patients who will be affected by it.

AN EXAMPLE OF CHALLENGES IN DETERMINING THE BEST USE OF INFORMATION: MEDICARE'S HOSPITAL READMISSION REDUCTION PROGRAM

In 2012, the Obama Administration and Centers for Medicare and Medicaid Services (CMS) wanted to show that it was possible to simultaneously implement health reform, reduce spending, and improve quality. Efforts to achieve that goal had been strongly encouraged by the Medicare Payment Advisory Committee (MedPAC) to CMS, which proposed reducing hospital readmissions as a way to achieve all three. Both CMS and MedPAC lobbied for the Medicare Hospital Readmission Reduction Program (HRRP) in the Affordable Care Act to financially penalize hospitals with higher than expected readmission rates. It was generally agreed that a high rate of readmissions adjusted for risk was a sign of poor quality

in a hospital, other things being equal. Readmission meant that the hospital's first effort to cure or stabilize a patient failed, and another costly attempt was necessary. The conclusion that fewer readmissions would save money was thought to be obvious. The conclusion that fewer readmissions would result in improved health outcomes followed from the definition of quality of care. There was evidence that other extant programs successfully achieved such outcomes; there was no evidence that financial penalties would cause all hospitals to choose to implement those programs.

Many hospitals did have readmission rates above what would have been expected based on their patient risk (though perhaps not based on the social vulnerability of those patients). However, there was little evidence that the hospitals affected by penalties would choose the best or least harmful methods if they did try to respond. Still, Congress proceeded with HRRP legislation, without waiting for evidence of effectiveness or safety from the policy it contained.[2]

The intervention relied on using financial penalties to offset what were viewed as distorted incentives to hospitals. After all, under Medicare's diagnosis-related group (DRG) based payment system, a hospital readmission meant more revenue and potentially more profit if the payment was above the hospital's marginal or opportunity cost. The relatively modest penalty at least reduced the financial reward even if it did not necessarily turn it into a loss. Why some hospitals responded to these incentives while others did not was also unknown. To make hospitals comparable, each admission was adjusted by Medicare's risk score (for higher risk and more costly patients within a DRG).

In all, there were some valid reasons why this policy seemed like a good idea at the time, even if not all analysts were in favor then and many have subsequently become more critical. Given what was known at the time, the policy was not unreasonable on its face – but it was risky. The need to make a good showing on what had been labeled the "Triple Aim"[3] (health care cost, quality, and access) – and the logic behind the financial and clinical arguments – won, and the program was implemented nationwide without a control group.

The initial results following implementation of this new program in October 2012 were striking. Almost immediately, the risk-adjusted

[2] James, J. (2013). Medicare hospital readmissions reduction program. *Health Affairs*, 34(2), 1–5.

[3] Berwick, D. M., Nolan, T. W., and Whittington, J. (2008) The triple aim: Care, health, and cost. *Health Affairs*, 27(3)(May/June), 759–769.

readmission rate fell, and eventually fell by 22 percent.[4] In fact, the rate had begun to fall for several years prior to the initiation of the program (or even the passage of the legislation). In addition, the measured readmission rate fell significantly just after the penalty went into effect, implying that opportunities to respond had been developed, stored away, and were able to be scheduled and implemented quickly.

Subsequent research has raised issues about the effectiveness of HRRP for spending reductions and health outcomes. Cutler et al. (2019)[5] noted that the onset of the penalty coincided with significant changes in the measured risk of hospitals subject to the penalty. These researchers speculate that there may have been no change in readmission behavior in such hospitals, but rather a change in risk that affected readmissions that would account for more as it was instituted, as well as linked improvements in mortality.

Some studies also found suggestive – though far from conclusive – evidence that the 30-day mortality rate may have risen soon after the program took effect at impacted hospitals.[6] These hospitals substituted longer stays in the emergency department ("observation stays") or transfers to post-acute care facilities for readmissions, and researchers inferred that death was more likely in such settings than if the patient had been readmitted as an inpatient. However, the 45-day mortality rate was unaffected, suggesting that any benefit from readmission was a few more days of survival. Thus, there was no rigorous prior evidence on the benefits and risks of HRRP, while subsequent evidence suggests that cost reduction may have been less than initially estimated and some adverse health outcomes may have occurred. In contrast, recent analysis by Gupta finds that the program did account for the bulk of the reduction in readmissions and was associated with some improvements in mortality rates.[7]

Hindsight cannot directly determine what should have been done. However, had this information been available at the outset, would

[4] Zuckerman, R. B., Sheingold, S. H., Orav, E. J. et al. (2016). Readmissions, observation, and the hospital readmissions reduction program. *New England Journal of Medicine*, 374(16), 1543–1551.

[5] Ody, C., Msall, L., Dafny, L. S. et al. (2019). Decreases in readmissions credited to Medicare's program to reduce hospital readmissions have been overstated. *Health Affairs*, 38(1), 36–43.

[6] Krumholz, H. M., Lin, Z., Keenan, P. S. et al. (2013). Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA*, 309(6), 587–593.

[7] Gupta, A. (2021). Impacts of performance pay for hospitals: The Readmissions Reduction Program. *American Economic Review*, 111(4), 1241–1283.

