Written Testimony of Brent C. James, MD, MStat

Hearing on
Reducing Health Care Costs: Eliminating Excess Health Care Spending and Improving Quality and Value for Patients

Senate Committee on Health, Education, Labor, and Pensions
Honorable Lamar Alexander, Chairman
Honorable Patty Murray, Ranking Member

17 July 2018

Summary

- **Health care delivery "waste" describes any consumption of resources that does not provide optimal benefit to a patient**. Under Deming's quality improvement theory, **higher quality eliminates waste**. This defines "value" – the best quality result at the lowest necessary cost. Deming's theories initially transformed manufacturing around the world. Starting in the late 1980s, clinical investigators demonstrated that Deming's theories apply in health care delivery.

- In 2010, an Institute of Medicine (IOM) expert panel conducted an evidence review. They estimated that **a minimum of 30 percent, and probably over 50 percent, of all money spent on health care delivery is waste recoverable through higher quality**. **Some analyses suggest that waste levels may be much higher**.

- This year, total expenditures on health care delivery in the United States will approach $3.6 trillion. Midpoint estimates suggest as much as **$1.8 trillion in recoverable waste**. More than **half of health care spending, and associated waste, is funded through government**.

- Most research on health care waste comes from the United States. However, evidence from other countries (e.g., Canada, Australia, and European democracies) suggests that **health care waste levels are similar across the world**.

- **The primary drivers of waste are** (1) care delivery execution that still relies primarily on **personal expertise and human memory** (the "craft of medicine"); (2) in the face of high and rapidly increasing **complexity** of clinical practice, that "exceeds the capacity of the unaided expert mind"; framed within (3) legal structures, cultural expectations, and **payment methods that actively encourage utilization**. Waste estimates include healthcare fraud and abuse. However, these factors are small compared to other sources.

- A series of at-scale projects have shown that **quality-based waste recovery is achievable using available tools**. For example, one Utah-based health system improved patient outcomes and thus reduced total operating costs 13 percent across 4 years ($688 million, through 2015). Other examples abound.

- Waste elimination through higher quality offers health care providers financial opportunities that dramatically exceed other sources. However, most care providers are not actively pursuing broad quality-based waste elimination. That is primarily because **payment mechanisms create misaligned financial incentives**.

  Improving quality to eliminate waste always requires that care delivery groups invest in new systems and change leadership. Under current payment mechanisms, the resulting waste savings often go to someone else. That can place the care delivery group under financial stress and leave them without resources for future projects. Under fee-for-service (FFS) payment only about 5 percent of quality-based waste elimination generates compensatory savings back to the care delivery group that must invest and make the change. Adding per case payment (DRGs) increases alignment to about 55 percent. The final 45 percent of potential savings requires that care providers bear financial risk – various forms of shared savings or directly capitated compensation.

- One person's waste is another person's income. Thus, **health care waste vigorously defends itself** – through traditional health management methods and, often, through political mechanisms.

A. Health care delivery "waste" describes any consumption of resources that does not provide optimal benefit to a patient.

Dr. W. Edwards Deming is the father of quality improvement theory. Quality improvement is the science of process management. Deming based his work on 3 foundational principles. The first 2 of Deming's 3 premises were:

1) Premise 1 – All productive human work, of any sort whatsoever, can be described as a process. The definition of a process: "A series of linked steps, usually but not always sequential, designed to create a product, transform an input into and output, produce an experience, generate information, or in some other way add value."

   On that foundation, Deming argued that any enterprise should organize literally everything around "value-added front line work processes," where "value-added" is defined by a customer.

2) Premise 2 – Every process produces 3 parallel classes of outcomes:
   a. A "physical outcome" is the product or service that the process was designed to create. In clinical care delivery, we call these medical or clinical outcomes.
   b. A "service outcome" describes the interaction that takes place between the producer of a product or service and the consumer of that product or service, as the transaction takes place, as experienced by the consumer. This is patient satisfaction – the care delivery experience.
   c. A "cost outcome" represents the resources consumed to operate the process. Treating cost as the outcome of a process, rather than as an input, made Deming's approach unique. It fundamentally redefined the concept of value, defined as the ratio of the quality of a product divided by its cost.

   The term "quality" describes the relative attributes of any outcome. Thus people speak of "clinical outcome quality" or "service quality." From a theoretic perspective, it is similarly appropriate to consider the "quality" of cost outcomes.

