International Journal of Production Research, 2015
Vol. 53, No. 24, 7213–7225, http://dx.doi.org/10.1080/00207543.2015.1005875

Managing nurse lines – practical challenges and the developing theory

Xiaofang Wang[a]*, Laurens G. Debo[b] and Alan Scheller-Wolf[c]

[a]*School of Business, Renmin University of China, Beijing, P.R. China;* [b]*Booth School of Business, The University of Chicago, Chicago, IL, USA;* [c]*Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA*

(*Received 1 May 2014; accepted 13 December 2014*)

As a fast growing specialty, medical telephone triage lines (nurse lines) play an important role in providing advice and recommendations to help ensure that medical care is delivered at the appropriate level. In this paper, we discuss the operations of these nurse lines, based on studies in the literature and interviews with nurse line researchers and practitioners. This leads us to identify issues relevant for the design and management of such lines, challenges that operations research (OR)/operations management (OM) can help address. We demonstrate that a classical call center modelling approach, focusing exclusively on congestion, will be insufficient in addressing these challenges. Thus, we motivate research on an integrated framework in which both congestion and decision-making accuracy are coupled. We summarise the limited extant research in OR/OM that is applicable to the unique challenges posed by nurse lines, highlighting these models' contributions, as well as the questions they leave unanswered. These questions provide significant opportunities for future research.

Keywords: health care; service operations; call center

1. Introduction

Mismatch of supply with demand in the healthcare domain can have a dramatic effect on costs and efficiency: According to statistics reported by the Centers for Disease Control and Prevention (CDC) in the United States (McCaig and Burt 2003), during 2003, of an estimated 113.9 million emergency room visits, 13% were non-urgent. At an estimated average cost of $300 per emergency room visit (Machlin 2006), a cost of $4.4 billion could have been managed more efficiently by directing patients to the appropriate care center. Such savings opportunities motivate medical practitioners and health insurance companies to invest in a type of call center with the goal of educating and informing patients about treatment options, in hopes of reducing the level of mismatch in care. These call centers, typically called nurse lines, have evolved into a recognised specialty. At these lines, telephone triage nurses provide advice and treatment recommendations over the phone, typically based on patients' answers to questions generated by diagnostic protocols. Effective nurse lines can relieve unnecessary traffic at expensive/urgent medical treatment centers, while also directing people who need intensive care to seek it, thereby reducing costs while maintaining quality and safety (Cariello 2003; Bogdan et al. 2004; Bunik et al. 2007). The objective of this paper is to identify the management concerns and research opportunities arising from the emerging practice of nurse lines. Specifically, we demonstrate why a classical call center modelling approach cannot effectively address challenges from nurse lines, especially the need to well balance clinical decision accuracy and operational efficiency. Based on detailed discussions on nurse line operations and related operations research (OR)/operations management (OM) work, we further propose and motivate research on an integrated framework coupling both congestion and decision accuracy, which complements the existing literature.

Many health organisations provide telephone triage lines, including HMOs and health insurance companies, Medicaid and Medicare programs, medical groups, and emergency departments. Some offer the services directly; others may contract with large organisations that provide telephone triage nurse services from call centers. For example, in Pittsburgh, one of the leading health insurers in Pennsylvania, Highmark, contracts with a company called Health Dialog to provide its more than three million members with access to nurse line services through a program known as 'Blues On Call' (Highmark 2005). Other examples are the two nurse lines whose directors we have interviewed: 'After Hours Telephone Care Program' is affiliated with the Children's Hospital in Aurora, Colorado; and 'NurseAdvice in New Mexico' was created by the Coordinated Systems of Care Community Access Program of New Mexico, a consortium of ten public and private health care organisations.

The performance and impact of nurse lines have been widely studied in the clinical literature. This literature has addressed several important issues in nurse lines, including evaluations of positive cost-savings that can be attributed to nurse lines

*Corresponding author. Email: xiaofang_wang@ruc.edu.cn

(Cariello 2003; Bogdan et al. 2004; Bunik et al. 2007), the safety/appropriateness of nurse line advice (Kempe et al. 2000), and nurse lines' performance with respect to wait times and call durations (Valanis et al. 2003; The Quality Improvement of Literacy, TeleCare, and Self Help (QUILTS) Collaborative 2006). More comprehensively, Poole (2003) provides a guide for developing a telephone advice system for a pediatric office practice; it describes the methods and detailed steps which have been developed and tested over 15 years in children's hospital telephone advice programs nationwide. All of these research studies are based on field experiments or empirical surveys, and focus on only one of the involved parties in a nurse line (e.g. Bunik et al. 2007 focus on the HMOs, Kempe et al. 2000 focus on the patients). As a complement to empirical studies, analytical models that study the impact of nurse line service on *all* involved parties are of vital importance. For example, such models would prove useful in generating insights into how different management options could affect different parties, in the hopes of making nurse lines more effective overall. Such models do not exist in the clinical literature, but have been successfully developed for *traditional* call centers in the OR/OM literature. We next discuss how traditional call center models fall short with respect to capturing nurse line operations, and how these traditional models have begun to be extended.

