Research Paper

Engineering perspectives on healthcare delivery: Can we afford technological innovation in healthcare?

William B. Rouse*

Tennenbaum Institute, Georgia Institute of Technology, Georgia, USA

This article considers the national challenge of controlling healthcare costs. It is argued that this system-level challenge is better addressed with an engineering approach rather than from the perspective of medical science that, quite rightly, is tailored to address clinical practice. The engineering approach is outlined and three models are proposed for controlling the costs of healthcare so that the growth of these costs tracks the growth of gross domestic product. These models provide insights into the magnitude of efficiency gains needed to accomplish this goal. Possible ways to achieve these gains are discussed.

Keywords Healthcare; complex systems; economic models; affordability

INTRODUCTION

The enormous cost of healthcare in the United States is often cited as a key national challenge (Economist, 2004; CBO, 2008). Healthcare is consuming an increasingly large portion of the gross domestic product (GDP). At the same time, there are concerns that the quality of healthcare in the US lags that of other countries (IOM, 2000, 2001). It is clear that substantial improvements in the delivery of healthcare value are needed and, it is argued, achievable via value-based competition (Porter and Teisberg, 2006). Of course, it should be kept in mind that our healthcare system did not become as it is overnight (Stevens et al., 2006).

A recent report published by the Congressional Budget Office (CBO, 2008) attributes 50% of the cost growth of healthcare over the past four decades to technological innovation. Science and engineering research has yielded a steady stream of innovations for detection, diagnosis and treatment, many of which have resulted in 10–15% annual growth rates in terms of number of uses. Compounding such annual percentages over 40 years results in very large numbers. In many domains, such as personal electronics or cellular telephony, such growth would be seen as an enormous success. However, the third-party payers of most healthcare bills see this growth as an ominous threat to their financial positions and the viability of the overall healthcare system.
In this article, this threat is addressed as an engineering problem, rather than a problem of medical science. First, the engineering approach is outlined, and contrasted with that of medical science. The CBO’s conclusions are then explored a bit more deeply. Three models are then proposed for controlling the costs of healthcare so that the growth of these costs tracks the growth of GDP. These models provide insights into the magnitude of efficiency gains needed to accomplish this goal. The article concludes with discussion of possible ways to achieve these gains.

ENGINEERING APPROACH

Determination of how best to control healthcare costs should be approached as an engineering problem rather than a medical science problem. Medical science has made enormous contributions to the content of healthcare in terms of detection, diagnosis and treatment of disease. However, this discipline has not focussed the same level of effort on the processes whereby healthcare is delivered. Perhaps the best evidence of this is the way in which the NIH budget is allocated. Efforts to improve the overall system of healthcare delivery have received little NIH investment. Consequently, the full returns on our massive investments in medical science have not been realized.

The potential of engineering to enhance healthcare has, of late, received increased attention (Reid et al., 2005). This potential can be understood in terms of the following levels of understanding of any phenomenon:

- **Describe** past observations
- **Classify** past observations
- **Predict** future observations
- **Control** future observations
- **Design** future observations.

Science progresses from describing and classifying past observations to predicting future observations. If these predictions turn out to be accurate, science concludes that the theory or model employed has credence. If not, the theory or model needs revision. The goal is to create valid knowledge.

Engineering builds on scientific knowledge, particularly in using models to predict. However, engineers are usually not content to just predict. They want to control the state of the system of interest. Better yet, if possible, engineers seek to design or redesign the system to facilitate better control. In some cases, the penchant for design and control has enormous societal implications (McPhee, 1990). The goal is to create high quality, affordable systems.

**Prediction**

Somewhat simplistically, there are two basic approaches to prediction. One approach is extrapolation. Models (i.e. equations) are fit to data collected under particular conditions and these equations are then used to project the outcomes for similar conditions. Statistical models, such as used in medicine for randomized clinical trials (RCTs), are examples of equation fitting. To the extent that the conditions of the trials adequately reflect the eventual conditions of use, we can be reasonably confident that similar outcomes will be attained.

RCTs work well, although slowly and expensively, when there are large populations that can be observed under controllable conditions. However, this approach cannot be employed for study of large-scale systems like the healthcare system. There are simply not sufficient numbers of healthcare systems to achieve statistical significance in the study of the large, systemic changes likely to be needed to control costs and enhance quality to the extent outlined earlier.