   The fact that every process always produces all 3 categories of outcomes means that the 3 classes of outcomes are intertwined. It is impossible to functionally separate them, from an operational perspective. For example, a physician may make a change to a treatment process with an aim to improve a clinical outcome. That will, by definition, also change the process's cost outcomes. Alternatively, an administrator may change that same process with an aim reduce costs. That will, unavoidably, change the process's clinical outcomes.

   Deming next began to explore the interactive relationships between physical and cost outcomes. A surprising finding emerged: The linkages between physical and cost outcomes were not always negative. To that point, everyone had always assumed that higher quality always meant higher cost. Deming showed that some major classes of process changes, when introduced to improve physical outcome quality, caused costs to fall. He identified 3 causal mechanisms by which physical and cost outcomes interact. The first 2 interactions form the basis for all quality-based definitions of "waste" (James, 1989):

a. **Quality waste** – A step in a process fails. Some proportion of the time (it doesn't have to be 100 percent), that process failure causes a physical outcome failure – a "quality" failure. When that occurs, the process operator has only 2 options:

   1. The process operator can repair the low quality product. This is called "**rework**" in quality theory. The problem: Rework – repairing the failed product – always involves additional time and resources. In other words, it always costs more.

      When a process operator detects a failure, the best response is to "move upstream" into the process, figure out where and how it failed, then fix the process so that it will not fail again. It is always cheaper to "do it right the first time" than to "fail then repair."

      For example, Reiss- Brennan *et al.* created a 3rd generation primary care medical home called Team-Based Care (TBC) (Brennan *et al*., 2016). They deployed chronic disease management, mental health integration, and care coordinators into primary care practices. As patients received better clinical management in a primary care setting, specialty visit rates fell by 21 percent, and hospitalization rates fell by 22 percent. Overall, deploying TBC cost $22 per person per year (a not-insignificant investment), but total medical expense fell by $115 per person per year (a five-times savings, compared to the investment). Seen through the lens of quality improvement, specialty referrals and hospitalizations represented failures of upstream primary care processes.

      Similarly, preventable care-associated patient injuries (patient safety) represent quality waste. It is nearly always much cheaper to avoid patient injuries from the start, than to treat them after they occur.

   2. If the failed outcome does not involve a human life, then the process operator could simply discard it. This is called "**scrap**" in quality theory. Obviously, all of the time and resources consumed to create the scrapped product are wasted.

      For example, a hospital clerk runs and prints a large report, only to discover that the date range used in the report was wrong. That wastes not just the discarded paper, but the human and computer time consumed to produce the report.

b. **Inefficiency waste** – 2 parallel processes produce identical outputs. One of those processes use significantly more resources to achieve that goal. The unnecessary use of resources represents waste.

   In the late 1980s, clinical research teams at Intermountain Healthcare examined treatment details for common conditions routinely managed in hospitals (transurethral prostatectomy (TURP), cholecystectomy (gallbladder removal), total hip arthroplasty (artificial hip joint implantation), coronary artery bypass graft surgery (CABG), community-acquired pneumonia, and implantation of permanent cardiac pacemakers) (James, 1995). For statistically identical patients with statistically identical clinical outcomes, they found about

a 2-fold difference in resources consumed. For example, when performing a TURP on a standard patient, one urologic surgeon consumed on average 1184 hospital dollars to achieve a good outcome (these were 1986 dollars – medical inflation has greatly increased those numbers across the years). Another surgeon in the same hospital averaged $2233 for an equivalent patient with the same good clinical outcome.

For both quality waste and inefficiency waste, process management offered an opportunity to reduce operating costs by producing better physical outcomes. Deming proved that better quality could drive lower costs.

There is a third way in which physical and cost outcomes interact, that fall outside Deming's "waste" mechanisms:

c. **Cost effectiveness** – In some circumstances, clinicians find a treatment process that produces better outcomes, but the new process appropriately consumes more resources to produce that result. When that happens, those who both stand to benefit and to compensate the required higher resource consumption face a choice: Does the amount of quality gained by using the new process justify its additional expenses. Obviously, this is a choice that health care consumers must ultimately make.

Under Deming's quality improvement theory, **higher quality can eliminate waste and reduce costs**. This defines "value" – the best quality result at the lowest necessary cost. Deming's theories initially transformed manufacturing around the world. Any company that could not master his process management methods to produce higher value – better quality at lower costs – could not compete with companies that could. It became a litmus test for survival in many industries, and produced a maxim: Do Deming or die.

Starting in the late 1980s, clinical investigators demonstrated that Deming's theories apply in health care delivery. They realized that Deming's approach took concepts found in preventive medicine, and generalized them.