The papers in the OR/OM literature have provided many analytical models for call centers incorporating different parties' perspectives. But most of them focus solely on congestion issues in service systems: The management objective is typically to achieve short mean waiting times (see references in Gans, Koole, and Mandelbaum 2003 and Aksin, Armony, and Mehrotra 2007). These papers do not address two of the distinguishing characteristics of nurse lines: (i) Agents are performing a diagnostic process in which they constantly balance decision accuracy and congestion; and (ii) use of nurse lines are not mandatory – patients may choose to call or not to call, which is important because the overall benefit of a nurse line is proportional to call volume aggregated from each individual customer's choice.

Regarding the first characteristic of nurse lines, a decision accuracy-congestion tradeoff is akin to the quality-speed tradeoff that has been studied in the queueing/dynamic control literature in the context of call centers. However, in that literature, the key concept as an important measure of service quality is callback, i.e. poor service quality results in a follow-up service request (retrial) which is called a callback. In addition, the major management decision is routing. These deviate from the focus of nurse lines on averting inappropriate treatment and maintaining precision of advice. For example, de Véricourt and Zhou (2005) study the routing problem to minimise the average total time of call resolution, including callbacks. They provide sufficient conditions under which it is optimal to route to the customer service representative with the highest call resolution rate (service rate times resolution probability). In addition, Armony (2005) suggests that the methodology presented in her paper can be extended to prove this routing rule is asymptotically optimal. But Mehrotra et al. (2012) show that such a routing rule may perform very poorly with respect to objectives that involve the call resolution. The importance of such objectives is also recognised in Hart et al. (2006). Thus, Zhan and Ward (2014) formulate an optimisation problem to determine the optimal routing with the dual performance objective of minimising average customer waiting time and maximising the call resolution. They solve this optimisation problem asymptotically and give insights via simulation. Note that callbacks are quite different from decision accuracy and cannot represent service quality in nurse lines. This is because in nurse lines the customers not calling back still incur the error costs due to inappropriate treatment decisions. In addition, the above papers do not model the agent's diagnostic process over the phone and thus cannot capture many details quantifying clinical decision accuracy and related error costs.

Recently, the quality-speed tradeoff has also received increasing attention in more general service settings other than call centers. When performing a 'discretionary task' (Hopp, Iravani, and Yuen 2007) or a 'customer intensive service' (Anand, Fazil Pac, and Veeraraghavan 2011), the service provider can terminate processing a task to make a good time/quality tradeoff when quality (value) of the task is an increasing function of service time. Kostami and Rajagopalan (2014) study this similar quality/speed tradeoff in a dynamic model where both speed and price are levers in determining demand levels. They capture the impact of speed on quality through a change in future market potential, which is a function of the speed in the current period. Tong and Rajagopalan (2014) investigate the optimal price and service time for such type of discretionary services when both service rate and demand are endogenous and functions of the pricing scheme. The authors evaluate two commonly used pricing schemes. In Alizamir, de Véricourt, and Sun (2013), they specifically model the diagnostic process in detail as a hypothesis test in a dynamic setting, which is close to the diagnostic process seen at nurse lines. However, all the above papers assume that the service system is centralised, without considering the second characteristic of nurse lines – customer choice to call or not. This deviates from the nurse line practice.

In the general settings of customer intensive services, Anand, Fazil Pac, and Veeraraghavan (2011) do incorporate both characteristics, the service quality/speed tradeoff and customer choices. But they abstract away the modelling details of services involving diagnosis and simply assume that service value is a monotonic function of the service time. The two papers that most closely model nurse line characteristics are Wang et al. (2010) and Wang et al. (2012). They model the nurse's diagnostic process in detail as a sequential testing process in which high decision accuracy is in conflict with the desire for short testing/service time. In addition, they model customer behaviours in choosing to call or not based on their perceptions of the quality (appropriateness of advice) and wait time of the nurse line.

The OR/OM papers described above provide initial steps towards characterising and claiming the societal benefits of nurse lines, but they still leave many questions unanswered. We motivate a more extensive consideration of nurse line operations, in the hope of facilitating more efficient and effective deployment of nurse lines. This would provide benefits for providers, patients and society as a whole. To do so, we first discuss the operations of nurse lines in detail in Section 2. Then in Section 3, we motivate the use of an integrated framework in which both congestion and decision-making accuracy are coupled to address these concerns, drawing on models from literature. In Section 4 we conclude the paper and present future research opportunities.

