ON THE STAFFING POLICY AND TECHNOLOGY INVESTMENT IN A SPECIALTY HOSPITAL OFFERING TELEMEDICINE

Hakan Tarakci* Melbourne Business School University of Melbourne Carlton, Victoria, 3053, Australia h.tarakci@mbs.edu

Zafer Ozdemir Farmer School of Business Miami University Oxford, OH, 45056, USA ozdemir@muohio.edu

Moosa Sharafali Lee Kong Chian School of Business Singapore Management University 50 Stamford Road Singapore 178899 sharafalim@smu.edu.sg

* Corresponding Author: Hakan Tarakci
Tel: +61 3 9349 8434

Full Postal Address: Melbourne Business School 200 Leicester Street Carlton, Victoria, 3053 Australia

1 The authors thank Alican Dalkilic, MD, MPH from Virginia Commonwealth University, for reviewing and editing the paper. The authors also gratefully acknowledge the many comments and suggestions from all the anonymous referees and the Editor, which very greatly improved the presentation of the paper.
ABSTRACT

We study a specialty hospital providing traditional face-to-face consultations by experts and telemedicine services by tele-specialists. As accuracy of diagnosis and treatment by tele-specialists are paramount in such a setting (unlike call center management), our main focus is to determine the optimal investment level in telemedicine technology with the trade off being between accuracy/quality and cost. Using a heuristic proposed in queuing theory, we provide the optimal investment in telemedicine technology together with the staffing policy, considering the various cost components, including staffing, technology investment, incorrect treatment, and waiting. The model also incorporates buy-in by the patients in the form of the arrival (show-up) rate dependent on the technology level established. We find that under certain conditions the hospital should not invest in telemedicine. Finally, we provide the optimal tele-specialist policy of the ratio of patients to treat via telemedicine and to refer to the face-to-face consultation. Our model also suggests that a policy of treating all patients via tele-medicine is never optimal.

Keywords: Telemedicine, e-health, information technology investment, queuing system
1. Introduction

Telemedicine is the delivery and provision of healthcare and consultative services to patients over distance using telecommunications technologies. New technologies continue to evolve over time, including real-time videoconferencing as well as store and forward technologies. Recent advances in information and communication technologies present enormous potential to offer telemedicine services throughout the world, expand access to primary, secondary, and tertiary levels of care, improve quality, increase efficiency, and decrease costs [2, 4, 29]. By providing greater access to medical expertise, telemedicine can reduce the geographical variability of diagnosis, treatment, and clinical management [19]. In particular, tele-consultations have been shown to change diagnoses and management recommendations, and also to reduce the long waiting times associated with access to high-demand specialty care [22]. A recent report by the Institute of Medicine, *Crossing the Quality Chasm*, states that “patients should receive care whenever they need it and in many forms, not just face-to-face visits…access to care should be provided over the Internet, by telephone, and by other means” [8, p. 61].

Telemedicine is the most promising alternative to the traditional mode of health care delivery. Patients in rural areas and in prisons as well as those that are home-bound are at a significant disadvantage in accessing traditional health care services. In addition, finding the right expertise in a certain geographical location is difficult at best for many regions of the country. However, this problem can be solved by employing the telemedicine applications embraced by the medical profession.

For example, some experts in the United States provide online second opinions for important diagnoses and treatments. The Cleveland Clinic, a well-known multi-specialty academic medical center in Cleveland, has established its online interface, the eClevelandClinic, to serve patients who need advice and possibly a major intervention, but who cannot easily access their doctors in person. The electronic service is limited to life-threatening and life-altering conditions that can be safely assessed online, such as a new cancer diagnosis, cardiac procedures, joint replacements, and neurological problems. Patients provide a personal medical history and the original diagnosis as well
as other relevant materials such as test results, MRI, films, x-rays, and a consent form. Three Harvard University teaching hospitals have initiated a similar service at [www.econsults.partners.org](http://www.econsults.partners.org). In addition, several leading oncologists render medical opinions at [www.mdexpert.com](http://www.mdexpert.com) and charge fees as high as $3,200. Experts at these renowned institutions perform their own examinations and review of tests, establish diagnoses and develop treatment plans. Patients typically follow treatments under the supervision of their local health care providers. A study reported that these experts agreed with the original diagnosis most of the time, but very frequently altered the original treatment plan [22].

In addition to the delivery of treatment plans that need to be carried out with the help of local providers, telemedicine is also used to provide care to remote patients needing specialist care. For example, when unable to recruit the sought-after intensive care specialists, St. Mary’s Health Center in Jefferson City, Missouri, decided to implement a telemedicine program where intensivists and critical care nurses could remotely care for patients at St. Mary’s and other community hospitals from an operations center in St. Louis. With the inception of the program, St. Mary’s has seen patient mortality drop by 24 percent [33]. Remote surgery is also becoming popular. Cone et al. [9] have used telemedicine to assist, monitor, and assess a surgery performed in Ecuador by a doctor in Virginia. A problem with the patient’s anesthesia was noticed and corrected by the doctor, who was assisting and monitoring via telemedicine. In another instance, a NSF meteorologist spending the winter in 2002 at Amundsen-Scott South Pole Station in Antarctica injured his knee and subsequently needed a surgical operation. Doctors were available at the U.S.-operated, year-round station in Antarctica, but the medical facilities there were not designed for surgery. For the first time in the program's nearly 50-year history, telemedicine has been used for surgery as doctors in Massachusetts helped a physician to surgically repair the damaged knee of the meteorologist.² These examples illustrate how medical experts in remote locations can help local providers deliver the best care to patients that they wouldn’t be able to do otherwise. Furthermore, remote robotics surgery is expected to allow tele-specialists to directly work on a patient in not so distant future. NASA has been experimenting with surgical robots to be used in medical procedures in space [27].