Engineering approaches to large-scale problems typically rely on models as a basis for prediction. These models are formulated from ‘first principles’ drawn from a range of scientific domains. These principles, usually stated as fairly simple mathematical relationships, become elements of much larger mathematical and computational models that are used to predict outcomes of different approaches to design and control of complex systems. This approach is illustrated later in this article, drawing upon knowledge developed in domains other than healthcare.
Engineering models are validated in several ways. The ‘first principles’ have almost always been scientifically validated in laboratory studies or possibly proven mathematically. The constructs that emerge from connecting these elements are shown, over time, to have predictive validity, i.e. the predictions of the models match actual observations. Ultimately, engineering models are shown to be useful by engines that run, airplanes that fly and bridges that remain standing.

The idea that models are validated ‘over time’ is central to the engineering approach. Engineering has been historically driven by practical needs to cross rivers, move goods and so on. Engineers could not wait for complete knowledge of strength of materials, for example, before building bridges. The experiences gained, including failures, often provided direction to modelling efforts, to identify engineering principles that would improve engineering practices. In this way, validation of engineering models is inherently an ongoing effort.

**Control**

Engineering the control of a system involves measurement, feedback and compensation so as to achieve system objectives. Measurement concerns ascertaining the state of the system. This, of course, requires defining system state variables, their units of measure and how such measurements can be made. Feedback involves comparing predicted and actual system states in order to correct errors as they occur. Such feedback results in a ‘control loop’ and, ideally, learning. Compensation concerns adding dynamic elements to the control loop to counteract delays and lags in system response.

In the context of healthcare delivery, knowledge of the state of the system is critical to controlling the system. Medical science has focussed quite effectively on defining, measuring and controlling the states of patients. However, the state of the overall system includes variables that medical science seldom addresses. Examples of such broader state variables include:

- The health state of each and every patient in the hospital at the moment, rather than on average.
- The distribution of labour hours per patient versus types of procedures and interventions today, rather than on average.
- The levels and locations of inventory for all consumables at the moment, rather than on average.

These types of variables are, quite rightly, not within the purview of clinicians providing patients with state of the art benefits of medical science. Yet, these types of variables have enormous impact on the costs of healthcare. Consequently, many have argued of late that we need a ‘science of healthcare delivery’ that draws upon best practices in systems engineering, in companion with medical best practices gleaned from medical science (Reid *et al*., 2005; Christenson *et al*., 2008).

**Design**

Engineering design involves problem analysis, solution synthesis, production of an artefact that embodies the solution and sustainment of the system in its use. Analysis of the problem of interest involves understanding input–output relationships, including uncertainties, and creating models as discussed above. Synthesis concerns designing input–output relationships to achieve system objectives. Production involves fabricating, constructing, programming, etc., to create systems that embody desired relationships. Finally, sustainment concerns creating mechanisms that assure future meeting of system objectives.

It can be argued that clinical practices have been designed in the sense that medical specialties continually research and refine the ways in which they detect, diagnose and treat patients with maladies falling within their specialities. However, the overall system of healthcare delivery was not designed – it emerged over time from the self-organization of large numbers of independent, individual practitioners (Rouse, 2008). The resulting system is highly fragmented and very inefficient, in part due to antiquated
information systems and incentives that are not aligned with providing quality, affordable care for everyone. This system of healthcare delivery needs to be designed – to be engineered – to create a seamless, high value healthcare delivery enterprise.

Summary

Engineering approaches to prediction, control and design have much to offer healthcare in terms of systemic improvement of the system by both decreasing costs and increasing quality. The engineering approach can take advantage of knowledge gained by medical science, but also can take advantage of other sources of knowledge to formulate the models needed for prediction, control and design of the system of healthcare delivery. The remainder of this article provides an illustration of how engineering can help tackle the challenge of healthcare.

HEALTHCARE COSTS

As indicated in the Introduction Section, the last four decades have seen enormous increases in healthcare costs. Specifically, real healthcare costs have tripled as a per cent of the GDP in the period 1965–2005, with half of this growth due to technological innovation (CBO, 2008). The magnitude of these increases has led some pundits to conclude that the nation is ‘running on empty’ (Peterson, 2005). There seems to be virtually unanimous agreement that something has to change significantly.