2. Nurse line operations, management concerns and key decisions

Organisations offering telephone triage services often provide toll-free numbers that quickly connect callers (who may be patients, parents, spouses or friends) to nurse line agents. Typically, but not exclusively, nurse line services refer to inbound phone calls.[1] Callers may call a nurse line agent 24 hours a day, 7 days a week when they need medical information or when they are trying to make a healthcare decision. Nurse line agents are usually registered nurses who are qualified to perform telephone triage. They have a range of titles: health coaches, advice nurses, telepractitioners, telenurses, telepractice nurses or consulting nurses. In an 8-hour shift, a nurse may take 60–80 telephone calls. Each call is processed within a brief time frame, usually 6 to 10 minutes on average (Wheeler 2003). The nurse first gathers required general information from the caller and then elicits the *symptoms* which prompted the call.

Nurses typically use computerised or paper guidelines (called *protocols*) to assess symptoms. The protocol can be viewed as an algorithm that guides the nurse through a sequence of diagnostic questions, based on the answers to earlier questions. The symptoms are classified according to body systems (cardiovascular system; eye, ear, nose and throat; gastrointestinal system; genital/urinary system; immune system; musculoskeletal system; neurological system). A protocol for a typical symptom is illustrated in Figure 1.

The end result of a protocol evaluation is a recommendation. Typically, this recommendation is a classification of the patient's symptoms into: Emergent, Urgent, Non-urgent, or Watch and Wait. This recommendation then leads to a selection of a treatment option by the patient, ranging from home care to dialing 911.

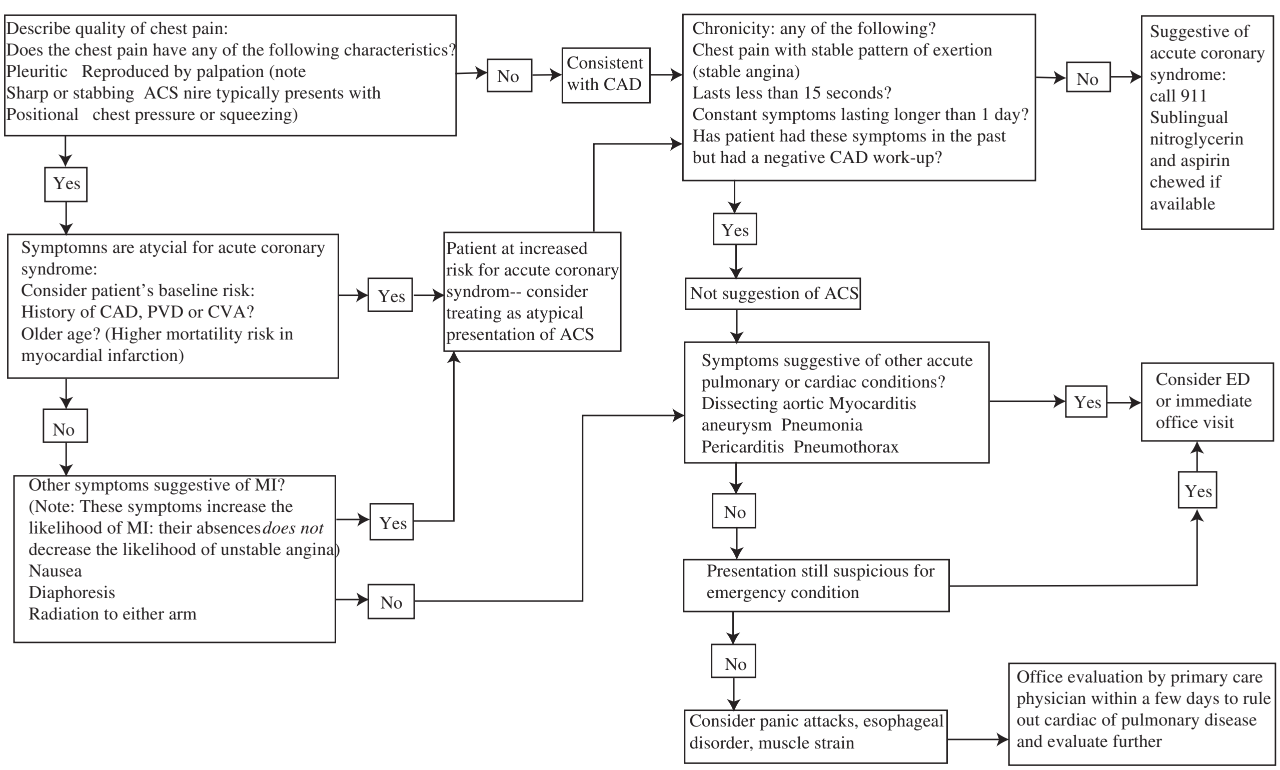

Figure 1. Protocol to follow for chest pain. Source: Reisman and Stevens (2002).