Other than enabling the timely delivery of specialty services to remote locations, telemedicine can also be cost advantageous. In the report titled “Telehomecare and Remote Monitoring: An Outcomes Review,” Stachura and co-authors cited a number of sources, which indicated the use of telemedicine technology can reduce costs. Stachura calculated the cost of an office visit at about $100 versus that of a “tele-visit” between $15 and $40.\(^3\) According to another study reported in the September/October 1997 Remington Report, telemonitoring of congestive heart failure patients “saved approximately $8,000 per patient because telehomecare reduced the need for conventional clinic visits” [31]. Further, telemedicine also yield to earlier discharge of patients from the hospital, thereby reducing the hospitalization cost, risks associated with hospitalization, and strain on hospitals needing beds. Thus, telemedicine presents policy makers and practitioners new opportunities, and warrants systematic investigation.

Despite the opportunity to improve quality of and access to health care, and decrease costs via telemedicine, there is little research that analyzes its economic viability with respect to traditional health care approaches using sound methodologies. Instead, research on telemedicine is mainly conducted to answer the question “Can we do this?”, while leaving the important question “Should we do this?” unanswered [22]. Cost information about telemedicine applications are very preliminary and are often concerned with making a case to proceed further [14]. Several economic studies indicate avoidance of travel or patient transfer as one of the most salient benefits of telemedicine, but most of these studies have methodological limitations, and the generalizability of findings is rather limited [15, 16]. Consequently, evaluating the effectiveness of telemedicine services is difficult at best.

Even when cost-effectiveness of telemedicine is evident, relevant frameworks that aid in defining the appropriate scope and application of telemedicine in different settings do not exist. In other words, the important question of “How should we do this?” is still unanswered. This is an important limitation because the adoption of telemedicine involves making decisions on complicated matters.

To address these issues, the entity taken up for discussion in this paper is a specialty hospital, which considers offering health care to remote locations via telemedicine. In such a situation, it should first decide on how much to invest in technology while taking into account concerns about both quality and cost. Furthermore, it should employ the appropriate number of qualified personnel who will serve via its face-to-face and telemedicine channels. Referrals among experts that serve face-to-face and via telemedicine should also be coordinated. While governments increasingly are trying to urge public hospitals to provide remote healthcare in order to ease the pressure on the limited resources, the response from the hospitals and public in general is lukewarm. Unless these important decisions are made with adequate understanding of the related issues, telemedicine programs and projects may fail to be implemented successfully. Hence, the research questions to be addressed in this paper which confront the above specialty hospital are:

(i) Should the hospital consider providing telemedicine services?

(ii) What should be the investment level in telemedicine technology and infrastructure?

(iii) How many tele-specialists and experts should the hospital employ?

We use a queuing framework to answer these questions. It should be pointed out that modeling the provision of telemedicine services as a queuing network is akin to running a call center because both types of services involve a diagnosis process as well as a treatment versus referral decision. However, the crucial differences between the two are: (i) medical experts need more time in solving problems due to their complexity while problems that can be solved by calling a call center are more trivial and take less time and attention than health problems that require licensed clinicians, multiple people, and specialty care, ii) a call center is usually an outsourced entity while it is not usually so with telemedicine services and (iii) the stakes for wrong diagnosis and/or treatment are very high in healthcare including telemedicine. Thus, investment in appropriate and up-to-date technology is paramount especially in telemedicine. Nevertheless, a search of the call center literature (see Aksin et al. [1] for example) reveals quite surprisingly no research on investment decisions. Hence, the main objective of this paper is to provide a decision framework and insights on when and how a specialty hospital should adopt telemedicine. Viewed from a social perspective, our
investigation also helps policy makers to determine when and how telemedicine can reduce total costs within the health care system. Our main contributions to the literature are the following:

1. We determine to what extent the hospital should invest in telemedicine technology to improve the delivery of tele-specialist services. To this end, the model we have formulated below is more general and realistic than hitherto considered in the literature. While consideration of investment in technology is new, we have also modeled the buy-in concept of telemedicine by patients by assuming the arrival (show-up) rate at the tele-specialist appointments to depend on the technology level available. This is based on the findings of Hu et al. [20, 21] that perceived service risks as well as attitudes toward technology are important determinants of telemedicine adoption.

2. By taking into account the various cost factors, we find the optimal staffing policy for the hospital (i.e., the number of face-to-face traditional experts and tele-specialists that need to be employed). We find the optimal referral policy for the tele-specialists (i.e., when they should attempt to treat a patient via telemedicine after the initial diagnosis and when they should invite the patient to the hospital). Given the cost of incorrect treatment, a model of the tele-specialist’s effectiveness, and the relative costs of service via the two channels, it is possible to identify a referral policy that is optimal for the system as a whole.

The rest of the paper is organized as follows: We review the literature in Section 2, we then set up the model in Section 3, and in Section 4 provides a sensitivity analysis that demonstrates the effects of various parameters on the optimal policy of staffing and technology investment. The discussion of the results, limitations, and future work is provided in Section 5.

2. Literature Review

The backbone to our study is the modeling of the patient flow through the system (see Figure 1) which is similar to what is considered in [18]. Naturally, the methodology we use in analyzing such a flow is queuing theory. Queuing theory is a predominant tool used not only to study systems

---

4 Technically, we find the heuristic solution pair to the approximate cost function given in (10). However, as [3] and [18] demonstrate, the heuristic solution is a very good approximation to the exact optimal solution which can only be found through numerical methods. Hence, for the sake of brevity, we adopt the use of “optimal” solutions as opposed to “heuristic solutions to the approximate cost function”
in manufacturing (see Buzacott and Shanthikumar [5]) and call center management but also healthcare management. For an excellent, exhaustive and very recent review on call center management, we refer the reader to Aksin et al. [1] and for a similar review on healthcare management we refer the readers to Fomundam and Herrmann [11].

Despite the documented benefits of telemedicine, such as reducing patients’ travel, work, and family expenses [7] and improving diagnoses and treatment plans [6], there is limited decision support systems research on managing the adoption and implementation of this particular modality of healthcare. Hence, our aim is to contribute to the knowledge on managing telemedicine implementations from the perspective of a specialist hospital.