Medicare's managerial decision have been different? One question pertains to the perceived likelihood of what actually happened at the time of implementation: How likely was it that the program would have gone off without a hitch? No one knows the answer. There were no prior estimates of cost savings that informed the decision. The threat of worsening mortality rates was discounted based on the observation that low readmission hospitals did not have significantly worse outcomes. The possibility that hospitals differed in managerial skills and styles rather than responsiveness to financial incentives was not considered.

We may be able to provide some information on yet another question: Would recommendations for HRRP have gone ahead if decision-makers knew in advance it might lead to an increase in short-term mortality and only moderate cost savings? Would evidence suggesting this possibility have been enough to influence the decision? This is the larger question at play here – as is the question of whether gathering such information by a prior trial rather than by before-and-after observations with no controls would have been worth the time delay and the research costs.

In this initiative, we do not believe that the answer is obvious because we do not know how decision-makers would value these trade-offs. But this example does show the preventive benefits as well as the need for a study to provide prior evidence in decisions about program initiation.

Would having this information, or better and more reliable information, have changed the decision to implement HRRP? The embarrassment of demonstrating worse health outcomes might have been the deciding factor. How much worse would the outcome had to have been, and how firm would the evidence need to be?

This example illustrates several points that will be further discussed in this book: (1) If good information on the effectiveness of an intervention can be obtained with limited time or resource investment, decisions will be improved; (2) However, information obtained at great cost and time will be of limited value unless unanticipated side effects and outcomes are found that would change substantially the estimated or suspected balance between costs and benefits; (3) Regardless, the importance placed on the costs and benefits will drive decisions. In the case of HRRP, If developments after implementation proxy what would have been found with evidence development, it is not certain that the adoption decision would have been different.

The primary conclusion is that the key factor is an "unknown unknown." What impacts are likely to be turned up by better evidence as

compared to current evidence? Some decision-maker has to make an estimate of this factor, but the task will be challenging.

CORONAVIRUS 2020

The onset of the novel coronavirus epidemic has exacerbated the improper and controversial use of evidence in decision-making. This new threat to health has increased ("exponentially," as they say) the number of impending decisions, of which a large share rely on evidence that is either absent or inconclusive. Swift decisions have had to be made without gold standard, randomized controlled trial evidence, given the novelty of the virus. Throughout the health care system, protocols have been overturned, drugs have been redirected, and conjectures about what does or does not seem like a good idea for management or treatment have proliferated.

With the novel coronavirus, the model of obtaining rigorous evidence before decision-making has been seriously and appropriately challenged by the immediacy of making crucial decisions about new drugs, new uses for old drugs, changes in care provision, and the need for staff protection. The decision model we will describe in this book can be used to guide best practices in this or any bad situation – and it will yield the common-sense conclusion that neither waiting for bulletproof evidence nor going on a hunch will be optimal. The case for the strongest model of evidence-based decision-making, such as that used by the US Food and Drug Administration, will need to be modified.

The dust still has to settle on a tumultuous period made worse by partisan criticism. Some innovations will turn out to be mistakes, others that work will have been delayed too long, and much will remain unknown and unresolved. We intend to avoid advocacy and criticism, but inevitably, in a world where politicians assume positions without appropriate facts, the actual rigorous research will dispel some of those positions. Moreover, the subjective nature of decisions when neither facts nor probabilities are known – situations of true (or, in the title of a recent book, "radical")[8] uncertainty mean that differences of opinion that cannot be settled by "science" will necessarily remain. We will use and describe the use of as much evidence as possible to circumscribe the limits on reasonable differences of opinion. After that, readers will have to make their own bets on which sources of evidence and authoritative speculation to trust.

[8] Kay, J., and King, M. (2020). *Radical Uncertainty: Decision-Making beyond the Numbers.* New York: W. W. Norton & Company.

OUTLINE OF THE BOOK

The next two chapters provide general observations on the "unknown unknowns" problem. One chapter looks at aggregate data on spending levels, growth, outcome levels, and changes over time. It asks what evidence we currently have for these patterns, and what evidence remains conjectural. It further questions how to relate decisions on specific innovations to the visible, although sometimes misleading, aggregate patterns. Can we relate the trends in total spending to reports on transformations that have lowered spending without harming health care quality – or is there no connection? The next chapter models the managerial problem of when to seek more rigorous information on the likely outcome of an intervention. It then goes on to discuss the subsequent question of how to respond to that information. It concludes with the issue of how to tell after the fact whether better information helped.

The remaining chapters are case studies of particular classes of decisions where information has or has not been available and/or used. In each case, the core question is: Given a set of decisions to be made, would those decisions have been different and/or better had they been made with more (or less) rigorous evidence? The goal here is to draw conclusive statements in each case about "evidence on the role of evidence" in this particular setting, or, if not to provide statements about actual evidence, at least offer a discussion of reasonable expectations about how much benefit there was or would have been from better evidence – and how that compares with the cost of that evidence. We think it is important for leaders in health care management and policy to be convinced that there is a problem with a lack of good evidence for decisions they make on innovations in care delivery and financing. Once they decide no good evidence exists, it is important to identify where that problem is most severe. But in addition to this task, we will offer ideas on what can be done – some based themselves on evidence of interventions that have worked or on efforts that have identified the blockages in obtaining and using evidence. Other ideas are generated based on the logical diagnosis of the malfunction and sensible correctives. There are informed actions that can be taken, and we will tell you what they are.