As an example, a set of protocols used in a nurse line we interviewed (Fiorenzio 2009) is *Barton Schmitt* in the computer form. *Barton Schmitt* is one of the highly regarded peer-reviewed pediatric protocols; it is used nationwide (AnswerStat 2003). First, after the greeting, the nurse recognises *symptoms*. Typical and often recurring symptoms are abdominal pain, nausea and vomiting, or respiratory problems. The nurse interviews the patient and writes down a description of the problem in the patient's own words. Then, the system prompts the nurse to ask 6–7 general questions to gather more information, recording the patient's responses. Based on the *keywords* in these pieces of information, the software system automatically searches and suggests the *appropriate protocols*. There may be multiple protocols appropriate for one patient; the nurse then chooses which she thinks is most appropriate. This is an important point where *nurses' experience/skill* plays a role (Fiorenzio 2009).

Within a certain protocol, a series of Yes/No questions are asked and then the final recommendation is generated (911/ED/urgent care/doctor visit/home care). The nurse must follow the protocol exactly, but the nurse can end a protocol before it makes a recommendation, if she thinks a different protocol is more appropriate. Such a protocol change may be suggested by the protocol itself: Protocols may involve branches – for example, the vomiting protocol may ask if there was a recent head injury. If the answer is yes then it advises the nurse to switch to the head injury protocol, and asks if she would like to do so (Fiorenzio 2009). Once she concludes one protocol, the nurse may not ask any more questions unless she starts a new protocol; in approximately 10–20% of the cases, more than one protocol is selected (Fiorenzio 2009; Hegarty 2009; Massaro 2009). The nurse can override the recommendation of a protocol only to recommend more urgent treatment (this happens about 2% of the time) (Fiorenzio 2009); when multiple protocols are selected, the most urgent outcome of the protocols is recommended. At the end of the call, the nurse files the call.

Next, we first describe the management concerns at nurse lines and then examine the crucial factors in determining nurse line performance: nurse line staffing, the skill sets required for telephone triage, and the different aspects of diagnostic protocols.

2.1 *Management concerns at nurse lines*

The following *management concerns* distinguish nurse lines: Safety and appropriateness of advice, wait times and liability issues. We discuss each of these in turn, and conclude with a brief discussion of the primary interactions between these factors.

2.1.1 *Safety and appropriateness of advice*

Not surprisingly, patient safety and appropriateness of the recommended care are the primary concern of the practitioners with whom we have spoken. Most nurse lines track the accuracy of the nurses' advice, in terms of *under-referral* errors and *over-referral* errors, either separately or in aggregate. Under-referral means referring severe patients to the non-urgent treatment options, and over-referral means referring non-severe patients to the expensive/urgent treatment options. In general, under-referral is more difficult to detect, because a follow-up with the caller needs to be done about the health outcome of the patient who was not recommended to seek urgent care. In addition, it is difficult to distinguish between an actual error due to a nurse misdiagnosing the condition, and a situation in which the condition of the patient deteriorated after the call. In contrast, over-referral is easier to detect because if the caller visited the ED or called 911 it will be logged, and it can be verified by the emergency physician whether the patient was really in need of urgent care or not (Massaro 2009).

Incidents of under-referral are typically followed up more severely (possibly with disciplinary actions) than over-referral. Thus nurses tend to err on the safe side. This may explain why some clinical studies (Bunik et al. 2007) indicate that there are very few cases of under-referral. However, over-referral is costly to the patient because of the unnecessary travel costs and the risks of undergoing unnecessary care and/or picking up diseases in the hospital. Obviously it is also costly to the health service provider.

Typically there exists a standard (target) for the appropriateness of advice at nurse lines. A common target is to make 95% appropriate decisions in aggregate (Massaro 2009). Even though this is an average of both over- and under-referral counted together, as the number of under-referrals is much smaller than over-referrals, the bulk of these tracked errors come from over-referrals. One nurse line we interviewed performs regular quality improvement projects to follow on up their accuracy records. Nurse lines may also undertake a project if a physician notifies them of a referral that was found to be inappropriate (Massaro 2009).