In achieving this objective, our paper closely follows a methodology similar to that of [18], which optimizes the total costs of a two-tier system by finding the cost-minimizing referral rate as a function of arrival and service rates. Hasija et al. [18] show that as the arrival rate increases, the optimal referral rate for the stochastic case converges to the optimal referral rate for the deterministic case. Another finding is that if the gatekeepers’ (the first tier in the system, akin to tele-specialists in our model) skill level is not high, a direct access system is always the preferred option. In our paper, we assume that the tele-specialists and experts have comparable skill levels and that the discrepancy between their successful treatment rates is mostly determined by the level of investment in tele-medicine technology. This investment level in technology is analogous to the skill level in [18] with the difference that the investment level in our paper is a decision variable whereas the skill level is an exogenous variable in [18]. So, the natural questions that arise are (i) what technology level should the gatekeepers (tele-specialists) be able to utilize, and (ii) is it worth investing in? Unless using the appropriate level of technology, tele-specialists may misdiagnose and treat patients incorrectly, potentially leading to suffering, complications, and litigation costs. Hence, it is very important for the hospital to invest in the latest technology to diagnose the patients’ conditions, and treat them, when necessary. Obviously this investment in technology requires capital. But, the literature on healthcare management is mainly focused on capacity management in the form of determining the staffing levels and/or the number of beds required etc (see Fomundam and Herrmann [11]). As pointed out above, management of telemedicine is similar to call center management. But, a good review of the call
center management research (see Aksin et al. [1]) for works on investment decisions on technology shows that no papers have been written on this topic to our knowledge. We therefore introduce technology investment costs to the model of [18], and analyze whether a telemedicine system consisting of the traditional experts and tele-specialists is preferred over the usual direct access system given the cost of technology. In modeling the investment cost function, we draw from similar ideas used in research in manufacturing. For example, the usefulness of set-up time reduction is well documented in the literature. Porteus [30] was the first to analyze benefits and costs of setup reduction. This requires investment in set-up time reduction and to model this aspect researchers (see Liu and Cetinkaya [25] and Leschke and Weiss [24]) have proposed exponential functions. We also differ from [18] in modeling the arrival (show-up) rate at the tele-specialists.

Methodologically, this paper builds on the literature on capacity planning in queuing networks. Halfin and Whitt [17] identify and compare heavy-traffic approximations for stochastic limits of multiple-server queuing systems. Whitt [34] proposes a simple square-root staffing principle for the same problem, and Borst et al. [3] show the asymptotic optimality of this principle using a cost function involving both waiting and staffing costs. We apply the square-root staffing rule to a two-tier queuing network to find the optimal number of tele-specialists and experts. Taking the level of investment in telemedicine technology and referral policy of tele-specialists as the decision variables, we then determine the total cost of operating such a system and minimize that cost.

Another related area of research is the design of gatekeeper mechanisms for multi-tier service systems. In health care, primary care physicians (as in the U.S.) and general practitioners (as in Britain) act as gatekeepers of specialist services. The relationship between the existence of gatekeepers, their financial incentives, and the level of use of specialist services has received much attention. Ferris et al. [10] challenge the importance of gatekeeping, while Gaynor et al. [12] empirically demonstrate that incentives to primary care physicians do influence patients’ utilization of resources. Given the importance of incentives, Malcolmson [26] investigates the optimal contractual mechanisms for gatekeepers, and Shumsky and Pinker [32] examine the optimal referral policy for a deterministic two-tiered system and the incentive issues that affect gatekeepers’ performance. We
adopt the treatment function used both in [18] and [32] in our analysis (as discussed in the next section) without explicitly considering any incentive problems or contractual issues.

Finally, Ozdemir et al. [a] analyzes the optimal online consultation strategies of experts, while Walczak et al. [b] investigates the performance of a decision support tool in allocating hospital bed resources. These two studies are relevant because the former focuses on using information technologies to deliver health care from a distance and the latter investigates the role of an information system in optimally allocating hospital resources.

3. The Model

Typically, health care systems around the globe exhibit a two-tiered structure where primary care physicians serve as gatekeepers to medical specialists. In countries such as Australia, Norway, and the United Kingdom, general practitioners act as gatekeepers for specialty services. Almost all managed care plans in the United States use gatekeeper arrangements, either by requiring a referral from a specified primary care physician before consulting a specialist [13] or by fully insuring the provision of care only if it is supplied, or authorized, by the responsible health maintenance organization [26]. The tiered provision of health care may take many forms and is implemented primarily to better utilize the medical experts’ time and other resources in the hospital system, as well as preventing non-emergency patients from clogging the system. Hence, in the modeling and the analysis to follow, we take into account this type of tiered provision of health care services.

Specifically, we consider a specialty hospital to which referrals have been made by the General Practitioners (GPs). Therefore, our decision analysis space includes only the treatment of the patients by the specialty hospital rather than whether or not these patients should have been referred to the specialty hospital in the first place. We make this assumption to work with a two-tier system rather than a three-tier one and hence to simplify the analysis. However, it should be noted that the methodology used in this paper can be extended to treat the referral rate by the GPs as a decision variable. We assume that the specialty hospital may employ a pool of experts who provide traditional face-to-face consultations and a pool of tele-specialists who have comparable skills and provide
medical services remotely using the available technological means. Figure 1 illustrates the patients’ flow through the system which forms the basis of our queuing network model.

Patients are always seen first by the General Practitioners (GP). In the model, we focus only on the complicated cases where the GP needs to refer the patient to a specialist. The patients either call the tele-specialist or request a personal appointment with one of the expert doctors in the specialty hospital. We assume that the proportion of patients seeking tele-specialists (denoted by \( k \)) depends on the level of investment in telemedicine technology (denoted by \( b \)). In other words, the hospital can improve its tele-specialists’ ability to successfully diagnose and treat remote patients through costly investment in technology. We normalize \( b \) to a range of 0 to 1, where 0 represents no telemedicine resources and no tele-specialists and 1 represents, for example, a state-of-the-art medical video conferencing system, such as the one described in LeRouge et al. [23].

