Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

OUTCOMES AS OUTPUT: A NEW APPROACH TO THE PRODUCTION FUNCTION FOR MEDICAL SERVICES

Eric Nauenberg

Working Paper No: 150013

www.canadiancentreforhealtheconomics.ca

August 5, 2015

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 150013
August 5, 2015

Outcomes As Output: A New Approach to The Production Function for Medical Services

Eric Nauenberg[a,b,c]

[a] *Institute of Health Policy, Management and Evaluation, University of Toronto*
[b] *Canadian Centre for Health Economics, Toronto, Canada*
[c] *Toronto Health Economics and Technology Assessment Collaborative, Toronto, Canada*

JEL Classification: D24; I11; I12
Keywords: productivity; potential Pareto improvement

This paper redefines the output of medical care services in terms of outcomes and suggests some policy implications stemming from optimization of resources. In the past, output of health care has been specified as an institutional measure such as number of visits or admissions. In fact, these are really quasi-output measures in that they are inputs to the true output or outcome of providing medical care. The model developed extends standard economic theories of production to respecify a production function for medical care services that expands on the work of Michael Grossman (1972) and others.[1] It further satisfies a need set out by Marchildon and Di Matteo (2015) which state.

> ...“decision-makers now want to bend the health-care cost curve in a way that will not block access or damage quality. Instead, they want permanent and persistent efficiencies, which will require that they...provide more appropriate but lower-cost services, substitute providers where possible, and rein in provider remuneration...”[2]

The model set out is particularly timely in that it is estimated that between 30% and 50% of health care expenditures are unnecessary to achieve current health care outcomes.[3]

In standard economic theories, two inputs, capital and labor, are used to produce final outputs. In health care, outputs are not as clearly defined as in the manufacturing sector. Since healthy individuals are the ultimate goal of a medical care system, the best measure of output would therefore be a measure of well-being or outcome at different points in time. Futhermore, the inputs used in producing a healthy individual are more complex than in standard theories of production: the output in health care requires the active participation of the individual in his/her own production of health stock to produce a tripartite relationship with labor and capital. That is, in addition to capital and materials (pharmaceutical agents, technology, etc.), and labor (medical care professionals), the party on whose behalf the medical care complex is working must cooperate to some extent

[1] Grossman, Michael. ”On the concept of health capital and the demand for health.” The journal of political economy 80; 1972: 223-255.

[2] Marchildon G, Di Matteo L. Bending the Cost Curve in Health Care. Toronto: Univ. of Toronto Press. 2015.

[3] Cutler DM. The Simple Economics of Health Reform. Economists’ Voice, The Economists’ Voice 7.5; 2010: 1-5.

for optimal results to occur. This input, often referred to as level of patient compliance, has been noted in the literature as a significant factor in patient care especially in the medicare population.[4] Further, the relationship between these three inputs may not be linear and independent given the power of providers/labour in many aspects of decision-making with regard to treatment protocols regarding both at the intensive margin (e.g., the amount of treatment delivered within a given visit) and the extensive margin (e.g., further visits and diffusion of new technology/decommissioning of obsolete technology). The essence of the power vested in providers has been definitively measured since Rice's (1983) initial work on supplier-induced demand continuing on until the present day.[5,6] Given the concentration of decision-making vested in health care providers, a natural question arises as to whether productive efficiency can experience potential improvements-similar to those theorized by Kaldor and Hicks in the realm of allocative efficiency-through proper financial incentives and redistribution of resources.

Whereas previous models have incorporated the tradeoff between surgery and drug therapies in such areas as cardiac care, there has been no previous effort to include the tradeoff with patient effort/adherence which can also be influenced by providers through reminder systems, and increasingly so, in the era of smart-phones.[7,8] This model completes a triad incorporating the costs and benefits of each of three inputs into a final product or health care outcome illustrated in the diagram below.

As outlined in a technical appendix, the model seeks to minimize the cost of care given a goal

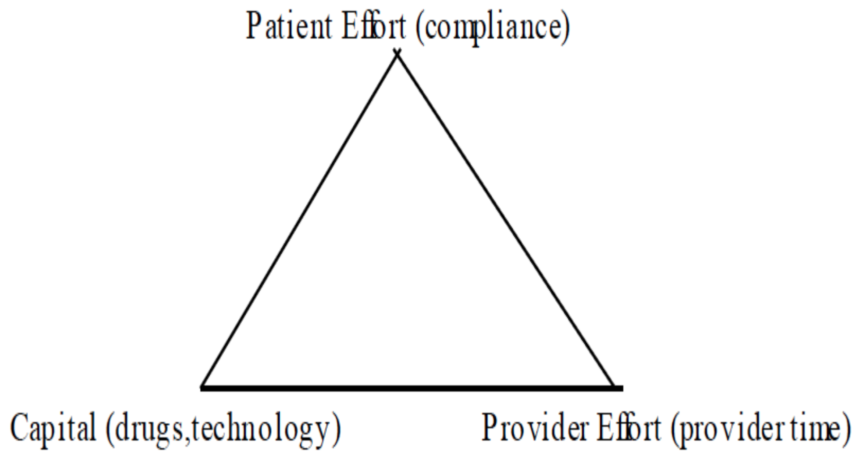
[4] Maronde RF; Chan LS; Larsen FJ; Strandberg LR; Laventurier MF; Sullivan SR. Underutilization of antihypertensive drugs and associated hospitalization. Medical Care, 12; 1989: 1159-66.