Clinical quality improvement's prevention-based approach raises 2 questions:

- How much quality and inefficiency waste exists in health care delivery?

- While theory is useful, it does not always accurately reflect implementable reality. Do these principles apply and produce expected results in real care delivery experience?

B. In 2010, an Institute of Medicine (IOM) expert panel conducted an evidence review (IOM 2010). They estimated that **a minimum of 30 percent, and probably over 50 percent, of all money spent on health care delivery is waste recoverable through higher quality**. **Some analyses suggest that waste levels may be much higher**.

C. In 2018, total expenditures on health care delivery in the United States will approach $3.6 trillion. Midpoint estimates suggest as much as **$1.8 trillion in recoverable waste**. More than half of health care spending, and associated waste, is funded through government.

D. Most research on health care waste comes from the United States. However, evidence from other countries (e.g., Canada, Australia, and European democracies) suggests that **health care waste levels are similar across the world**.

The 2010 IOM report is currently the best published citation for waste in health care delivery. Subsequent reports derive from it. It started with Deming's ideas of quality waste and inefficiency waste, then catalogued specific examples of care delivery waste that various researchers had documented.

Outside the U.S., most countries lack the detailed financial data that make direct waste estimates possible. However, care delivery systems in other countries have studied clinical process failures and generated estimates of failure rates. That is the basis for asserting that financial waste rates in other countries mirror those seen in the United States. The waste results from the process failures. While the U.S. invests much more heavily in health services research than other countries, there is sufficient evidence to support the conclusion that care delivery process performance, and process failures, are similar.

Several years later, James & Poulsen published a financial model that aligns Deming's waste categories to health care delivery operations (James & Poulsen, 2016). That approach included additional categories of waste that did not appear in the 2010 IOM report, and so produced higher estimates of the amount of waste that currently exists in care delivery operations.

The James/Poulsen financial model has 3 tiers:

Tier 1 waste (the base) – A "unit of care" is any granular service or supply provided to a patient during a care delivery encounter. Examples of "units of care" include things like a single dose of a specific medications; a single specific type of imaging exam; a lab test; an acuity-adjusted hour of nursing time; a 6-minute block of physician time, adjusted by specialty; an acuity-adjusted minute in a procedure room (such as an operating theater); a bedpan; a box of tissue (Kleenex). U.S. hospitals maintain master lists of all possible "units of care" that they could possibly supply to a patient. Depending on other internal features, these lists are called "charge masters" or "cost masters." As a patient receives care during a clinical encounter, the treating facility records each unit of care consumed. Mapped through the charge or cost master, this allows the treating facility to create a detailed bill for all services provided during a clinical encounter.

One can usefully think of a care facility as a business that obtains, creates, maintains, and supplies these "units of care." In Tier 1, "waste" refers to any costs associated with any "unit of care" in excess of an absolute necessary minimum. The associated organizational function is called "supply chain." It is that part of a care delivery organization that obtains, assembles, and supplies to the point of actual care delivery all necessary supplies, including equipment and personnel.

Figure 1 labels Tier 1 "**Efficiency** (cost per unit of care)".

Tier 2 waste – During a clinical encounter physicians, nurses, and other clinicians select and apply different units of care, including their own time, to address patients' health care needs. "**Within-case utilization**" refers to the specific type and number of "units of care" used during a care delivery episode. Detailed studies of variation in care delivery typically focus at this level. Such studies identify cohorts of similar patients being treated for the same clinical problem, then track the type and

number of units of care used. They break total health care costs into two parts: The type and number of units of care consumed (utilization), and the true cost of acquisition of each unit of care (cost per unit).

Tier 3 waste pushes the idea of utilization a level higher. "**Case-rate utilization**" describes how often specific treatments are used in defined population.

	Waste rate within category (% of all cases)	Percentage of total health care costs to which this category applies	Percentage of total cost recoverable w/in category	Remaining value
				100.0%
3. Case-rate utilization (# of cases within a population)				
Inappropriate care	20%	100%	100%	80.0%
Care patients would not have selected if given a fair choice	40%	15%	100%	75.2%
Avoidable care	15%	100%	80%	65.6%
2. Within case utilization (# & type of units per case)				
Initial misdiagnosis, delayed diagnosis	15%	100%	25%	63.1%
Avoidable care-associated patient injuries	26%	10%	70%	62.0%
Variation in care delivery not driven by patient need	33%	100%	80%	45.6%
Operational inefficiency for health professionals	35%	40%	50%	42.4%
Avoidable administrative overhead	30%	15%	50%	41.5%
Excess insurance company profits	50%	20%	70%	38.6%
1. Efficiency (cost per unit of care)	5%	60%	100%	37.4%
			Proportion waste in care delivery:	**62.6%**

Figure 1. Breakdown of total health care delivery costs into 3 tiers, that build one on top of another.