Patients remotely contacting the specialty hospital wait for an available tele-specialist, who first tries to diagnose the ailment. If remote treatment is possible then she does so; otherwise, the patient is immediately referred to the expert pool to be physically seen by an expert at the hospital. Unfortunately, medical errors happen regardless of the mode of visit. According to the Institute of Medicine, medication errors harm at least 1.5 million people every year, causing an extra medical cost of $3.5 billion a year to treat these drug-related injuries occurring in hospitals\(^5\). Therefore, we assume that remote interactions with the tele-specialists may end unsuccessfully, in which case the system incurs a cost for incorrect treatment. We further assume that experts, who treat the more difficult cases, can also commit costly errors. The incorrect treatment by the tele-specialists is modeled as in [18]. We assume that patients contacting the tele-specialists have ailments at complexity/difficulty level \( X \), which is a random variable following a uniform distribution on the unit interval [0, 1]. Higher complexity levels imply more difficult cases that require higher investment in technology. Given the complexity level \( x \), we adopt the treatment function \( f(x) \), as defined by [18, 32], to be the

probability that a tele-specialist successfully treats the patient. We also assume infinite waiting space at both the tele-specialists and the expert doctors. Specifically, we use the notation given in Table 1.

\[ \text{Table 1} \]

Note that the technology investment cost function \( C(b) \) is assumed to be convexly increasing in the level of investment in telemedicine technology \( b \); hence, \( C'(b) \) and \( C''(b) \) both >0. Note also that the proportion of patients selecting tele-specialists over in-person consultation with experts is given by the function \( k(b) \). The function \( k(b) \) has the following properties:

(i) For \( b = 0 \), that is, when there is no investment in tele-specialist technology, no patient would select tele-specialists \( (k(0) = 1) \).

(ii) As the investment level \( b \) increases, patients' confidence in the tele-specialists also increases and thus the proportion increases as well. However, this increase in proportion occurs at a decreasing returns to scale, implying \( k'(b) > 0 \) and \( k''(b) < 0 \).

(iii) Even at the maximum level of investment (i.e., when \( b = 1 \)), there would still be patients who would not want to go through tele-specialists; in other words, \( k(1) < 1 \).

These assumptions are realistic based on the observations of Hu et al. [20, 21].

The level of investment in telemedicine technology determines diagnosis and treatment rates of tele-specialists. The maximum rates of diagnosis and treatment for the tele-specialists are attained when \( b = 1 \). Accordingly, we have \( \mu_d(b) = \mu_d b \) and \( \mu_t(b) = \mu_t b \), where \( \mu_d \) and \( \mu_t \) are the respective theoretical maximum rates. To maintain consistency, we assume all the cost parameters are given in cost per unit time.

Clearly, the telemedicine system as defined above is a queuing network. We assume the overall arrival rate has a Poisson distribution. While the service time distribution is assumed to be exponentially distributed for the expert doctors, it is not so for the tele-specialists. A tele-specialist’s service time is exponentially distributed when she diagnoses a patient and immediately refers the patient to the experts, and is the sum of two different exponential distributions when she decides to treat the patient following the diagnosis. So, effectively the service time distribution of a tele-
specialist is a double exponential distribution (i.e., a mixture of exponential distributions). Similar to [18], we use two approximations to analyze the queuing network described above.

(i) We assume that the effective service rate of a tele-specialist is approximately exponentially distributed with the effective rate given by

\[
\Pi(b) = \left( \frac{1-r}{\mu_d(b)} + \frac{r}{\mu_i(b)} \right)^{-1}.
\]

(ii) The approximation (i) above makes the subsystems of the tele-specialist and expert doctor pools to be \( M/M/n \) queuing systems where the arrival rates into these subsystems are as shown in Figure 1. One of the decisions to be made is the staffing levels of tele-specialists and expert doctors. To find these optimal staffing levels, we use the heuristic suggested by [3]. For a brief description of the heuristic we refer the reader to [18] which shows the closeness of the heuristic solution to the exact optimal solution. Below we present the relevant formulas.

For a \( M/M/n \) queuing system with offered load \( \rho = \lambda / \mu \), [3] shows the asymptotically optimal staffing level to be

\[
\hat{n} = \rho + y^*(c, w)\sqrt{\rho},
\]

where \( c \) and \( w \) are the unit costs of staffing and the unit cost of waiting. The above expression has the interpretation that \( \rho \) is the minimum staffing level required, and the second term is the safety staffing level due to the uncertainty in the system. Here, \( y^*(c, w) \) is the minimizer of the function

\[
\alpha(y, c, w) = cy + \frac{w\pi(y)}{y},
\]

where

\[
\pi(y) = \left[ 1 + \frac{y\Phi(y)}{\phi(y)} \right]^{-1},
\]

and \( \phi(y) \) and \( \Phi(y) \) are the probability distribution function and the cumulative distribution function of the standard normal distribution, respectively. Thus,
\[ y^*(c, w) = \arg \min_{y > 0} \alpha(y, c, w). \]  

(5)

It is known (see [18]) that \( \alpha(y, c, w) \) is unimodal, and hence \( y^*(c, w) \) can be determined using a numerical procedure. Further, the approximate total cost of staffing and waiting is known to be

\[ TC^{app}(\rho, c, w) = c\rho + \alpha(y^*(c, w), c, w)\sqrt{\rho}. \]  

(6)

In the sequel, we will use expression (6) in our total cost calculations.

We are now ready to derive the total cost function. The system now consists of two \( M/M/n \) queuing systems. Based on queuing theory, we have

\[ TC_1(n_1, b, r) = \lambda k(b) w_1 W(n_1, \lambda k(b), \mu(b)) + c_1 n_1, \]  

(7)

and the combined total cost function is

\[ TC = TC_1(n_1, b, r) + TC_2(n_2, b, r) + m_1 \lambda k(b)[r - F(r)] + m_1 p \lambda [1 - k(b)F(r)] + C(b). \]  

(9)

If we use the approximations (2) and (6) in (9), the total cost function just becomes a function of \( b \) and \( r \) only. Thus, the total cost function (9) approximates to

\[ STC^{app}(b, r) = \begin{cases} 
TC^{app}(\rho_e, c_e, w_e) + m_e p \lambda, & \text{if } b = 0 \\
TC^{app}(\rho_e, c_e, w_e) + TC^{app}(\rho_e, c_e, w_e) + m_1 \lambda k[r - F(r)] + m_1 p \lambda [1 - k(b)F(r)] + C(b), & \text{o/w}
\end{cases} \]  

(10)

where

\[ \rho_e = \frac{\lambda k(b)}{\mu(b)}. \]  

(11)

and

\[ \rho_e = \frac{\lambda [1 - k(b)F(r)]}{\mu_e}. \]  

(12)

We note that the decision to not invest in telemedicine technology \( (b = 0) \) implies that the system will have no tele-specialists and will have only expert doctors to attend to the patients \( (r = 0) \). We next present our first proposition. All proofs are in the Appendix.
Proposition 1: The optimal level of investment in telemedicine technology and the optimal fraction of patients treated by tele-specialists, given by $b^*$ and $r^*$, respectively, are one of the following three combinations:

(i) $b^* = 0$ and $r^* = 0$, implying that no tele-specialists should be employed.