[5] Rice, TH. "The impact of changing Medicare reimbursement rates on physician-induced demand." Medical care 21; 1983: 803-815.

[6] Clemens J, Gottlieb JD.. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?." The American economic review 104; 2014: 1320-1332.

[7] Vervloet M, Linn AJ, van Weert JC et al. "The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature." Journal of the American Medical Informatics Association 19; 2012: 696-704.

[8] Dayer L, Heldenbrand S, Anderson P.. "Smartphone medication adherence apps: potential benefits to patients and providers." Journal of the American Pharmacists Association: 53;2013: 172-181.



of returning patients to a specific level of functioning or outcome at a particular point in time. The level of outcome is affected by the amount of each of the three inputs applied to an individual's medical care. As well, each of the three major inputs are each affected by a number of factors. With regard to provider effort, possible factors include the number of ancillary personnel available to assist in a case and the provider reimbursement structure. Regarding patient compliance, possible factors affecting effort include the amount of out-of-pocket expenses, patient health status, and socioeconomic factors (e.g., culture and income).

Furthermore, neoclassical economic theory has posited that productive efficiency is a parallel concept to that of allocative efficiency. With regard to the latter, it is well known that efficiency is attained either when (1) trading between parties does not improve the lot of at least one of them without worsening the lot of the others or (2) when somebody who becomes worse off as the result of a forced trade is compensated for their loss to maintain their well-being ceteris paribus. Likewise, neoclassical economic theory posits that productive efficiency is attained when it is not possible to further substitute between factor inputs (i..e, capital and labour) in order to either increase output holding costs constant or lower costs holding production constant. This inherent desire for parallel structures in neoclassical economic theory, however, misses an important point with regard to productive efficiency: It is often not the tradeoff between capital and labour that is often key to maximizing production but rather the interaction between capital and labour that is more important. The argument is that neoclassical theory should be reformulated to reflect realities not only in the health care sector but in all sectors whereby improving productive efficiency often hinges on either changing the way that labour works with capital or by labour choosing to decrease the use of various forms of capital altogether.

Neoclassically, Capital and labour are conceptually connected to total costs (TC) or expenditures through a linear cost function:

$$TC = wK + rL$$

in which w and r represents wages (unit price of labour) and capital rents (unit price of capital) In neoclassical economic theory, the objective would be to minimize costs subject to maintaining output or the dual problem of maximizing output subject to a cost constraint. The assumption that factor inputs are linear in their impact upon costs is under challenge herein. Specifically, a more realistic cost function would be formulated as follows:

$$TC = wK + rL + f(K, L)$$

In this reformulation, there is an extra term that reflects a potential non-linear interaction between labour and capital that is pivotal in determining costs (See Technical Appendix).

The policy implications of such a model are varied. Primary among these is the ability to determine the optimal mix of resources to achieve stated goals and how best to achieve this optimal mix. The model will determine the optimal contribution of each of pharmaceuticals and technology, physician time, and level of patient compliance in order to minimize the cost of care. Further, the model will be suggestive of how physicians and other providers can be properly incented to be cost-effective with regard to the use of capital and prescription drugs while optimizing their efforts alongside those of their patients. While at certain times, each of the inputs can be complimentary, such as when technology assists a physician in performing diagnostic tasks, at other times, each of these inputs serve as substitutes for each other. One example of the latter is the choice of whether to treat a cardiac patient with anti-hypertensive drugs, with surgical procedures, or with lifestyle changes. Different combinations of these three choices are warranted in different situations. Therefore, one possible use of the model is to determine the conditions under which different combinations are most cost-effective.

