Modelling Shared Efficiency between Diagnostic and Non-Diagnostic Revenue-Producing Hospital Cost Centers

Dan Friesner *, Matthew Q. McPherson

A game-theoretic model of hospital production is developed which allows for sequential and, if appropriate, simultaneous production across departments and/or cost centers within a hospital. Decisions made in two units: one whose work is primarily completed by physicians, and one whose work is completed by diagnostic technicians (which may be a laboratory or an imaging department), and whose efforts may fundamentally alter the types of output produced in the other department. This model predicts that inefficiency in the diagnostic unit impacts inefficiency in the physicians’ unit, but inefficiency within the physician’s unit does not impact inefficiency in the diagnostic unit.

Keywords: efficiency, sequential production, simultaneous production, games, hospital theory

Introduction

The production of hospital services is unique in that, when admitted for care, a patient receives a bundle of interrelated, jointly produced goods and services. For example, a single patient who undergoes hip replacement surgery may receive services – in addition to the surgery itself - from physical therapy, diagnostic imaging, pharmacy, housekeeping and dietary services. Some services, (i.e., housekeeping and dietary services) are distinct from the surgery. In such cases, the production of surgical, dietary, and housekeeping services can be modelled analogously to multiple-output producers in other industries.

The production of other medical services, including (but not limited to) diagnostic services provided by imaging and medical laboratory departments, are more challenging to conceptualize within a multiple-output production process. The provision of diagnostic services are unique services, for which hospitals may bill patients and insurers. However, clinicians in other units (especially physicians) may use diagnostic services when making diagnoses, or determining the appropriateness of additional medical services provided to a patient. Thus, the production of diagnostic services, in part or whole, defines the service that is provided (or the manner the service is provided) in other units (Leaven, 2015; Friesner et al., 2019).

Few studies in the economics or management science/operations research (MS/OR) literatures address the efficiency of hospital production at the level of the hospital department. The economic and MS/OR literatures on efficiency analysis in health care includes only 13 hospital efficiency studies that used something other than the hospital as the unit of observation (Hollingsworth, 2003). Few, if any, of these 13 studies examined departments providing diagnostic services. More recent reviews of the economics and MS/OR efficiency analysis literatures indicate that the hospital continues to be the most common unit of observation (Stefko et al., 2018 and Leleu et al., 2018).

Additional studies, especially those of a theoretic nature, are warranted to conceptualize efficiency-related decisions and predict behavior at the level of the department, rather than the hospital as a whole. Because such studies generate hypotheses that must be empirically evaluated at the level of the department, the development of a theoretic model of efficient production within or across hospital departments provides an opportunity to reframe the economic and MS/OR literatures to a unit of observation that is consistent with the context in which efficient decisions are actually made.

The environment in which hospital care is delivered is unique and worthy of examination. To this end, and consistent with the economic and MS/OR literatures, a game-theoretic model of hospital production that describes the optimal strategies of various hospital personnel who work in different departments is presented (Marco, 2001; Rosenman and Friesner, 2004). This model allows for sequential and, if appropriate, simultaneous production across departments or cost centers within a hospital. The model examines decisions made in two units: one whose work is primarily completed (or determined) by physicians and other providers, and one whose work is primarily completed by diagnostic technicians (which may be a laboratory or an imaging department), and whose efforts may fundamentally alter the types of output produced in the other department. The predictions developed in the modelling process provide direction for future empirical research assessing the efficiency of production within and across hospital departments.

Literature Review

The economic and MS/OR literatures can give insights in developing a game-theoretic model of efficient, department-level hospital production. A general summary of the economic and MS/OR efficiency analysis literature across all applications and industries shows the wide-spread and diverse use of efficiency analysis (Emrouznejad and Yang, 2018). Focusing specifically on healthcare, the vast majority of the literature is empirical in nature, and focuses on the hospital as the unit of observation (Hollingsworth, 2003 and 2008; Stefko et al., 2018; and Leleu et al. 2018).

Diagnostic hospital departments

Within the context of diagnostic hospital departments, the literature is both expansive and sparse. There are literally thousands of published studies that examine the impact of specific protocols, practices, or interventions to improve the efficiency or effectiveness of specific aspects of the department’s production. For example, Analytical Hierarchy Process (AHP) methodology has been used to determine which specific areas of medical laboratory operations are most likely to impact the efficiency of the entire department’s production (Leaven,
The most efficient means to schedule phlebotomists; one type of medical laboratory employee, has also been examined (Leaven and Qu, 2014). Additionally, regression analysis has been used to predict the amount of tests performed by a laboratory (Leaven, 2016). None of these studies assess the efficiency of a department as a unique decision-making unit, nor do they examine the implications of departmental efficiency (again, using the entire department as the unit of observation) on the hospital as a whole or other departments within a hospital.