(ii) $r^*$ is the internal solution(s) satisfying the first order condition $\frac{\partial}{\partial r} STC^{app} (b, r) = 0$, implying that tele-specialists should be hired but experts are still needed ($r^* < 1$); $b^*$ is given by the solution(s) of $\frac{\partial}{\partial b} STC^{app} (b, r) = 0$, suggesting a positive level of investment in telemedicine, but not necessarily acquiring the most advanced technology available ($b^* < 1$).

or

(iii) $r^*$ is the internal solution(s) satisfying the first order condition $\frac{\partial}{\partial r} STC^{app} (b, r) = 0$ and $b^* = 1$, implying the optimality of acquiring the most advanced telemedicine technology available.

Proposition 1 characterizes the optimal solution, which is found by comparing the values of the total cost function at the extreme values for the decision variables along with any internal value that satisfies the first-order conditions. Further, we observe that the optimal level of investment is always less than the upper bound (i.e. $r^* < 1$), which implies that the tele-specialists never completely replace expert doctors. This is probably because the probability of successful treatment by tele-specialists, although positive for all cases, approaches zero as the complexity level increases, even with the maximum level of investment in telemedicine technology.

In the next section, we perform a sensitivity analysis and illustrate the above method through a numerical example. We also identify the conditions that call for introducing the telemedicine service.

4. Sensitivity Analysis
In analyzing the effect of the model parameters on the optimal decisions, we first provide the results in the form of comparative statics analysis and then discuss them with a numerical analysis.

4.1. Comparative Statics

We use the implicit function theorem to derive the theoretical effect of the model parameters on the optimal outcomes. This method gives us comparative statics results; i.e. the change in the optimal $b$ and $r$ values when one of the model parameters is altered. This analysis is very valuable in terms of determining whether more or less technology investment in telemedicine ($b$) is needed and/or how the optimal proportion of patients treated by tele-specialists ($r$) vary as a result of changes in various parameter values. Proposition 2 outlines the results of this analysis (obtained by applying the implicit function theorem to $STC^{\text{app}}$). Please note that we limited the comparative statics analysis to internal solutions only (i.e., for $0 < r < 1$ and $0 < b < 1$).

Proposition 2. The (internal) optimal tele-specialist treatment ratio ($r^*$) is increasing in the maximum diagnosis plus treatment rate of tele-specialists ($\partial r^*/\partial \mu_t > 0$), the waiting cost at experts ($\partial r^*/\partial w_e > 0$), the probability and the cost of incorrect treatment at experts ($\partial r^*/\partial p > 0$, $\partial r^*/\partial m_e > 0$) and the cost of staffing of experts ($\partial r^*/\partial c_e > 0$), while it is decreasing in the cost of staffing of tele-specialists ($\partial r^*/\partial c_s < 0$), the incorrect treatment cost at tele-specialists ($\partial r^*/\partial m_s < 0$), the maximum diagnosis rate of tele-specialists ($\partial r^*/\partial \mu_d < 0$), the treatment rate of experts ($\partial r^*/\partial \mu_e < 0$) and the waiting cost at tele-specialists($\partial r^*/\partial w_s < 0$). Finally, the relationship between $r^*$ and the arrival rate of patients is inconclusive since $\partial r^*/\partial \lambda$ has a complex, non-monotonous pattern.

Proposition 2 states that as $c_s$, $m_s$, $\mu_d$, $\mu_e$ and $w_e$ increase, the optimal fraction of patients to be treated by tele-specialists decreases. This implies that for large values of these parameters, it is optimal for tele-specialists to treat less difficult cases. On the other hand, as $c_s$, $m_s$, $p$, $\mu_e$ and $w_e$ increase, the optimal proportion of patients to be treated by tele-specialists increases as well. Hence, for large values of these parameters, tele-specialists should treat more difficult cases, which will result in a higher proportion of patients treated by tele-specialists and fewer patients referred to the experts.

The implicit assumption in Proposition 2 is that the optimal level of investment in telemedicine technology remains the same. Of course, except for the case of a corner solution, this
will not be true when the model parameters change. Therefore, we need to keep this in mind when interpreting Proposition 2. Nevertheless, combining the results of Propositions 2 and 3 (presented below) provides us with valuable insights into the health care provider’s optimal strategies.

**Proposition 3.** The (internal) optimal level of investment in telemedicine technology ($b^*$) is increasing in the cost of staffing of both tele-specialists and experts ($\frac{\partial b^*}{\partial c_s} > 0, \frac{\partial b^*}{\partial c_e} > 0$), the waiting cost at both tele-specialists and experts ($\frac{\partial b^*}{\partial w_s} > 0, \frac{\partial b^*}{\partial w_e} > 0$), the probability and cost of incorrect treatment at experts ($\frac{\partial b^*}{\partial p} > 0, \frac{\partial b^*}{\partial m_e} > 0$) and the arrival rate of patients ($\frac{\partial b^*}{\partial \lambda} > 0$), while it is decreasing in the cost of technology parameters ($\frac{\partial b^*}{\partial C_1} < 0, \frac{\partial b^*}{\partial C_2} < 0$) and the treatment rate of experts ($\frac{\partial b^*}{\partial \mu_e} < 0$). We also find that the optimal investment level is decreasing in the incorrect treatment cost at tele-specialists if the investment level is already low, but increasing if the level is already high ($\frac{\partial b^*}{\partial m_s} < 0$ for small $b^*$ and $> 0$ for large $b^*$). The optimal level of investment is increasing in maximum diagnosis and diagnosis plus treatment rates of tele-specialists at low technology levels, but decreasing at high levels ($\frac{\partial b^*}{\partial \mu_d}, \frac{\partial b^*}{\partial \mu_t}$ are both $> 0$ for small $b^*$ and $< 0$ for large $b^*$). Finally, the relationship between the optimal technology investment level and the ratio of patients choosing tele-specialists is inconclusive since $\frac{\partial b^*}{\partial K_1}$ has a complex, non-monotonous pattern.