Another possible use of the model is to determine optimal pricing strategies and incentives to achieve stated goals. Given the dependence of both provider and patient effort on financial incentives, it might be possible to determine whether out-of-pocket expenses and provider reimbursements are determined so as to encourage optimal behavior. It is certainly possible that current incentive structures produce suboptimal levels of effort.[9] The key to producing potential improvements in productive efficiency is ϵ_2 (provider effort) since the bulk of decision making with regard to resource allocation reside with health care providers. As a result, proper financial incentives are needed to both neutralize supplier-induced demand at both the intensive and extensive margins as well as to motivate patients to achieve maximal compliance with treatment regimens. Given Cutler's (2010) finding of vast wastage of health care resources, K is hypothesized to decrease relative to ϵ_1 and ϵ_2 under a scheme of potential Pareto improvements in productive efficiency. In

[9] Wedig, GJ Ramsey Pricing and Supply-Side Incentives in physician markets. Journal of Health Economics. 1; 1993; 365-384.

this regard, Nauenberg (2014) has shown that further reducing the level of costs without impacting outcomes (functional health) requires containing the growth of technology, but also increasing the use of prescription medications.[10]. If so, the natural question is how can this shift be operationalized? A potential is to use gain-sharing as proposed by Gaynor, Rebbitzer and Taylor (2004) with an additional feature.[11] Instead of funds shared solely between the provider and payer alone, some portion of the savings should be redistributed to programs with established high value (e.g., childhood vaccination programs). In this way, a potential Pareto improvement in productive efficiency may be possible as resources are redistributed between different health care services to produce higher overall productivity. The most recent evidence regarding the performance of U.S. Accountable Care Organizations– able to retain some of the savings they create through reductions in areas like hospital-acquired infections and readmission rates - indicate that 24% of 220 ACOs produced savings of $700 million nationally of which approximately half was retained by the ACOs.[12] Some of these savings were then reinvested by some ACOs in housing and other social programs noted to have a large impact on the health of populations serviced by these organizations.[12] With respect to gain-sharing, some portion of provider and/or healthcare organization remuneration could be withheld in some type of fund-holding scheme. These funds would then be released contingent upon providers and/or health care organizations:

- selectively decommission expensive, ineffective technologies,
- limit the use of discretionary technologies, and/or
- employ cost-saving technologies/Px Drugs
- Increase patient compliance/adherence through reminder systems

[10] Nauenberg E. Changing Healthcare Capital-To-Labor Ratios: Evidence and Implications for Bending the Cost Curve in Canada and Beyond. Int J Health Care Finance Econ. 14; 2014: 339-353.

[11] Gaynor,M., Rebitzer, J., Taylor, L. (2004). Physician incentives in health maintenance organizations. Journal of Political Economy, 112; 2004: 915-931.

[12] Abrams M, Nuzum R, and Zezza M, et al. Realizing Health Reform's Potential. The Commonwealth Fund. 2015

Further research could be conducted on a variety of outcome measures. Several of these indices, such as the activities of daily living index, perceived quality of life indices, and Center for Epidemiologic Studies-Depression score, are suitable options for establishing a scale to measure outcomes. The next step for this project is to find data to test the assumptions of the model. This would require a data set with measures of outcomes or health status taken at various points in time. Secondly, data would be needed on pharmaceuticals administered, procedures performed, provider charges, levels of patient compliance, and type of insurance coverage. Lastly, data on patient and provider characteristics are needed for model specifications.

Technical Appendix

Outcomes are specified as a function of capital (K) and effort, ϵ_1 and ϵ_2, exerted by both patients and providers. In turn, patient effort (i.e., level of compliance) is function of sociodemographic variables, financial incentives (e.g., deductibles and copayments), difficulty of treatment regimen, and health status. Provider effort is a function of hours worked, intensity (i.e., throughput), reimbursement structures (e.g., capitation or fee-for-service), and level of expertise. The objective is to minimize costs constrained by the goal of attaining a certain level of outcome.

Suppose cost is a function of capital, patient effort and provider effort, C(K,ϵ_1,ϵ_2). The problem is to minimize C subject to the production constraint, θ(K,ϵ_1,ϵ_2):

$$Min\ \ C = rK + t_1\epsilon_1 + t_2\epsilon_2 + f(K, \epsilon_1, \epsilon_2)$$

$$s.t.\ \ \varnothing = K^{\alpha}\epsilon_1^{\beta}\epsilon_2^{\gamma} + g(K, \epsilon_1, \epsilon_2)$$

Where: r=capital rents and Px drug costs, t_1=patient time, out-of-pocket costs and transaction costs, t_2=provider(s) time/fees received, transaction costs, ϵ_1 = h(TD,HS,SD,I), ϵ_2 = m(L,In,R,Ex), Ex = *level of provider expertise*, HS = *patient health status*, I = patient financial incentives (i.e., copayments & deductibles), In = level of work intensity (i.e., throughput), L = man-hours provided, R = provider reimbursement structure, SD = sociodemographic variables, and TD = treatment difficulty level.

Some common assumptions are made about the cost and production functions: the cost function is continous, nondecreasing in factor prices and homogeneous of degree one. Production technology is monotonic, convex, regular and continuous. For purposes of evaluation, a Cobb-Douglas production function ($\alpha + \beta + \gamma = 1$) is adopted as well as a cost function where cost is partially a linear function of independent inputs as well as a nonlinear interdependency between them..