2.1.2 *Wait times and liability issues*

Using a nurse line service is voluntary: Potential callers choose to use the service or not, primarily based on their perception of the nurse line's quality (appropriateness of advice) and wait time. To entice callers, health officials and organisations seek to improve responsiveness without sacrificing accuracy of advice (Guadagnino 1998; The Quality Improvement of Literacy, TeleCare, and Self Help (QUILTS) Collaborative 2006).

There exist industry standards which nurse lines strive to meet for wait times. These standards are typically set by the nurse line partners and the Department of Health and Human Services. For example, abandonment rates should be below 5% and wait times should be less than 30 seconds[2] (Fiorenzio 2009). As wait times are heavily influenced by call durations, these are monitored as well. Typically, the call duration target is set between 6 and 10 minutes (Wheeler 2003): Nurses must thus balance satisfying patients' needs with processing the call backlog in a timely manner. This may be difficult because patients are often emotional.

Thus managers, when designing nurse lines – and nurses, when operating nurse lines – must seek the optimal trade-off between quality and waiting times. To find this balance, managers and triage nurses must understand the nature of this trade-off with respect to safety, service and liability. This is one of, if not *the most*, fundamental questions facing the nurse lines.

The issue of liability is also a significant concern for nurse lines. Such liability arises from incorrect recommendations; we discovered no liability issues specifically due to long wait times. These liability cases have mostly arisen because nurse lines did not use professional nurses, or because the nurses did not use an evidence-based protocol, which allegedly led to erroneous dispensations (Fiorenzio 2009).

2.1.3 *Interactions between responsiveness and the appropriateness of the nurse line recommendation*

In nurse line operations, minimising wait times may interact/conflict with the safety and appropriateness of advice: In general, longer service times entail both greater accuracy and longer waits. Thus jointly determining accuracy and wait time targets becomes paramount. This trade-off can be seen in setting high-level goals for staffing and average call times. It also manifests itself in specific operational choices. We provide two examples of this below.

2.1.3.1 *Use of multiple protocols*. As we mentioned before, with a frequency of 10–20%, nurses follow two (or more) protocols for a caller. This is typically the case when symptoms are somewhat ambiguous, or when there are multiple symptoms. The most severe disposition of the two (or more) protocols is then taken. Using multiple protocols may be discouraged by the nurse line manager due to the *trade-off* between gaining more accuracy and decreasing waiting time. An interviewee (Fiorenzio 2009) mentions that during busy times, multiple protocols may be followed less frequently (to reduce call backlogs).

In this example the nurse's experience is very important, as this helps the nurse determine which protocols to select first, and also whether to use multiple protocols. Experienced nurses may know common protocols by heart, which speeds the diagnostic process. Thus, here, the interplay of accuracy, waiting time and skill level is manifested.

2.1.3.2 *Coping with emergency events*. Whenever an emergency event (for example, the H1N1 virus, the avian flu, etc.) occurs, nurse line volumes typically increase significantly. This might severely strain the staffing at nurse lines. Moreover, in the face of such emergent medical situations, protocols may need to be updated, to guarantee *appropriateness of the recommendation*. Prior to receiving an update, nurses err on the safe side and typically take a *longer* time to dispense callers (they also may have to look up other sources of information – such as from the CDC – rather than operating from a self-contained protocol). Thus again the trade-off between accuracy and wait time is apparent.

Note that protocol choice directly influences this tradeoff. Using a protocol that is regularly updated to incorporate emergent conditions eases this trade-off between accuracy and wait time. Equipped with new protocols, nurses can maintain high accuracy of advice without sacrificing efficiency (waiting time) in the face of emergency events.

2.2 *Nurse staffing and skills required*

Nurse line managers determine staffing based on statistical analyses of historical call data, including demand volumes and trends. In general, managers follow traditional call center staffing rules, but as shown in Poole (2003), the staffing rule in practice is often simplified: the product of the call volume and the expected service time. Given the variability of call volumes during the day/season/holidays, nurse lines also commonly use surge staffing – they have employees who are on-call and available to answer calls from home, where they have appropriate protocol implementations. If call volumes are unexpectedly high, calls may be routed to these nurses to provide additional staffing.

Nurses must have the ability to think critically and quickly, apply deductive reasoning and multi-task. Extensive nursing experience is required, but not sufficient, for tele-triage, because bed-side decisions can be made much more easily than decisions over the phone. So specialised training in call center services is needed. Nurse skill also impacts the speed of serving the patient. For example, one nurse line we interviewed (Fiorenzio 2009) stated that a well-trained nurse can serve 6 calls/hour while a new nurse can only serve 4 calls/hour. Thus, a nurse with a higher skill level works both faster and more accurately.