Proposition 3 states that as $c_s, c_e, \lambda, w_s, w_e, p$ and $m_e$ increase, the optimal telemedicine technology investment increases, implying that for large values of these parameters, the hospital should invest more in telemedicine technology, possibly even in the most advanced technology available ($b^* = 1$). Note that an increase in these parameters would also increase total cost. Thus, the hospital mitigates this increase in total cost by investing in better technology. On the other hand, as $\mu_e, C_1$ and $C_2$ increase, the optimal $b$ decreases. Thus, for large values of these parameters, investment in the most advanced telemedicine technology is not economically justified ($b^* < 1$).

Combining the results of Propositions 1, 2 and 3, we find that for sufficiently high values of the cost of staffing of tele-specialists, the cost of incorrect treatment at tele-specialists, the treatment rate of experts, the cost of tele-specialist technology investment and/or the cost of waiting at tele-specialists, the hospital prefers to employ only experts (Proposition 1.i applies; $b^*, r^* = 0$). On the
other hand, for sufficiently high values of the treatment rate of tele-specialists, the cost of waiting at experts, the cost of staffing of experts, the probability of incorrect treatment at experts and/or the cost of incorrect treatment at experts, the hospital prefers to employ tele-specialists as well as experts but it does not invest in the most advanced technology (Proposition 1.ii applies; \(0 < b^*, r^* < 1\)). Finally, for even higher values of this last set of parameters (except for the treatment rate of tele-specialists), the hospital invests in the most advanced technology available (Proposition 1.iii applies; \(b^* = 1, r^* > 0\)).

4.2. Numerical Analysis

We now demonstrate numerically the effect of nine parameter values on the optimal number of experts and tele-specialists, as well as total cost. We selected the parameter values based on the numerical examples provided in [18]. For example, our expert treatment rate, tele-specialist diagnosis rate, tele-specialist unit staffing cost, tele-specialist waiting cost, tele-specialist incorrect treatment cost are identical to the ones provided in Figure 2 of [18]. We slightly changed other parameter values and introduced new parameters to take into account the differences in the two models (for example, our paper assumes technology investment is costly and a decision variable while Hasija et al. [18] take the skill level as a “free” exogenous variable) and to present a “non-trivial” baseline example (where \(0 < b^*, r^* < 1\)). For our baseline model parameters, we used the following values: \(\mu_e = 1, c_e = 10, w_e = 100, w_s = 5, c_s = 1, \lambda = 30, \mu_d = 5, \mu_t = 3.5, C_1 = 25, C_2 = 2, K_1 = 1.5, m_s = 1, m_e = 5, p = 0.1\). The results and detailed analyses are given below.

4.2.1. The tele-specialist staffing cost (\(c_s\))

Here, we use the model parameters listed above and varied \(c_s\) to observe the changes in optimal values and total costs. These results are presented in Figure 2. As expected from the results of Proposition 2, the (internal) optimal tele-specialist treatment ratio decreases with \(c_s\), implying that higher tele-specialist staffing costs mean more patients will be referred to experts. On the other hand, as suggested by Proposition 3, the (internal) optimal \(b^*\) increases in \(c_s\), implying that investing in more advanced and efficient technology becomes more important as the staffing costs rise. Note also that \(b^*\) and \(r^*\) remain internal solutions until \(c_s\) becomes too large for the hospital to afford tele-specialists. Beyond a certain threshold, it is optimal to employ only traditional face-to-face experts. Finally, as expected, the optimal number of tele-specialists decreases and the total cost increases with \(c_s\).
However, a large increase in $c_s$ minimally affects the total cost. For instance, even a fourfold increase in tele-specialist cost (say, from 0.5 to 2) increases the total cost by a mere 2.7%. This means that our model is robust against sizable changes in the staffing cost. Of course, once $c_s$ becomes large enough to justify only traditional experts, further increases in $c_s$ have no effect on the total cost or the number of tele-specialists employed.

4.2.2. The cost of waiting at the tele-specialists ($w_s$)

The optimal outcomes are presented in Figure 3. Consistent with Proposition 2, as waiting cost at the tele-specialist increases, the (internal) optimal tele-specialist treatment ratio decreases. Also, similar to the effect of tele-specialist staffing cost, the (internal) optimal level of investment in telemedicine technology increases with $w_s$, which is consistent with Proposition 3. We note that it is optimal to employ tele-specialists as long as the waiting cost is not significantly large. The effect of waiting cost on the total cost is minimal. For instance, it is still optimal to staff tele-specialists when waiting cost rises 1,400% (say, from 1 to 15) because even in this case the increase in total cost is only 2.5%. We believe this minimal effect of waiting cost on total cost is due to the fact that the waiting time per patient is kept to a small level. Of course, employing tele-specialists would no longer be optimal when waiting cost is prohibitively high.