Figure 1 adds specific subcategories within each tier, cataloguing known classes of waste in health care delivery. The columns to the right summarize data from the published literature regarding measured within-class rates of waste; the proportion of total health care costs to which that category of waste apply; and estimates of how much of that waste should be amenable to extraction using current technologies. The table correctly adjusts – for example, it applies waste savings that could be obtained by eliminating "within case utilization" only after all inappropriate care and avoidable care have been removed. Many of the estimates in the table are just that – the author's own estimates – and will be modified as better data and expert opinion become available.

Figure 1 adds one important category that other studies of health care waste did not include. Wallace and Savitz adapted Toyota Production System (TPS) Lean Observation to a health care setting (Wallace & Savitz, 2008). They tracked work performed within more than 60 different health profession roles, such a pharmacists, nurses of various specialties, hospitalist physicians, and central supply staff workers. They directly assessed those roles in 4 integrated care delivery systems (Intermountain Healthcare, Providence Health Systems, University of North Carolina Health Care, and University of Virginia). Every task performed was classified as "value adding" or "non-value adding" (waste) by expert observers, in real time.

The proportion of health worker's time judged "waste" ranged from 20 percent to over 70 percent. Overall, non-value adding activities – waste – comprised on average more than 35 percent of all health professional work time. Extrapolating to the entire health care workforce, their findings are summarized as "Operational inefficiency for health professionals" in Figure 1. By way of illustration, waste levels were placed at 35 percent (column 1), as estimated in the study. Worker salaries are estimated to comprise about 40 percent of all health care costs (column 2). Based on experience gained while addressing associated processes, the model estimates that 50 percent of such waste could be recovered with current technologies.

This model, while extending beyond those included in the 2010 IOM report, still leaves some sources of waste unaccounted. For example, it does not include estimates of clinician inefficiencies that track back to the structure of current electronic medical record (EMR) systems (Sinsky et al. found that physicians spend about 2 hours performing EMR-based administrative tasks for every hour they spend with patients – Sinsky et al. 2016).

This model and associated argument is the source, in conclusion (B) above, of the statement that "some analyses suggest that waste levels may be much higher"

E. The primary drivers of waste are (1) care delivery execution that still relies primarily on **personal expertise and human memory** (the "craft of medicine"); (2) in the face of high and rapidly increasing **complexity** of clinical practice, that "exceeds the capacity of the unaided expert mind"; framed within (3) legal structures, cultural expectations, and **payment methods that actively encourage utilization**. Waste estimates include healthcare fraud and abuse. However, these factors are small compared to other sources.

Arguments supporting this assertion are beyond the scope of this testimony. The author refers interested parties to James & Savitz, 2011 and James & Lazar, 2007; or invites those parties to contact the author directly.

F. A series of at-scale projects have shown that **quality-based waste recovery is achievable using available tools**. For example, one Utah-based health system improved patient outcomes and thus reduced total operating costs 13 percent across 4 years ($688 million, through 2015). Other examples abound.

A great many examples of clinical projects that show lower costs associated with better clinical quality are present available in the peer-reviewed medical literature. For purposes of this testimony, however,

a recently published experience at Intermountain Healthcare provides a solid "at scale" example (James & Poulsen, 2016 – *op. cit.*):

Intermountain Healthcare is a non-profit system of 22 hospitals, more than 190 outpatient clinics, and an associated HMO-model health plan. It supplies more than half of all health care services in Utah and some areas in surrounding States.

In 2010, Intermountain's Chief Financial Officer set a goal: In order to keep health services affordable, and thus accessible, to the patients Intermountain served, he asked the care delivery system to limit total health care costs increases to consumer price index inflation plus 1 percent (CPI+1). Intermountain's Finance department modelled that goal using "best estimates" of prior consumer price index inflation and health care total cost growth rates. Their estimates are shown in Figure 2. It required that Intermountain reduce its total operating costs by 13 percent across the next 5 years, through the end of 2016.

To achieve that goal, Intermountain's clinical leadership launched 5 major quality improvement projects, with an aim to control health care costs through better clinical outcomes. The heavy green line shows results for the first 4 years of the project. Across those 4 years, operating costs fell by $688 million – a 13 percent reduction in the system's expected total operating costs.