Figure 1 summarizes the overall phenomena discussed in the CBO report. Technological inventions become market innovations as they increase in effectiveness and associated risks decrease. This results in increased use that leads to increased expenditures. In parallel, increased efficiency via production learning (more on this below) leads to decreased cost per use, although not enough to keep up with growing use rates in healthcare. Finally, increased use yields improved care that leads to longer lives and increased chances of again employing the technology of interest.

The concern in this illustrative example is how to control the phenomena as depicted in Figure 1. More specifically, what efficiencies must be realized to steadily decrease the cost per use to offset the constantly increased use to enable affordable healthcare? In typical engineering fashion, we approach this control problem with a series of models, beginning with a very simple model and elaborating as the limits of each model become clear.

Model 1: Growth

The first model considers what efficiencies are needed to counteract the growth in Figure 1. We start with a simple equation.

\[
\text{Cost}(1 - \alpha) \times \text{Use}(1 + \beta) = \text{Total}(1 + \delta)
\]  

(1)

where \(\alpha\) is the annual rate of cost reduction, \(\beta\) is the annual rate of usage growth and \(\delta\) is the annual allowable total growth. A bit of algebra shows that the annual rate of cost reduction required is given by the following:

\[
\alpha = \frac{\beta - \delta}{\beta + 1}
\]  

(2)

Table 1 shows the cost reductions needed for five of the technologies discussed in the CBO report, assuming zero allowable growth. These
are rather significant decreases. However, these decreases are more instructive than definitive because of the simplicity of the model. In particular the model is quite limited in that it provides no mechanism for achieving cost reductions and does not differentiate elements of healthcare delivery processes. Thus, we need to elaborate our model.

Model 2: Learning

The second model considers production learning, a well-understood concept in industrial engineering (Hancock and Bayha, 1992). Quite simply, as we produce more of an item, we get better at it and unit costs decrease. In some industries, such as semiconductors, these decreases are a primary source of profit margins. Many manufacturing industries employ production learning curves to predict costs and hence profits. The notion of learning curves has been explored in healthcare in the context of cost accounting (e.g. Ahrens, 1999). This second model applies this concept to the overall healthcare system.

Model 2: Learning

The basic learning equation is given by

\[
\text{Cost}(t=T) = \text{Cost}(t = 0) \times \text{No. of Uses}(t = T)^{-\text{Rate}}
\]

(3)

This learning phenomenon is usually discussed in terms of ‘per cent curves’. For example, a 70% curve means that after each doubling of number of units produced, unit costs are 70% of what they were after the last doubling. Table 2 provides a few examples of rates needed in Equation (3) to achieve different per cent curves.

Most learning curves fall in the 70–90% range. This range reflects the experiences of many industries, including airplanes, automobiles and electronics. Percentages below 70% are quite rare. As our later results show, we may need to get significantly below 70% in healthcare if costs are to be controlled. This represents a very significant challenge.

Figure 2 shows learning curves for the three learning rates in Table 2, assuming a 10% annual rate of growth in usage. Note that the initial conditions were 100 uses at $100 per use, yielding an initial total expenditure of $10 000. Figure 3 shows the growth of total expenditures, again assuming 10% annual growth in usage.

Table 3 shows the overall results for annual growth rates of 5 and 10%, assuming a 70%
leaning curve. Unit costs have dropped significantly, but the growth of usage has overwhelmed these efficiencies. Overall, this model exhibits impressive cost reductions due to production learning, but it does not suggest where and how this learning happens. Further, the model does not reflect the process whereby healthcare is delivered.

Model 3: Process

The third model explicitly considers the process whereby healthcare service is provided. As shown in Figure 4, this process includes multiple stages and differentiates labour from technology. The value of thinking in terms of healthcare processes has been found to be useful, for example, in the context of management accounting in healthcare (Hankins and Baker, 2004).

We can define the learning rates for technology from a rich experience base. This allows us to, somewhat optimistically, set the technology learning rate at 70%. The question, then, is what labour learning rate is needed to control growth to acceptable levels.1 The goal is not to predict cost reductions, but instead determine what savings have to be achieved to achieve affordability.