4.2.3. The arrival rate ($\lambda$)

Here, we analyze the effect of the patients’ arrival rate to the system and present the optimal values in Figure 4. Initially, at low level of arrivals ($\lambda = 20$), the system can handle all patients without any need for tele-specialists ($r^* = 0$). As expected, we observe that as more and more patients arrive, the optimal number of experts increases. Once the arrival rate reaches a critical value (for instance, when $\lambda = 27$), the hospital finds it optimal to employ both experts and tele-specialists. Further increases in the patient arrival rate are accompanied by hiring more of both experts and tele-specialists. We note that the relationship between the (internal) optimal tele-specialist treatment ratio ($r$) and the arrival rate is not monotonous, unlike the relationship of the treatment ratio with the tele-specialist staffing cost ($c_s$) and the cost of waiting for the tele-specialist ($w_s$). On the other hand, the
optimal level of investment in telemedicine technology increases with the patient arrival rate. This result, suggested in Proposition 3, is not surprising since the arrival of more patients require faster treatment rates, which is directly related to the level of technology investment ($b$). The optimal investment reaches its maximum level beyond a threshold level of arrival rate. Finally, we note that an increase in the arrival rate has a significant effect on the total cost. For instance, a 100% jump in the arrival rate (say, from 20 to 40) increases total cost by around 55%. On the other hand, the service cost per patient decreases due to economies of scale.

Figure 4

4.2.4. Maximum tele-specialist diagnosis rate ($\mu_d$) and maximum tele-specialist diagnosis plus treatment rate ($\mu_t$).

Figures 5 and 6, respectively, present the optimal outcomes. We use the base model parameters for both variables; however, to be able to analyze the effects of $\mu_t$ on the optimal outcomes, we set $\mu_d$ to a large value (100,000) given $\mu_t < \mu_d$. We see that the optimal tele-specialist treatment ratio decreases in $\mu_d$ but increases in $\mu_t$, which is supported by Proposition 2. On the other hand, we observe that the optimal level of investment in telemedicine technology is decreasing in both parameters, which is in line with Proposition 3 (note that $b^*$ is “large” – it is close to 1). A final observation in both tables is that the change in either of these rates has a minimal effect on the optimal investment and referral values as well as on total cost. For instance, even when $\mu_d$ approaches infinity, total cost decreases by less than 1%! We observe a similar trend for $\mu_t$. These observations suggest that high tele-diagnosis and tele-treatment rates have very limited effect on the total cost. We suspect that this is because the total cost is dominated by staffing costs rather than waiting time costs.

Figures 5 and 6

4.2.5. The cost of telemedicine technology parameters ($C_1$ and $C_2$)

The results, presented in Figures 7 and 8, respectively, exhibit patterns similar to one another. In both cases, as these technology cost parameters increase, the optimal level of investment in telemedicine technology decreases, which is also in line with Proposition 3. Thus, as the cost of technology becomes dearer, it is optimal for the hospital to decrease its investment in it, possibly to the point where the hospital does not invest in telemedicine or hire any tele-specialists. The results
also suggest that as the cost of technology increases, the hospital initially needs to hire more tele-specialists and more experts while decreasing its investment in technology, in effect replacing costly technology with more staff. Of course, no tele-specialist is hired when the cost of technology is too high.

____________ Figures 7 and 8____________

4.2.6. The ratio of patients selecting tele-specialists ($K_1$)

Please recall that the proportion of patients initially selecting tele-specialists is given by $k(b) = 1 - e^{-K_1b}$, which is increasing in $K_1$. We observe that the optimal level of investment in telemedicine technology initially increases and then decreases with $K_1$. This suggests that as more patients initially choose the tele-specialist route, it becomes optimal to invest in better technology. However, as the proportion of tele-patients increases beyond a certain threshold, the technology investment needs to be reduced. This is because the benefit of economies of scale dampens the benefit of technology investment. The results for this section are presented in Figure 9.

____________ Figure 9____________

4.2.7. The cost of incorrect treatment at tele-specialists ($m_s$)

Figure 10 supports Propositions 2 and 3. In particular, we observe that the optimal tele-specialist treatment ratio ($r$) decreases with the incorrect treatment cost. We also see that the optimal level of investment in telemedicine technology increases with the cost of incorrect treatment when this cost takes relatively low values. These observations suggest that as the cost of an incorrect tele-treatment increases, more patients need to be referred to experts (higher “$1 - r$” value) for a face-to-face consultation. In addition, the investment in telemedicine technology should also increase if this cost is not prohibitive. Of course, if the incorrect treatment cost at tele-specialists is too large, the hospital should not invest in telemedicine technology and should hire only traditional experts. Finally, we observe that the effect of the increase in incorrect treatment cost on the total cost is not substantial; even an 8-fold increase in incorrect treatment cost (say, from 0.25 to 2) raises total cost by just 2.9%.

____________ Figure 10____________
Finally, we would like to point out that the “kinks” (the points of sharp increases/decreases) in the $b$ and $r$ curves in Figures 2-10 (except for Figure 5) indicate that the system cost with telemedicine ($b, r > 0$) is lower or higher than the system cost without telemedicine ($b, r = 0$), thus causing a switch into or out of a system with telemedicine. The sharp changes in the $b$ and $r$ values as we switch from a non-telemedicine to a telemedicine system, for instance, imply that if hospitals would like to invest in telemedicine, the investment level should be significant rather than marginal. We believe this is caused by two factors: i) the hospitals need to make a substantial investment in telemedicine technology to replace the more skilled experts with telemedicine specialists without loss of capacity and quality, and ii) investment in high levels of telemedicine technology allows the hospital to gain economies of scale.

5. Conclusion

Telemedicine is of great importance not just for technological, but also for economical reasons, especially at a time when healthcare costs have been increasing at a much faster rate than inflation, while some remote/rural areas in even developed countries lack good quality specialty healthcare services. Therefore, in this paper, we have studied the strategies of a specialty hospital regarding the provision of remote telemedicine (via tele-specialists) services. Specifically, we have used queuing theory to determine the number of experts and tele-specialists the hospital should employ. Our analysis reveals that tele-specialist services can never completely replace face-to-face services, even when we allow for medical errors in the physical encounter and invest in the most advanced telemedicine technology available. The results of our study suggest that unlike some other sectors in the economy, the remote channel will likely continue to be complementary and secondary to the traditional channel in health care.

The most important decision we have analyzed is the optimal investment level in telemedicine in a specialty hospital. Presumably, the hospital can invest in technology that will improve the diagnosis/treatment accuracy of tele-specialists. For instance, the hospital can invest in state-of-the-art videoconferencing technology and/or powerful medical equipment to provide better telemedicine services to patients. Of course, the more advanced the technology becomes, the more expensive it is
going to be. Therefore, we determined the optimal investment in technology level that provides the optimal balance between cost and quality of service.