The literature examining efficiency using the diagnostic hospital department as the unit of observation is sparse and primarily empirical in nature (Leaven, 2015). The impact of budgetary allocations to the medical laboratory department on the hospital as a whole has been examined.

Specific types of budgetary allocations to the department impact different aspects of the hospital’s financial standing (profitability, capital structure, etc.) (Friesner et al., 2019). Additionally, evidence suggests hospitals which implemented health information technology tracking systems within their medical laboratory operations achieved improved hospital financial performance (Zhao et al., 2019). Further, panel data methods have been used to examine the intensity of computed tomography (CT) and magnetic resonance imaging (MRI) use and overall cost efficiency in a panel of Chinese hospitals. A positive relationship between MRI and CT use and overall cost inefficiency is found (Wei et al., 2018). While an important first step, these studies say little about the allocation of resources involved in the provision (and possible overuse) of diagnostic imaging services, and their impact on efficiency in other hospital units.

An empirical methodology that identifies the amount of productive inefficiency that is shared between hospital departments has been developed (Murphy et al., 2011). However, this methodology does not characterize the means by which inefficiency is shared. This is important, especially if inefficiency in diagnostic units mitigate or exacerbate inefficiency in other departments.

**Physician Behavior**

The economic and MS/OR literatures using game theory to model physician behavior primarily focus on decisions related to the patient-physician relationship and care decisions, rather than decisions related to efficient hospital production. For example, a game-theoretic model of decision-making in the context of care-related decisions for patients with chronic liver failure has been developed (Diamond et al., 1986). The model differs from traditional (unilateral) decision analysis because the actual outcome depends on the choices of both the physician directing the most appropriate treatment path and the decisions of the patient in response to that prescribed treatment path. Game theory has been used to examine decision-making and choices in the context of brain death (Riggs 2004). The prisoner’s dilemma framework was used to illustrate the potential relevance of using game theory to inform the clinical practice of medicine and medical ethics. In the context of acute stroke care, evidence suggests game theory may improve the understanding of complex medical situations and help clinicians make practical decisions under uncertainty (Saposnik and Johnston, 2016). Additionally, an investigation to examine the dilemma faced by physicians in prescribing antibiotics illustrates the usefulness of game-theoretic approach. Analysis of equilibrium strategies, evolutionarily stability, and replicator dynamics show rational doctors, motivated to attain the best outcomes for their own patients, will prescribe antibiotics irrespective of the level of antibiotic resistance in the population and the behavior of other doctors. This is the case even though physicians would achieve better long-term outcomes if they were more restrained in prescribing antibiotics (Colman et al., 2019).

Other studies focus on modeling the dynamic environment in which health care is provided, including physician behavior. A controlled experiment analyzing the impact of professional norms on prospective physicians’ tradeoffs between their profits, the patients’ benefits, and the payers’ expenses for medical care suggests that professional norms derived from the Hippocratic tradition shift weight to the patient in physicians’ decisions while decreasing their self-interest and efficiency concerns (Kesternich et al., 2015). Game theory has also been used to provide a representation of the interactions that typically take place in the operating room environment. The prisoner’s dilemma game (and variations) can be used to illustrate common interactions in the operating room between staff (physicians and nurses) and stakeholders (administration) (Marco, 2001).

A crucial gap in the literature is a theoretical basis for the understanding of how inefficiency might be shared across different departments within a hospital. This requires a focus on the department as the unit of observation. Since production processes may be sequential or simultaneous in nature, the consideration of strategic decision-making across departments is necessary. This approach allows for inefficiency to be shared across departments.

**Conceptual framework: a model of hospital production**

A simple theoretical model of hospital production is presented which attempts to fill this gap in our understanding of how inefficiency might be shared across departments within a hospital. For simplicity, the model proceeds directly from the economic and MS/OR efficiency analysis literatures rather than from theories (for example, resource dependence and natural selection) drawn from the organizational and operations management literatures (Hurley and Kaluzny, 1987; Alexander and Wells, 2008; Yeager et al., 2014). While the latter are valid, reliable, and informative, they are less amenable to game theoretic modelling which seeks to generate very precise predictions. Rather, these theories are valuable in interpreting the parameters of the model and using those interpretations to guide subsequent empirical research.