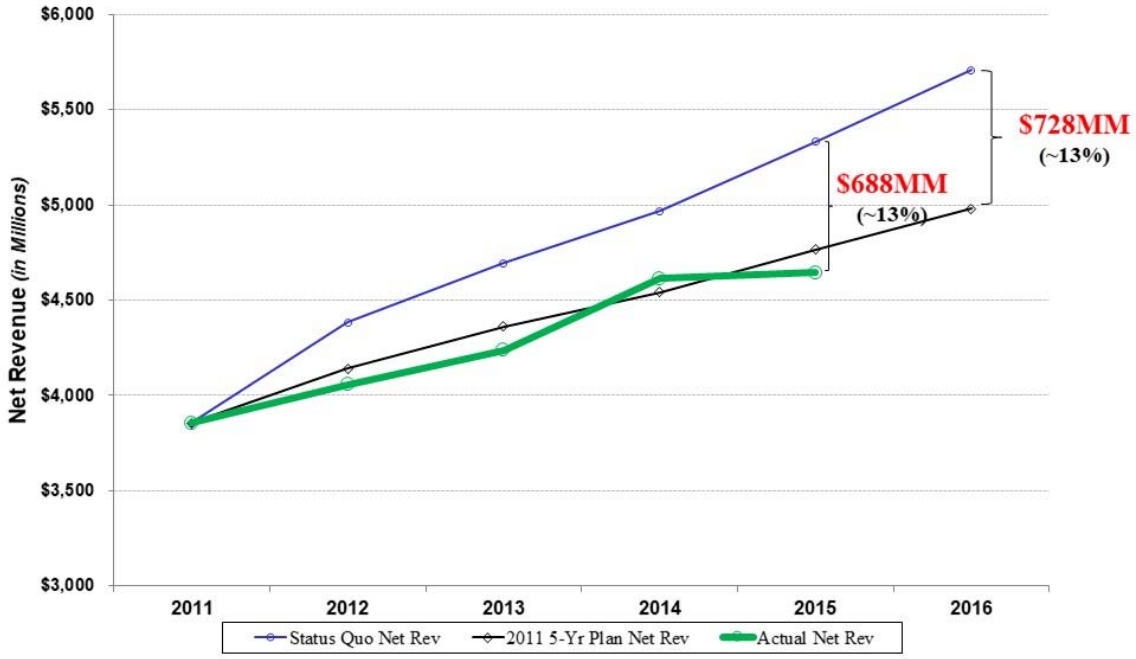

Figure 2. Financial consequences of waste elimination at Intermountain Healthcare from 2011 to 2015, achieved through clinical quality improvement. The solid blue line shows expected total health care costs for Intermountain's service population, taking into account general population growth, aging of the population and other population-based epidemics (e.g., Baby Boom entering chronic disease years, the obesity epidemic), and introduction of new treatment technologies. The solid black line shows 'allowable' growth in health care costs needed to achieve CPI+1 – a 13 percent reduction in total operating costs through 2016. The green line show actual total costs.

These results echo findings of waste elimination and cost savings demonstrated by a long list of other projects, at Intermountain and many other U.S.-based care delivery systems. They demonstrate, using current tools, it should be possible to dramatically reduce growth rates in health care costs.

G. Waste elimination through higher quality offers health care providers financial opportunities that dramatically exceed other sources. However, most care providers are not actively pursuing broad quality-based waste elimination. That is primarily because **payment mechanisms create misaligned financial incentives**.

As noted above, this year the U.S. will spend almost $3.6 trillion on health care delivery services. A midpoint estimate that about 50 percent of that spending is waste suggests a value opportunity of about $1.6 trillion. That dwarfs, by at least a factor of 100, any other opportunities for health care provider income growth. Return on investment for waste elimination projects are typically significantly larger than those for traditional service expansion approaches, as well (as above, the details of this analysis are left for another settings). Why, then, do such levels of waste continue? Why aren't health care markets driving care providers to very vigorously address and remove waste in the health care delivery system?

The reason: Financial incentives for waste elimination do not align. Figure 3 shows how the tiered classes of health care waste, defined in Figure 1, align with payment mechanisms.