The last major decision we considered is regarding the behavior of tele-specialists. When a tele-specialist evaluates a patient, he or she has two options to choose from: either attempt to treat the patient if it seems feasible or refer the patient to one of the experts for a face-to-face consultation. Therefore, we focus on the important question of “What percentage of patients should be treated by the tele-specialists and what percentage should be referred to the experts?” In this paper, we characterize this optimal referral rate, a decision that depends on the complexity of the disease and the level of investment in telemedicine technology.

There are certain limitations given the nature of this work. For example, in real life consumers may visit providers at settings different from specialty hospitals, some of which may offer their services through the Internet. Our model also does not incorporate strategic issues regarding patients’ choice of channel type because we took the perspective of a specialty hospital/social planner and aimed to understand the impact of referral policies and investment in telemedicine technologies on the total cost of the system as a whole. In addition, our simplified model also does not take into account all aspects of patient care (such as insurance coverage, the type and severity of disease(s), and past medical history) and of the provider (such as bed capacity and discharge rate). These simplifications are made to obtain tractable analytical results.

There are a number of avenues for future research. One immediate extension to the current model would be to include the General Practitioner (GP) as another decision maker, in effect creating a 3-tier process: the GP, the tele-specialist, and the expert. Another idea would be to include public benefits, costs and policies in our model. For example, the hospital could be required to provide a minimum medical service level to remote areas, in which case tele-specialists would be hired even if doing so were not economically feasible. A third avenue of exploration would be on revenue management models. Specifically, medical providers may choose to offer high-value services face-to-face while other services through telemedicine, and may differentiate their rates based on channel, which would impact usage rates different channels by patients. An investigation of these issues may provide additional insights about the overall adoption problem.
Appendix: Proofs of Propositions

Proof of Proposition 1. Similar to [18], the proof is based on usual calculus-based arguments by studying the behavior of first and second order conditions. We find that the second derivative of the cost function with respect to $r$ is not always positive nor negative. Further, we find that the first derivative with respect to $r$ evaluated at the extreme point $r = 1$, given below, is always positive, but it could be positive or negative evaluated at the extreme point $r = 0$. Combining these facts, we know that the optimal $r$ value cannot be 1; it must be either zero or the (internal) $r$ value(s) that satisfies the first order conditions (FOC). A similar analysis with respect to $b$ shows that the first derivative, given below, evaluated at the extreme points 0 and 1 can be both positive or negative; therefore, the optimal $b$ value could be 0, 1, or an internal solution satisfying FOC. The equations we obtain in doing this analysis are quite complicated and are available upon request. For illustration purposes, we include the first derivative of the cost function with respect to $r$ and $b$:

$$
\frac{\partial \text{STC}^{app}}{\partial r} = \frac{\lambda c_e}{\mu_e} (k(b)F^r(b,r)) - 0.5\alpha_e (k(b)F^r(b,r)) \sqrt{\frac{\lambda}{\mu_e(1-k(b))F(b,r)}}
$$
$$
+ c_e \lambda k(b) \left(1 - \frac{1}{b\mu_i} \right) + m_e \lambda k(b)(1 - F^r(b,r)) - m_e p \lambda k(b)F^r(b,r)
$$
$$
+ 0.5\alpha_e \left(1 - \frac{1}{b\mu_d} \right) - \frac{\lambda k(b)}{1 - r + \frac{r}{b\mu_d}}
$$

$$
\frac{\partial \text{STC}^{app}}{\partial b} = -\frac{\lambda c_e}{\mu_e} (k(b)F^b(b,r) + k'(b)F(b,r)) - c_e \lambda k(b) \left(1 - \frac{1}{b\mu_d} \right) + \frac{r}{b\mu_d}
$$
$$
+ c_e \lambda k'(b) \left(1 - \frac{1}{b\mu_d} + \frac{r}{b\mu_i} \right) - 0.5\alpha_e (k(b)F^b(b,r) + k'(b)F(b,r)) \sqrt{\frac{\lambda}{\mu_e(1-k(b))F(b,r)}}
$$
$$
+ 0.5\alpha_b \left(1 - \frac{1}{b\mu_d} \right) + \frac{k'(b)}{b\mu_i} - \left(1 - \frac{1}{b^2\mu_d} + \frac{r}{b^2\mu_i} \right) k(b) \sqrt{\frac{\lambda}{k(b)(1 - r + \frac{r}{b\mu_d})}}
$$
$$
+ m_e \lambda (k(b)F^b(b,r) + k'(b)(r - F(b,r))) - m_e p \lambda (k(b)F^b(b,r) + k'(b)F(b,r)) + C'(b)
$$
where \( \alpha_i = \alpha(y_i, c_i, w_i) \) for \( i = s, e \), and \( F'(b, r) \) is the first derivative of the function \( F(b, r) \) with respect to \( j = b, r \).

**Proof of Proposition 2.** Similar to the methodology in [15, 26], we first find the internal optimal tele-specialist treatment ratio (\( r^* \)) that satisfies the first order condition \( \partial STC_{\text{opt}} / \partial r = 0 \). We then apply implicit differentiation to this equation with respect to the parameters considered in the sensitivity analysis to obtain the properties of \( \partial r^* / \partial \mu_t \) and other partial derivatives. For example, using this approach, we observe that \( \partial r^* / \partial \mu_t \) is universally positive for all the parameter values under consideration. The equations we obtain in doing this analysis are quite complicated and are available upon request.

**Proof of Proposition 3.** As in Proposition 2, we first find the internal optimal level of technology investment in telemedicine (\( b^* \)) that satisfies the first order condition \( \partial STC_{\text{opt}} / \partial b = 0 \). We then apply implicit differentiation to this equation with respect to the parameters considered in the sensitivity analysis to obtain the properties of \( \partial b^* / \partial c_s \) and other partial derivatives. For instance, using this approach, we observe that \( \partial b^* / \partial c_s \) is universally positive for all the parameter values under consideration. The equations we obtain in doing this analysis are quite complicated and are available upon request.
References