Therefore, a complete model of nurse line staffing should include both base and surge staffing by nurses, of possibly different skill levels. Moreover, a manager not only must decide on staffing levels, but also on the skill level of nurses to employ. While such questions are considered in the call center literature where they affect service speed (Gans, Koole, and Mandelbaum 2003; Aksin, Armony, and Mehrotra 2007), their joint impact on the unique management concerns of nurse lines has not yet be studied.

2.3 *Protocol and its implementation*

As an important high-level decision, nurse lines select protocols and decide on how these protocols are to be implemented; for example, whether to use a book form or to invest in a computer system. Protocols implemented in a nurse line typically comprise a *set* including many individual protocols for different symptoms; different implementations may use the same protocol set. For example, the protocol set of *Barton Schmitt* is implemented in different computer formats: McKesson, LVM Systems, Epic, Intellicare and Fonemed. Together these systems support over 400 call centers (AnswerStat 2003). In addition, the *Barton Schmitt* book format is used in an estimated 10,000 pediatric offices (AnswerStat 2003). The nurse lines we interviewed (After Hours Telephone Care Program at the Children's Hospital, Aurora, CO, and NurseAdvice New Mexico) both work with Barton Schmitt protocols.

Table 1 highlights the variation in page length and software features of some currently available protocol sets and their implementations. Given the variation, it is agreed that different protocol implementations may result in different diagnostic performance for a given symptom (Wheeler and Siebelt 1997) *as well as* call duration (Bunik et al. 2007). For example, compared to a non-computerised protocol implementation, a computerised protocol implementation can bring up the most relevant protocols faster, as it is based on automated keyword searches instead of relying on a nurse's memory. It is believed that a low-skilled nurse equipped with an advanced computer system can typically perform as well as a highly-skilled nurse using a book, in terms of speed and accuracy (Fiorenzio 2009; Massaro 2009).

Thus, a complete model of nurse line staffing should also take into account protocol choice *and* implementation, and their effect on call times and accuracy. Moreover the interaction of these effects with staffing must also be modelled.

Table 1. Protocols and their implementations: B means book form, SW means software, and ± indicates an estimate.

Name/Vendor	Form	Protocols	Features
Simonsen	B	± 94 pages	Adult and pediatric
Reisman & Stevens	B	± 412 pages	Adult and pediatric
Briggs	B	± 668 pages	Adult and pediatric
Lafferty and Baird	B	± 416 pages	Adult and pediatric
Schmitt	B	± 450 pages	Pediatric
Brown	B	± 320 pages	Pediatric
Katz	B	± 260 pages	Pediatric
Clinical solutions	SW	Teleguide™ Algorithms	Seamlessly transfer patient data
HealthLine Systems, Inc. Sharp Focus	SW	The Cleveland Clinic Foundation	Control hours of operation, information and staffing
LVM Systems	SW	Schmitt/Thompson	Frequently used protocols, adjustable views of protocols, remote access
McKesson's ASK-A-NURSE program	SW	Schmitt	
Fonemed, LLC	SW	Schmitt/Thompson	Staff have less to remember, thus mistakes are less likely to occur

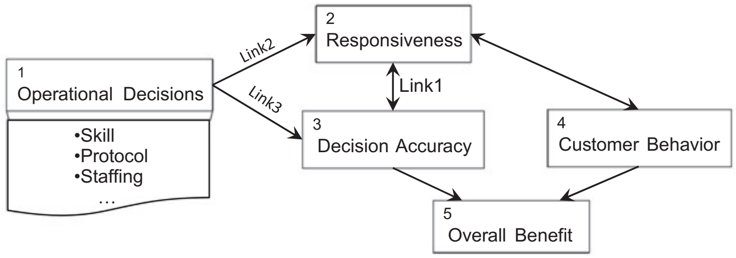

Figure 2. An integrated framework for nurse lines.

3. Integrated analysis of nurse lines

In this section, we develop an integrated framework in Figure 2 to summarise key features and the connections among operational decisions, performance measures and customer behaviours in nurse lines. Link 1 describes the central tradeoff in nurse lines between two major performance measures, responsiveness and decision accuracy. These two performance measures are influenced by operational decisions via Link 2 and Link 3, respectively. Link 2 (the impact of operational decisions on responsiveness) has received much attention among traditional call center managers and has been widely studied in OR/OM. However, Link 3 (the impact of operational decisions on decision accuracy) is generally missing in traditional call center models. We include Box 4 because customer behaviours are important in determining the overall benefits (Box 5) of nurse lines. At nurse lines the overall benefits are proportional to the volume of calls that successfully averted inappropriate treatment and the call volume is aggregated from each individual customer's choice of using nurse lines or not.

In summary, although extant service models in OR/OM literature have addressed one or some pieces of this framework, more analytical work is needed to integrate all the boxes and links together to promote a better understanding and a more efficient use of nurse lines.