Aside from Marco (2001), few studies have assessed collaborative efforts by different health care personnel to provide efficient care. Marco’s study focuses primarily on simultaneous decision-making by various clinicians. Such a model is inappropriate for the interaction between physicians and diagnostic laboratory or imaging personnel, as the use of diagnostic information impacts physician decisions. Our model allows for sequential and, if appropriate, simultaneous production across departments. Production within one unit (which produces diagnostic services) may fundamentally alter the types of output produced in another department. The model provides a theoretical basis for future empirical studies to assess shared inefficiency in hospitals.

**The model**

For simplicity, consider a simplified version of health care production that involves two types of clinicians: “providers” and “diagnostic staff”. The former include physicians, nurse practitioners, and physician assistants. The latter include technical personnel whose services are billed directly to patients or third parties. Providers and diagnostic staff are assumed to provide ethically appropriate, patient-centered care; however providers and diagnostic staff exercise some degree of power in determining resources involved in providing care to patients.

When a typical patient seeks care, the provider has an initial patient consultation, inclusive of physical assessment, patient history, and
diagnostic orders. Let $D$ be a variable that defines the intensity of resources used in this endeavor, and let $P^T$ be the (real) input price for $D$. The patient is referred to diagnostic staff for care, who provide diagnostic services using $L$ units of service at a (real) price of $P^T$ per unit. Lastly, the provider uses that information to provide $T$ units of treatment at a (real) input price $P_T$. This sequential approach allows both types of personnel to define the quantity of services provided as well as to help define the services provided. For simplicity, the production of patient care ($Y$) is assumed to be linear in inputs:

$$Y = a + bT + cD + fL$$

(1)

where $a$, $b$, and $c$, and $f$ are parameters (all non-negative in value) and the remaining variables are as previously defined.

Each type of staff operates within a cost center or department, and must adhere to a budget constraint. Assume that all providers act collectively, or there is a unit administrator who makes decisions concerning evidence-based practice and budgetary decisions. Let the provider’s budget constraint be given by:

$$B = PrT + PoD + M$$

(2)

where $B$ is the overall provider budget, $M$ is the amount of “inefficient slack” or excess non-pecuniary expenditures made by providers (Friesner and Rosenman, 2002), and the remaining variables are as previously defined. The diagnostic staff has an analogous approach to decision making and an analogous budget:

$$\beta = P^T_L + S$$

(3)

where $\beta$ is the department or cost center budget and $S$ is the amount of inefficient spending on the part of the diagnostic staff.

Each type of clinician has an exogenously defined “target” level of patient care and inefficiency, respectively (Rosenman and Friesner, 2004). These targets are agnostic to the intended level of inefficiency or profit generated by the hospital, and represent a very flexible means to address a wide array of objectives. Note that if the clinician(s) in question cares about efficiency, the target level of inefficiency is zero. If the clinicians attempt to maximize hospital profit, the targets are jointly set at the profit-maximizing level of patient care, and zero, respectively. Each type of clinician attempts to minimize a weighted average of squared deviations from these target values, subject to operating within their budget. For providers, the objective is:

$$\text{minimize}_{T,D} \ A = w(Y - Y^*)^2 + (1 - w)(M - M^*)^2$$

subject to $B - P_T^0T - P_D^0D - M = 0$  

(4a)

where $w$ is a proper proportion ($0 \leq w \leq 1$). Substituting (1) into (4a) yields:

$$\text{minimize}_{T,D} \ A = w(a + bT + cD + fL - Y^*)^2$$

+ $(1 - w)(M - M^*)^2$

subject to $B - P_T^0T - P_D^0D - M = 0$  

(4b)

The objective for diagnostic personnel is:

$$\text{minimize}_L \Delta = \theta(Y - Y^*)^2 + (1 - \theta)(S - S^*)^2$$

subject to $\beta - P_L^0L - S = 0$  

(5a)

where $\theta$ is a proper proportion ($0 \leq \theta \leq 1$). Substituting (1) into (5a) yields:

$$\text{minimize}_L \Delta = \theta(a + bT + cD + fL - Y^*)^2 + (1 - \theta)(S - S^*)^2$$

subject to $\beta - P_L^0L - S = 0$  

(5b)

