The Road to Better System Performance – Enabling Factors and Stumbling Blocks to Improve Overall Quality of Care

(AKA From Volume to Value – Aspiration Meets Reality)

Robert Berenson M.D.

Institute Fellow, Urban Institute

rberenson@urban.org

Zi Health Services Research Conference 2017:

Learning from the Regions – Benefit for all?

Berlin 13 Sept. 20017

The presentation will:

- Review the roots and concepts underlying “value-based payment” in the US;
- Explore the various conceptual and operational challenges in reasonably assessing value;
- Provide and discuss a broadly used US scheme of the continuum of alternative payment models -- and why it is flawed
- Conclude with a contrarian viewpoint on value and how to achieve it

Geographic Variations in Spending

Dartmouth research and the Dartmouth Atlas represent seminal work exploring spending variations in Medicare and demonstrating large cost variations with little or no differences in quality or patient experience

When properly adjusted for approved variations in input prices, graduate education, etc., there is a 30% variation in spending (service use) between the 10th and 90th percentile across US regions

But in the US multi-payer context, things are always more complicated

- For private insurers who have to negotiate prices with providers (hospitals, physicians, and vertically integrated health care systems), prices, rather than service use drive spending variations
- Prices for hospital services, such as DRGs, vary by 100% between the 10th and 90th percentile – due to provider concentration and more negotiating leverage
- Hospital prices charged to private insurers exceed Medicare prices by a lot more than physician fees

Paying for value, not volume, is one of the few policy areas that gets bipartisan political support

- "The most powerful way to reduce costs (and make room to expand coverage) is to shift away from 'volume-based' reimbursement (the more you do, the more money you make) to 'value-based' reimbursement." – Bill Frist, M.D., then the Republican leader of the Senate during debate on the Affordable Care Act (AKA Obamacare).

- "Doctors and hospitals will have to be paid differently. Not simply for procedures – the more they do the more they make – but for outcomes." – Dr. Tim Johnson, ABC News, the night the ACA passed Congress

The failure of the “sustainable growth rate” was a catalyst for measuring performance at the hospital/clinician level

- The SGR provided a spending cap for physician services, based on both Gross Domestic Product and population growth. If spending exceeded the target, fee increases were then supposed to be limited to pay back the excess
- The SGR did not work – for many years service volume growth was not restrained, and the SGR-mandated fee cuts were postponed each year
  - “The tragedy of the commons”
- Policy makers concluded that value has to be measured at the individual, not aggregate, level to avoid perverse behavior

The Triple Aim

- "The Institute for Healthcare Improvement's Triple Aim is a framework that describes an approach to optimizing health system performance. It is IHI's belief that new designs must be developed to simultaneously pursue three dimensions, which we call the "Triple Aim":
  - Improving the patient experience of care (including quality and satisfaction);
  - Improving the health of populations; and
  - Reducing the per capita cost of health care."
    - IHI website

Policy priorities in Medicare as reflected in legislation over 10 years

- Pay-for-performance (labeled “value-based purchasing”) for most provider systems, including individual physicians
- New payment models to be tested in demonstrations: e.g., shared savings, bundled episodes, mixed fee schedule and “capitation”
- New organizational delivery models, esp. “accountable care organizations” and patient-centered medical homes”

Issues in Measuring Value in Health Care

There is a disagreement over the role of measurement in value-based payment

- For many, value-based payment means literally measuring quality and costs and directly rewarding the result, where Quality/Costs = Value
- For others, it mostly means adopting payment methods that have a greater demonstrated relationship to desired outcomes (the Triple Aim) and using measures more opportunistically, relying more on the design of payment to affect value

What do we really mean by value?