3.1 *The tradeoff between decision accuracy and responsiveness (Link 1)*

In nurse lines, a central tradeoff is between minimising waiting times and increasing the appropriateness of the nurse's recommendation as shown in Link 1. Fundamentally, nurse line agents are making discretionary decisions and their decision accuracy increases with diagnostic time (service time), which entails a time/quality tradeoff. This similar tradeoff has raised increasing interests in more generalised service settings. Under different names, 'discretionary task' in Hopp, Iravani, and Yuen (2007) and 'customer intensive service' in Anand, Fazil Pac, and Veeraraghavan (2011) emphasise that the value obtained from service increases with service time and completion of the service process is subjective. They directly model this time/quality interaction as a monotonic functional relationship. Motivated by triage nursing and engine maintenance systems, Alizamir, de Véricourt, and Sun (2013) develop a discrete elicitation sequence of binary probabilistic cues in a dynamic decision framework.

Only very few articles focus on this tradeoff in call centers. Motivated by nurse lines, Wang et al. (2010) and Wang et al. (2012) abstract the diagnostic process into a continuous Brownian motion with unknown drift, which corresponds to the patient's unknown pathology. Nurse line agents' discretionary decisions are modelled according to what they call the *service depth*. This captures the different protocol-related choices affecting the accuracy/response time tradeoff that the nurse line can make, for example, whether/when to discourage multiple protocols.

Previous research work shows that once this time/quality tradeoff exists, service managers should strive to optimise performance resulting from the *overall* effects of service efficiency and service value (accuracy): When reducing congestion is a dominant factor for profitability, the added capacity should relieve congestion, but may reduce service value (accuracy). Conversely, when increasing service value (accuracy) is a dominant factor for profitability, the added capacity should be used to improve value, but may increase congestion. In nurse line practice, unfortunately, this insight has not been rigorously verified empirically. Accuracy and response time targets are typically set independently; their interaction is generally ignored in management (Fiorenzio 2009). Thus, extant nurse line implementations likely fall short of securing the maximum possible benefit. In addition, how to make these relationships as accurate as possible, and how this may affect individual nurse's decisions, remain open analytical and empirical questions.

3.2 *Operational decisions and responsiveness (Link 2)*

The second link in Figure 2, representing the impact of operational decisions typically made by call center managers on responsiveness-related performance measures, has been widely studied in OR and OM. These decisions include determining how many agents to hire and scheduling available agents based on demand forecasts for a given time period. There may be additional short-term decisions including forecast updating, schedule updating and call routing. Staffing decisions may be made in multi-skill settings, that is, in call centers where calls of different types are served by service representatives with different skills. Once the arrival process, queueing discipline and service mechanism of the call center system are given, based on an appropriate queueing or simulation model, one can derive or estimate the responsiveness-related performance measures such as average waiting time, average queue length and the utilisation of the server. Gans, Koole, and Mandelbaum (2003) and Aksin, Armony, and Mehrotra (2007) provide good surveys about these models.

3.3 *Operational decisions and decision accuracy (Link 3)*

The third link connects typical nurse line operational decisions such as skill, protocol, staffing to service value related performance measures. This link is generally missing in traditional call center models. Next, we will explain both the practical and the theoretical values of this link, drawing on relevant OR/OM models.

3.3.1 *Skill*

In all our discussions with practitioners, the nurse's skill level and experience are mentioned as key factors in their human resources policy, and key determinants for the speed and accuracy of the recommendation. Modelling this aspect will broaden the traditional definition of agents' skill which is normally modelled as the ability to handle calls of different types (e.g. Harrison and Zeevi 2004; Armony 2005; Wallace and Whitt 2005; Bassamboo, Harrison, and Zeevi 2006; Cezik and L'Ecuyer 2008, Pot, Bhulai, and Koole 2008; Chevalier and den Schrieck 2009). For example, Alizamir, de Véricourt, and Sun (2013) and Wang et al. (2010) broaden the definition of skill into a measure of productivity: given other parameters, more highly-skilled nurses perform diagnosis faster and more accurately. Specifically, Alizamir, de Véricourt, and Sun (2013) characterise skill effect as the increase in cue validity and the increase in the cue elicitation rate (thus, improving the diagnostic speed and accuracy) to characterise the skill effect resulting from hiring more experienced employees, training employees, or improving technology. Wang et al. (2010) capture this by modelling nurses' skill level as the difference between two drifts (two pathologies) of the Brownian motion, which determines the diagnosis speed and accuracy. However, none of these papers directly address skill-level decisions and their impact on this tradeoff; they take skill levels as exogenously fixed. Questions concerning skill-assortment as a decision variable are important future directions.