This model is formulated in terms of Equations (4)–(8)

\[
\text{Cost}(t) = \text{Cost of Labour}(t) + \text{Cost of Tech.}(t) \tag{4}
\]

\[
CTOT(t) = \text{CPUL}(t) \text{ NU}(t) + \text{CPUT}(t) \text{ NU}(t) \tag{5}
\]

\[
\text{CPUL}(t) = \text{CPUL}(1) \text{ NU}(t)_L^{\text{Rate}} \tag{6}
\]

\[
\text{CPUT}(t) = \text{CPUT}(1) \text{ NU}(t)_T^{\text{Rate}} \tag{7}
\]

\[
\text{NU}(t) = \text{NU}(1) (1 + \beta)^{t-1} \tag{8}
\]

where CTOT, CPUL and CPUT denote total costs, labour cost per unit and technology cost per unit, and NU denotes number of units.

Figure 5 shows the required efficiency to control increases of healthcare costs and to track increases of GDP. The best case is for 4% GDP growth and 5% usage growth, which yields a need for a 70% learning curve for labour. This magnitude of learning is imaginable. The worst

---

1In this analysis, we have not differentiated the labour associated with the four stages of Figure 4, as we expect such differentiation would depend on the health issue being addressed.

---

Figure 3. Expenditure growth at 10% annual growth in use

Table 3. Impacts of production learning

<table>
<thead>
<tr>
<th>Rate (%)</th>
<th>Results at 30 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of uses</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>5</td>
<td>412</td>
</tr>
<tr>
<td>10</td>
<td>1586</td>
</tr>
</tbody>
</table>

Figure 4. Service delivery process model

Copyright © 2009 John Wiley & Sons, Ltd.

DOI:10.1002/sres
case is for 0% GDP growth and 15% usage growth, which yields a need for a 40% learning curve. This level of learning has never been achieved in any domain.

Implications

The implications of the results with these three models are quite clear. In order to limit the growth of total healthcare spending to the growth of the GDP, some combination of three things is needed:

- **Limiting the growth of technology use:** As Goodman *et al.* (2006) report, lower use rates are achievable as demonstrated by integrated medical centres.
- **Limiting the cost of technology use:** Here we can build on industry experience of learning as production quantities increase.
- **Decreasing the cost of labour associated with use:** Increased labour productivity is achievable through several means as discussed below.

Overall, the savings due to learning are the key to affordability. This will, however, be a significant challenge as learning rates less than 70% are very difficult to achieve.

Sources of Learning

In the industries where production learning curves have long been used, the sources of learning identified include labour efficiency, changes in personnel mix, standardization, specialization, methods improvements, better use of equipment, changes in the resource mix, product and service redesign and shared best practices.

For the example pursued in this article, we can think of these possibilities in terms of less labour hours per use or less expensive labour hours per use, and possibly no labour per use. Fewer hours are achievable via individual learning. Less expensive hours are achievable by, for example, substituting assistant physicians or nurse practitioners for physicians. In this case, the experts can be used as orchestrators of cadres of much less expensive clinicians.

Labour elimination is often technology enabled. For example, web-based scheduling and account management can enable patients to substitute their labour for that of providers, as has been experienced in the airline, banking and retail industries. Customers often find this much more satisfactory than dealing with multi-level phone systems typical for many health providers. For example, patients with chronic diseases such as diabetes are likely to benefit from online tools for self-management of their diseases. The results could be better care, greater patient satisfaction and lower costs.

The Commonwealth Fund has recently published their recommendations for ‘bending the curve’ (Schoen *et al.*, 2008). Based on extensive economic analyses, they propose the following:

1. **Producing and Using Better Information**
   1.1. Promoting Health Information Technology
   1.2. Center for Medical Effectiveness & Health Care Decision Making
   1.3. Patient Shared Decision Making
2. **Promoting Health and Disease Prevention**
   2.1. Public Health: Reducing Tobacco
   2.2. Public Health: Reducing Obesity
   2.3. Positive Incentives for Health
3. **Aligning Incentives with Quality and Efficiency**
   3.1. Hospital Pay-for-Performance
   3.2. Episode-of-Care Payment
3.3. Strengthening Primary Care and Care Coordination
3.4. Limit Federal Tax Exemptions for Premium Contributions