**Sequential, three-stage decision making**

The game has three stages. In the first stage, providers choose $D$. In the second stage, diagnostic personnel produce $L$, and in the final stage, providers put forth resources ($T$) to treat patients. The game is solved via backward induction using the Mathematica Version 11.2 software package. To solve stage 3, the constraint in (4b) is solved for $M$ and substituted into the objective function. Taking the partial derivative of this function with respect to $T$ yields:

$$\frac{\partial U}{\partial T} = 2bw(a + bT + cD + fL - Y^*) - 2P_T^0(1 - w)(B - P_T^0T - P_D^0D - M^*) = 0$$

(6)

Note that the second order condition to ensure a minimum holds. Equation (6) yields a solution for $T$:

$$T = \frac{(1-w)Pr(Pr^2(Pr^2b^2+Pr^2-\theta \beta^2(Pr^2+\beta^2))}{b^2Pr^2(b^2+Pr^2(Pr^2+\beta^2))}$$

(7)

In the second stage of the game, the solution for $T$ is substituted into the diagnostic personnel’s objective function. The constraint in (5b) is also solved for $S$ and substituted into the objective function. Taking the first order condition for $V$ with respect to $L$: 

$$\frac{\partial V}{\partial L} = -2P_T^0(\beta - P_L^0L - S - \theta)(1 - \theta) + 2b((1 - w)P_T^0(B - P_D^0D - M^*) - \theta(b(a + cD + fL - Y^*)) + (Pr^2(1 - w) + b^2w)^2)$$

(8)

Note that the second order condition is unambiguously positive, ensuring that the optimum is a minimum. Solving (8) yields:

$$L = \frac{(1 - \theta)P_T^0(\beta - S)^2(Pr^2(1 - w)b^2w^2 + Pr^2(Pr^2+\beta^2))}{(1 - \theta)P_T^0((1 - w)b^2w^2 + \beta^2w^2)^2}$$

(9)

In the final stage of the game, the solutions for $T$ and $L$ are substituted into the provider’s objective function. The budget constraint in (4b) is solved for $M$ and substituted into the objective function, yielding an objective function with a single choice variable, $D$. The first order condition is:

$$\frac{\partial U}{\partial D} = \frac{2b^2P_T^0(b - cM)^2}{P_T^0(b - cM)^2}$$

(10)

This yields a solution for $D$:

$$D = \frac{bP_T^0(\beta - cM) + P_T^0(\theta a + \theta cM)}{P_T^0(\theta b + cM)}$$

(11)

Substituting (11) into (9):

$$L = \frac{\beta - S}{\theta \beta^2}$$

(12)

Working recursively, the optimal solution for $T$ is:

$$T = \frac{P_T^0(b - cM) + (\theta a + \theta cM)}{P_T^0(\theta b + cM)}$$

(13)

Substitution indicates that the optimal value for $Y$ is $Y^*$, and that the optimal values for both the provider and the diagnostic personnel are zero; that is $U = V = 0$.

The implications of the optimal solutions are straightforward. Greater slack from either clinician (or both clinicians) decreases the
quantity of patient services delivered by providers. This is true at the
time of diagnosis (D) and at treatment (T). Inefficiency on the part of
diagnostic personnel only impacts the inputs/effort of diagnostic
personnel, not provider services. Thus, the model provides a very
simple means to empirically test for sequential, shared efficiency.
Provider inefficiency should be uncorrelated with the quantity of
diagnostic inputs provided; however, diagnostic inefficiency should be
correlated with the quantity of provider inputs provided.

Two-stage decision making

The previous model assumes that providers, when
determining D, make decisions prior to laboratory personnel choosing
L. There are certain situations (transplant units, oncology units, etc.)
where L and D may be determined simultaneously. To account for this
possibility, the previous game is altered to allow only two stages. In
stage 1, providers and diagnostic personnel choose D and L
simultaneously. In the second period, the providers choose T. Solving
the game via backward induction, the providers’ second stage
decisions continue to be characterized by equations (6) and (7).

In the first stage of the game, both types of providers
(simultaneously) attempt to optimize their objective functions, given
their respective budget constraints, as well as the solution for T given
in (7). Substituting both constraints into the objective functions and
taking first order conditions yields:

\[ \frac{\partial L}{\partial \delta} = -2(\beta P_L - cP_L) (1 - w) w (P_L (a + fL + Y + cD) + b (R - P_D - M + s)) \]

\[ \frac{P_L^2 (1 - w) w}{(P_L (1 - w) w)^2} = 0 \] (14)

\[ \frac{\partial V}{\partial L} = -2P_L (\beta - P_L - L \ast (1 - \theta)) + 2f \beta (1 - w) w (P_L (a + fL + Y + cD) + b (R - P_D - M + s)) \]

\[ \frac{(P_L (1 - w) w)^2}{(P_L (1 - w) w)^2} = 0 \] (15)