- Value = Quality/Costs
- But there is no quantitative precision to the value equation
  - For example, is value increased when quality increases at a higher cost?
- So “value” more casually actually is used to suggest getting a “bigger bang for the buck”

The quality numerator

- Quality is measured differently for each measure of interest. e.g., % compliance with a quality standard, mortality rate for a condition or intervention. There is no common metric like quality-adjusted life years (QALYS) as used in cost-effectiveness analysis
- There are huge measure gaps such that what we do measure may not reflect aggregate quality at all
  - Important gaps include diagnosis accuracy, appropriateness of procedures and other treatments, management patient with multiple chronic conditions

The cost denominator

- Costs are usually measured as dollars spent but can also represent the *rate of increase* in dollars spent (we often refer to “bending the curve” of health spending
- And even with something as seemingly straight-forward as dollars spent, there are disagreements on how to measure and report costs (beyond the common error in the US of mistaking charges or payments with actual costs)

The guiding mantra – “If you can’t measure it, you can’t manage it”

- And its close cousin, “If something… cannot be measured, it cannot be improved.”
- Called a “truism,” the quote is commonly attributed to W. Edwards Deming, who was a widely revered expert in management and management science, and instrumental in the economic recovery and success of Japan after WWII.

What Deming actually wrote and believed

- "It is wrong to suppose that *if you can't measure it, you can't manage it* – a costly myth."
  - *The New Economics*, 1994, page 35.
- So not just taken out of context, but an overt misquote – or in Trumpian reality, "alternative facts"
- Other consistent Deming quotes (of many available):
  - "The most important figures one needs for management are unknown or unknowable, but successful management must nevertheless take account of them." *Out of the Crisis*, 1982, p 121

The dueling slogans

- *"If you can't measure it, you can't manage it"*
- *"Not everything that can be counted counts, and not everything that counts can be counted."*
  - Commonly attributed to Albert Einstein, it was actually coined by a sociologist named William Bruce Cameron, writing after Einstein had died

No wonder the US doesn't do evidence-based policy making

It can't even get quotes right

It is important to distinguish measures for public reporting & P4P and for internal QI

- Of course, for internal process improvement, having data is often highly desirable or even necessary – both outcomes and to measure reliability of the processes
- My concern is about the public policy infatuation with public reporting and P4P, not with how organizations use measurement as part of internal efforts to produce reliable processes (what Deming actually emphasized)

Conceptual concerns about P4P

- P4P is too often presented as having “compelling logic” and “face validity” – not so fast
- Behavioral economists now have started to weigh in, arguing that P4P can “crowd out” intrinsic motivation, with an overall negative impact
- Across education, health care, other sectors, concern about “teaching to the test”
- Together, these two could result in overall performance decline even if incentivized performance improves -- we measure much less than what we care about

Accurate measurement is difficult

- Even relying on a seemingly simple and important metric, such as a hospital's readmission rate, can be misleading – this is a policy priority in Medicare
  - Like when a health care system successfully reduces both readmissions and admissions – the change in the ratio – the readmission rate -- may not reflect its success
  - Like when hospitals serving very different patient populations in geographic areas with different resources are compared to each other in a "tournament" P4P
  - Like when there may be an opportunity for hospitals to engage in "regulatory evasion" by calling an admission an "observation" stay (this one is in dispute in fact but not in theory)

Other operational challenges with P4P

- Major gaps in available measures, which rely largely on claims data, with no other reliable and affordable sources on the horizon
- Small numbers, often making statistically valid inferences of individual clinician performance problematic
- High administrative costs – a *Health Affairs* (Casalino and others, Mar, 2016) paper estimated $15.4 billion/year just for physician practices' reporting
- Provider "gaming" behavior in response to P4P, to the detriment of patient care, e.g.,
  - Avoiding complex patients, when inadequate case-mix adjustment
  - An increase in 31-day mortality rates

Proliferation of “information brokers” which rate physicians, hospitals and health plans

- But there are serious questions about validity and reliability of these ratings and rankings
- There was disagreement across four prominent rating systems, with each identifying different sets of high- and low-performing hospitals – limited overlaps across the 4 ratings
  - each evaluation system uses its own rating methods, has a different focus to its ratings, and stresses different measures – Austin, et al. Health Affairs, Mar. 2015

There's a growing body of good studies and lit reviews finding that hospital and physician P4P don't produce outcomes better than secular trend. What's not clear is how much public reporting -- a form of P4P in its possible impact on moving market share -- contributes to the improved secular trend

Nevertheless, despite a raft of conceptual and operational concerns, and the unimpressive empirical findings, Congress has increased the importance of and reliance on P4P in payment policy

The current policy infatuation with public reporting and, especially, P4P has lead to this perverse policy result:

What we measure publically is considered important and demanding attention while

What we can't or don't measure is marginalized or ignored altogether -- like diagnosis errors, a largely ignored quality problem or the workforce needs for an aging population

How to improve the use of measurement in public policy (from Berenson, Pronovost, and Krumholz, 2013)

- Use measures strategically as part of major quality improvement initiatives, not as ends in themselves;
- Measure at the level of the health care system and then the organization, not the clinician ("information brokers" do that);
- Expedite moving from processes to outcomes (but not easy);
- Place greater emphasis on patient experience and patient-reported outcome measures as important in themselves;
- Invest more in the "basic science" of measurement development, tasking a single entity with defining standards for measuring and reporting performance:
  - to improve the validity and comparability of publicly-reported quality data and
  - to anticipate and prevent unintended adverse consequences

The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)

The MACRA *Quid Pro Quo* for repeal of the Sustainable Growth Rate

- The actual substantial cost of the not reducing physician fees by >20% was paid for through a long-term schedule of nominal fee increases
- The *quid pro quo* in the legislation was moving payment from “volume to value”
- There are two arms of the strategy – the Merit-based Incentive Payment System and Alternative Payment Models. Clinicians (or their organizations) are given a choice

MIPS assessment categories

- Quality (30%)
- Resource Use (30%)
- Electronic Health Record Performance (25%)
- Clinical Practice Improvement Activities (15%)
  - Such as expanding practice areas, population management, care coordination, beneficiary engagement, patient safety

(percentages when fully phased in in 2022)

MIPS payment adjustments

- Those performing at 0-25% of target thresholds get maximum negative adjustment
  - 2019: - 4%
  - 2020: - 5%
  - 2021: - 7%
  - 2022: - 9%
- Positive adjustments
  - Maximum: 3 X annual cap for negative adjustment – so theoretically as much as 27% bonuses (I am not kidding)
  - Eligible for additional payment if 25% above performance threshold
- The negative adjustments basically fund the bonuses

The food here is terrible -- and such small portions

-- a 1920's joke (used by Woody Allen in "Annie Hall")

Alternative Payment Models

CMS/LAN APM Framework

Some Observations About the CMS/LAN* Framework

- Emphasizes theoretical incentives in payment methods, mostly ignoring the design and operational issues that working together influence clinician behavior
- Assumes that value derives only from 1) use of quality measures and 2) financial risk-bearing
- In short, the Framework that actually classifies 28 payment models is useful for presenting a continuum of payment method structural elements (measures and risk) but errs in implying that value follows the same continuum
- **Any payment method can be designed to produce more or less value – and that includes classic fee-for-service, in the US case, the Medicare Physician Fee Schedule**

* LAN = Learning Action Network

Attributes of fee schedules (for example)

Advantages

- Rewards activity, industriousness
- *Theoretically can target payment to promote desired behavior*
- Implicitly does case-mix adjustment
- Commonly used by payers and physicians

Disadvantages

- Can produce too much activity, physician-induced demand
- *Maintains fragmented care provided in silos*
- High administrative and transaction costs
- *What is not defined as payable is marginalized*
- Complexity makes it susceptible to gaming and to fraud
- Susceptible to pricing distortions that alter impact

G. B. Shaw, the Doctor's Dilemma (1909):

"That any sane nation, having observed that you could provide for the supply of bread by giving bakers a pecuniary interest in baking for you, should go on to give a surgeon a pecuniary interest in cutting off your leg, is enough to make one despair of political humanity. But that is precisely what we have done. And the more appalling the mutilation, the more the mutilator is paid. He who corrects the ingrowing toe-nail receives a few shillings: he who cuts your inside out receives hundreds of guineas, except when he does it to a poor person for practice."

Alternatives to P4P Would Emphasize

- Reducing the power of the incentives in volume-based payment that produce too much care instead of trying to counter these dominant “FFS” incentives with small, P4P dollars
- Alternative Payment Models potentially try to do that, but most are thin layers on top of FFS
- *“The most powerful methods for reducing medical harm are: feedback, learning from the best, and working in collaboration”*
  - Lucian Leape, M.D. commenting on the Michigan Keystone Project eliminating CLABSI in MI hospitals

Thank you