4. Correcting Price Signals in the Health Care Market

4.1. Reset Benchmark Rates for Medicare Advantage Plans
4.2. Competitive Bidding
4.3. Negotiated Prescription Drug Prices
4.4. All-Payer Provider Payment Methods and Rate
4.5. Limit Payment Rate Updates in High-Cost Areas

The report ‘Bending the Curve’ provides analytic projections of the savings that could be realized by adopting these recommendations. In the context of the models discussed here, seven of these recommendations (1.2, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2) would tend to reduce use rates. Nine of these recommendations (1.1, 1.2, 3.3, 3.4, 4.1–4.5) focus on reducing costs, either by increased efficiency (1.1, 1.2, 3.3) or by increased market-based competition (3.4, 4.1–4.5). The efficiency-based recommendation may be central to providers being able to respond successfully to increased competitive forces.

Note that reduced use rates decrease total costs, but also result in less process learning and hence higher relative costs per use. On the other hand, we might characterize increased wellness as system learning rather than process learning. Thus, for example, as people learn to either not start smoking or stop smoking, the demands on the processes associated with lung cancer decrease, resulting in this process not becoming as efficient as it could have been where there more demand placed on it.

Argyris and Schon (1978) term this phenomena ‘double loop learning’. Single loop learning involves getting better and better at something, as reflected in the second and third models presented earlier. Double loop learning, in contrast, involves questioning whether the thing you are trying to get better at you should be doing at all. Thus, in this example, rather than trying to get very good at treating lung cancer, we should find ways to eliminate it. We have to keep in mind, however, that single loop learning may be our only choice in circumstances where we cannot figure out how to accomplish a double loop change.

CONCLUSIONS

This article has illustrated an engineering approach to addressing the complex problem of escalating healthcare costs. Ironically, this has been done in the context of an engineering phenomenon, namely, successful technology innovation leading to growing markets and increased revenues. The problem in healthcare is that increasing revenues to innovators translate into increasing costs to payers. Such growth is viewed more favourably when individuals pay, rather than third parties.

It may be possible to devise market-based mechanisms to control the growth in demand. De facto rationing is also likely, although we do not like to talk about use of this mechanism. The other primary mechanism, which was the primary focus of this article, is increasing system efficiency to lower supply costs and hence prices. Such efficiency is needed to assure the afford-
ability of technology innovation. While the required improvements are very substantial, the estimates of their magnitude provided here offers some guidance for how aggressive efficiency initiatives need to be.

In searching for such huge efficiencies, it is important to focus on the whole system (Rouse, 2008). Consider the architecture of healthcare delivery shown in Figure 6. The efficiencies that can be gained at the lowest level (clinical practices) are limited by the nature of the next level (delivery operations). For example, functionally organized practices are much less efficient than delivery organized around processes.

Similarly, the efficiencies that can be gained in operations are limited by the level above (system structure). Functional operations are driven by organizations structured around specialties, e.g., radiology. And, of course, efficiencies in system structure are limited by the healthcare ecosystem in which organizations operate. Differing experiences of integrated health systems in the US as well as in other countries provide ample evidence of this.

The fee-for-service model central to healthcare in the US assures that provider income is linked to activities rather than outcomes. The focus on disease and restoration of health rather than wellness and productivity assures that healthcare expenditures will be viewed as costs rather than investments. I suspect that recasting of ‘the problem’ in terms of outcomes characterized by wellness and productivity may enable identification and pursuit of efficiencies that we cannot imagine within our current frame of reference.

Recasting the problem may also lead us to conclude that increased healthcare expenditures – and wellness expenditures – are justifiable, indeed desirable, if we can capture the productivity gains possible with a healthier, albeit ageing, population (Hall and Jones, 2007; Rouse, 2008). There is some evidence that lost productivity following medical absences costs employers more than direct outlays for healthcare (Collins et al., 2005). As is often the case for complex systems, the best approaches to their management is highly influenced by the nature of the objectives adopted.

ACKNOWLEDGEMENTS

This article has immeasurably benefited from comments and suggestions by Rahul Basole, Douglas Bodner, Kenneth Boff, Denis Cortese, Robert Smoldt and Stephen Schoenbaum.

REFERENCES