WASTE REMOVAL LEVEL	% of all waste	PAYMENT METHOD		
		FFS	Per case	Provider at risk
3. Case-rate utilization *(# cases per population)*	45%	▼	▼	▲
2. Within-case utilization *(# and type of units per case)*	50%	▼	▲	▲
1. Efficiency *(cost per unit of care)*	5%	▲	▲	▲

Figure 3. Association of tiered waste categories with payment mechanisms. "FFS" stands for "fee for service" payment – still the most common method used to reimburse care delivery. Waste elimination always requires substantial investment, nearly always by care providers. The triangles in the table show who gets the savings when a waste-elimination project succeeds. Red triangles indicate that the savings go to payers, leaving those who must invest – the care providers – with no recompense for their initial investment.

Improving quality to eliminate waste always requires that care delivery groups invest in new systems and change leadership. Under current payment mechanisms, the resulting waste savings often go to someone other than those who must make that investment. That can leave care delivery groups under financial stress, without resources to fund future projects. Under fee-for-service (FFS) payment only about 5 percent of quality-based waste elimination generates compensatory savings back to the care delivery group that must invest and make the change. Adding per case payment (DRGs) increases alignment to about 55 percent. The final 45 percent of potential savings requires that care providers bear financial risk – various forms of shared savings or directly capitated compensation.

Figure 3 also summarizes the waste tiers in Figure 1, noting the total proportion of all waste opportunities associated with each tier. For example, about 45 percent of cost reduction opportunities function at the level of population health (Tier 3). Another 50 percent aligns to addressing variation in clinical practice, improving patient safety, and eliminating administrative overhead (Tier 2). Figure 3 assigns only 5 percent of total waste elimination opportunities to unit costs (Tier 1).

Health care delivery in the United States costs significantly more per person, and consumes more of total national wealth (as measured by percentage of Gross Domestic Product), than does health care delivery in other modern democracies. Papanicolas , Woskie, & Jha (Papanicolas, 2018) correctly note that unit costs, by themselves, explain the 2-fold difference health care spending seen in the United States as compared to other countries.

The reason that Figure 3 assigns only 5 percent of total waste elimination opportunities to unit costs (Tier 1) is because so many of the elements that drive higher unit costs are outside of the control of health care providers. We arrived at the 5 percent estimate based on observations of the gains achieved by successful supply chain operations in leading care delivery systems. We also note that, given the size of the U.S. health system, even a 5 gain is consequential.

H. One person's waste is another person's income. Thus, **health care waste vigorously defends itself** – through traditional health management methods and, often, through political mechanisms.

This argument, too, is left for further discussion beyond this document.

References

James, B.C. *Quality Management for Health Care Delivery* (monograph). Chicago, IL: Hospital Research and Educational Trust (American Hospital Association), 1989.

Reiss-Brennan B, Brunisholz KD, Dredge C, Briot P, Grazier K, Wilcox A, Savitz L, and James B. Association of integrated team-based care with health care quality, utilization, and cost. *JAMA* 2016; 316(8):826-34 (Aug 23/30).

James, Brent C. What is a TURP? controlling variation in the performance of clinical processes. *Improving Clinical Practice: Total Quality Management & the Physician* (ed: D.B. Blumenthal and A.C. Scheck). San Francisco, CA: Jossey-Bass Publishers, 1995 (Chapter 7).

Institute of Medicine Roundtable on Value and Science-Driven Healthcare. *The Healthcare Imperative: Lowering Costs and Improving Outcomes*. Yong, Pierre L., Saunders, Robert S., and Olsen, LeighAnne, editors. Washington, DC: National Academy Press, 2010.

James Brent C and Poulsen Gregory P. The case for capitation: It's the only way to cut waste while improving quality. *Harv Bus Rev* 2016; 94(7-8):102-11, 134 (Jul-Aug).

Wallace, C. Jane and Savitz, Lucy. Estimating waste in frontline health care worker activities. *J Eval Clin Pract* 2008; 14(1):178-80 (Jan).

Sinsky *et al.* Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Ann Intern Med* 2016; 165(11):753-60 (6 Dec).

James Brent C and Savitz Lucy A. How Intermountain trimmed health care costs through robust quality improvement efforts. *Health Affairs* 2011; 30(6):1185-91 (June).

James Brent C and Lazar Joel S. Sustaining and extending clinical improvements: A health system's use of Clinical Programs to build quality infrastructure. *Practice-Based Learning and Improvement: The Clinical Improvement Action Guide, 2nd Edition*. Nelson Eugene C. and Batalden Paul, editors. Chicago, IL: Joint Commission Press, 2007.

Papanicolas Irene, Woskie Liana R., Jha Ashish K. Health care spending in the United States and other high-income countries. *JAMA*. 2018; 319(10):1024-1039 (Mar 13).