Equations (14) and (15) can be used to identify Nash equilibrium
pure strategies:

\[ D = P_L (R - M + s) + P_L (aP_L - fS - P_L Y + \beta f) \]

\[ L = \frac{R - S + \theta}{P_L} \] (16)

\[ L = \frac{R - S + \theta}{P_L} \] (17)

Clearly, the solutions to both problems are identical, so no first or
second mover advantages exist in the game. All shared inefficiency,
whether it occurs simultaneously or sequentially, has the same impact
on the model, and on the provision of patient care in particular. Thus,
the testable hypotheses regarding shared inefficiency do not depend on
the timing of diagnostic services. This is a crucial implication since
specific types of patient care are more appropriately modelled by the
two stage game (i.e., care in organ transplant units) or the three stage
game (i.e. rural, primary care).

Discussion

This model of hospital production presents a number of interesting
deductions and predictions. First, shared inefficiency is a one-way
phenomenon. Inefficiency in the diagnostic department impacts
provider inefficiency, but inefficiency by providers fails to impact
diagnostic department inefficiency. Diagnostic departments
strategically adjust their practices to mitigate the effects of provider inefficiency.

Second, as long as treatment is provided to patients after their initial
physician consultation and diagnostic work has been completed, the
nature of shared inefficiency is unaffected by whether the physicians
perform their patient consultations simultaneously with, or prior to, the
work of the diagnostic unit. Thus, shared efficiency appears to be
embedded in hospital production regardless of the timing of decision-
making.

The most interesting aspect of these predictions is that they are
generated from a model based on economic and MS/OR rationale.
However, the model’s predictions are consistent with other, alternative
theories of efficient hospital production, especially resource
dependence theory (RDT). Under RDT, physicians must rely on
resources of resources external to themselves (and their department) to
produce their own patient care services (Yeager, et al., 2015). The
production of physician services is an open system, and thus impacted
by changes in the availability, pricing and other sources of uncertainty
associated with securing critical resources, of which diagnostic
information is one. Within the context of our model, diagnostic
department personnel act intentionally and strategically to ensure that
their work is an external resource to physicians. This allows
inefficiency to be embedded within overall hospital production
through the concept of shared efficiency. The implications of this
inference are also twofold. First, the model provides an argument
illustrating the importance of empirical hospital efficiency analyses
conducted at the level of the department, rather than the hospital.
Second, the model presents an argument illustrating the
complementarity of different models of hospital production.
Embedding principles of RDT into game theoretic models of hospital
productions may add additional (more accurate and precise) inferences
than what is contained in this model, and may be a useful extension of
the current work.

The model also allows for the appropriate characterization of shared
inefficiency which can be assessed by estimating a system of factor
demand equations based on (11)-(13). An obvious null hypothesis is
that the coefficient estimate characterizing the marginal impact of the
target level of physician slack on diagnostic input usage (L) is zero.
Another null hypothesis is that the coefficient estimate characterizing
the marginal impact of the target level of diagnostic unit slack on
physician input usage (D and/or L) is significantly different from zero.
In practice, many hospitals and health systems are decentralizing their
diagnostic and support services. Hospitals typically maintain a central
medical laboratory for non-urgent or standard testing procedures. However,
medical laboratory technicians and technologists may also
staff smaller laboratories housed directly in departments that provide
highly specialized types of care, to handle diagnostic information
needs specific to that unit. In such cases, if the null hypotheses
described above are rejected, a likely explanation is that diagnostic
services have been decentralized and incorporated into this unit.

Limitations and Future Research

While this model provides some interesting predictions, the
model’s structure is limited. The production of patient care is specified
as linear in variables and parameters, and may not accurately depict
the production of complex or acute patient care. The model also
requires a prior knowledge of the unit manager’s target levels for
inefficient spending. Lastly, the model assumes only two units or cost
centers. In reality, hospitals are complex environments with scores of
different units, each of which has slightly different productive
activities and incentives.

By characterizing efficiency via managerial slack/inefficient
spending, this model characterizes a specific type of inefficiency.
Other forms of inefficiency, such as technical inefficiency, are
unexplored. Given these considerations, this model is a first step
towards a comprehensive understanding of shared inefficiency in
hospital production.

One meaningful extension of the current work is using theories such
as RDT to provide a more specific, measurable, and clear interpretation of the
model’s parameters. This would allow model predictions which
exhibit greater precision and could be empirically tested in a more accurate and precise fashion.

References