To simplify the analysis, the calculation here is limited to main effects (i.e., only K, ϵ_1 and ϵ_2 as explanatory variables) and initially only examines linear relationships. The resulting first-order approximation follows:

The Lagrangian:

$$L = rK + t_1\epsilon_1 + t_2\epsilon_2 + f(K, \epsilon_1, \epsilon_2) + \lambda(\varnothing - K^{\alpha}\epsilon_1^{\beta}\epsilon_2^{\gamma} - g(K, \epsilon_1, \epsilon_2))$$

First-Order Conditions

$$\begin{aligned}
\frac{\partial L}{\partial K} =& r + \frac{\partial(f(K, \epsilon_1, \epsilon_2)}{\partial K} - \lambda\left(aK^{\alpha-1}\epsilon_1^{\beta}\epsilon_2^{\gamma} + \frac{\partial(g(K, \epsilon_1, \epsilon_2)}{\partial K}\right) = 0 \\
\frac{\partial L}{\partial \epsilon_1} =& t_1 + \frac{\partial(f(K, \epsilon_1, \epsilon_2)}{\partial \epsilon_1} - \lambda\left(BK^{\alpha}\epsilon_1^{\beta-1}\epsilon_2^{\gamma} + \frac{\partial(g(K, \epsilon_1, \epsilon_2)}{\partial \epsilon_1}\right) = 0 \\
\frac{\partial L}{\partial \epsilon_2} =& t_2 + \frac{\partial(f(K, \epsilon_1, \epsilon_2)}{\partial \epsilon_2} - \lambda\left(\gamma K^{\alpha}\epsilon_1^{\beta}\epsilon_2^{\gamma-1} + \frac{\partial(g(K, \epsilon_1, \epsilon_2)}{\partial \epsilon_2}\right) = 0 \\
\frac{\partial L}{\partial \lambda} =& \varnothing - K^{\alpha}\epsilon_1^{\beta}\epsilon_2^{\gamma} - g(K, \epsilon_1, \epsilon_2) = 0
\end{aligned}$$

Solving for the unknowns (ignoring, for the moment, interaction terms f(K,ϵ_1,ϵ_2) and g(K,ϵ_1, ϵ_2)):

$$\epsilon_1 = \frac{\varnothing(\beta t_2)^{\gamma}}{\gamma^{\gamma}t_1^{(1-\beta)}}, \quad \epsilon_2 = \frac{\varnothing\gamma^{(\alpha+\beta)}t_1^{\beta}}{(\beta t_2)^{(\alpha+\beta)}}, \quad K = \frac{\alpha\varnothing t_1^{\beta}t_2^{\gamma}}{\beta^{(1-\gamma)}\gamma^{\gamma}r}, \quad \lambda = \frac{t_1^{\beta}t_2^{\gamma}}{\beta^{(\alpha+\beta)}\gamma^{\gamma}}$$

Assuming diminishing returns, it is also possible to assume that the Hessian is positive semi-definite; therefore, the second-order conditions for a minimum are also satisfied. This specification of the problem makes possible a number of observations regarding the outcome function. First, existence can be proven as well as the uniqueness of an equilibrium level of capital, patient effort and provider effort. Second, the Lagrangian multiplier, λ, makes it possible to consider marginal costs at the optimum in terms of pairs of factor prices. Third, the first order conditions make possible the calculation of a solution for the equilibrium optimal level of effort and capital in terms of factor prices. For example, optimal patient effort is a function of the price of the provider's effort and the inverse of the price of patient effort. The optimal provider effort is a function of the price of patient effort and an inverse of price of provider effort. Cross-price effects, own-price effects and movement of changes in factor demands relative to changes in factor prices can also be analyzed through this model. Finally, the paper investigates the implication of the cost/production function duality.

When investigating the effects with the added impact of the nonlinear interdependencies on output/outcomes, the mix of capital (K) and effort, ϵ_1 and ϵ may shift. If $\frac{\partial(f(K,\epsilon_1,\epsilon_2))}{\partial K}$, $\frac{\partial(f(K,\epsilon_1,\epsilon_2))}{\partial \epsilon_1}$ and $\frac{\partial(f(K,\epsilon_1,\epsilon_2))}{\partial \epsilon_2}$ are all > 0 with $\frac{\partial(g(K,\epsilon_1,\epsilon_2))}{\partial K} \leq 0$, and both $\frac{\partial(g(K,\epsilon_1,\epsilon_2))}{\partial \epsilon_1}$, $\frac{\partial(g(K,\epsilon_1,\epsilon_2))}{\partial \epsilon_2} > 0$ then it is conceivable that potential improvements can be made in productive efficiency beyond what would occur voluntarily - ceteris paribus.