3.3.2 *Protocol*

Given the high variation among protocol implementations, how to implement protocols in the service process is just as important as other decisions such as staffing and training.[3] For example, similar to nurses' skill, a more advanced (computerised) implementation may improve both speed and accuracy. Wang et al. (2010) jointly study the effects of protocol implementation and staffing on accuracy and waiting time. The relative effect of different implementations and how a nurse line can choose the best protocol for its specific market and needs remain open questions.

3.3.3 *Staffing*

In the nurse line practice, staffing decisions are usually based on 'statistical analysis of historical demand data' (Massaro 2009), which typically follow call center theories without incorporating appropriateness of advice or the discretionary service property. This fails to account for the fact that, due to the discretionary nature of diagnostic service, adding capacity can actually allow for longer service, which increases accuracy, thus having an impact on both congestion and the error rate. Therefore, different from other service systems, in nurse lines, adding capacity (increasing worker productivity or number on staff) may increase congestion or increase error rates. This result has been numerically observed in more general discretionary services by Hopp, Iravani, and Yuen (2007) and Alizamir, de Véricourt, and Sun (2013). Wang et al. (2010) show analytically that in nurse lines, this observation remains true.

Most papers in the OR/OM literature address staffing decisions at the tactical level, and their focus is to design staffing algorithms to optimise various wait-related measures (e.g. Aksin and Harker 2003; Armony, Plambeck, and Seshadri 2009; Baron and Milner 2009). While wait-related and value(accuracy)-related measures are optimised together, Wang et al. (2010)

establish a staffing result extending the well-known principle regarding capacity choice, the *Square Root Rule for Safety Staffing* (Whitt 1992). We discuss this result in more detail below.

Let $R = \frac{\lambda}{\mu}$ denote the (average) offered load, where λ is the average call arrival rate and $\frac{1}{\mu}$ is the mean call duration. The *Square Root Staffing* principle is as follows: for moderate to large values of R (or equivalently moderate to large λ assuming that μ is fixed), the appropriate staffing level is approximated by

$$M = R + \beta\sqrt{R},$$

where β is a positive constant that depends on the desired level of service. Resulting from this principle is the concept of *statistical economies of scale*: The required excess capacity $\beta\sqrt{R}$ beyond nominal requirements R to achieve the target service level grows less than proportionately with the load of calls to be handled.

Note that the staffing rule used in nurse line practice is a simplified Square Root Rule for Safety Staffing (Poole 2003): $M = R$, ignoring the excess capacity. The staffing rule developed in Wang et al. (2010) is a modified version of the square root rule, capturing the excess capacity needed due to both accuracy concerns and stochastic variability. Furthermore, the underlying statistical economies of scale leads to their 'capture-all-demand' staffing result: The optimal staffing is such that the nurse line captures the highest demand it can, because the increase in staffing cost grows less than proportionately with the load of calls to be handled (and thus so does the revenue from the demand served). The intuition behind this result is that the nurse line manager's staffing decision must focus particularly on increasing the customer patronage because the nurse line demand is more elastic to staffing than a traditional call center: At nurse lines, staffing influences the demand rate both through its impact on waiting times *and* through its impact on accuracy. In addition, the benefit generated by nurse lines largely depends on the number of users (aggregate demand).

In summary, extant work has shown the potential of integrated analysis in generating insights into staffing decisions in nurse line practice. In addition, it also confirms that joint decisions on staffing and service depth, or in other words, modelling both accuracy and wait time concerns, complements the traditional staffing conclusions in the call center literature.

3.4 *Customer behaviours (Box 4)*

In traditional call center models, customer behaviours are typically abstracted into inputs to queueing models, in the forms of arrival process and abandonment with given distributions. Individual behaviour models for customers are rarely seen in these models. Gans, Koole, and Mandelbaum (2003) have pointed out that it is one of the most challenging aspects in developing queueing models for call centers is the incorporation of human factors for both customers and agents in a practical manner. They suggest that an appropriate framework is that of a game-theoretic or economic equilibrium, arrived at through learning and self-optimisation. Along this line, some strategic queueing models have been successfully applied to call center applications when customer behaviours are emphasised. In these papers, the demand rate is aggregated from an individual utility optimisation, which balances perceived waiting costs against the perceived benefits of service. For example, Hassin and Haviv (1995) consider an $M/M/1$ queue in which the reward of service completion for an individual reduces to zero after some time. Customers compare the waiting cost and the reward to decide whether to join, and if they join, when to renege. Mandelbaum and Shimkin (2000) consider an $M/M/m$ queue, in which they derive a rational decision framework for determining the abandonment times of waiting customers based on an individual utility optimisation. From this the patience (patience is the time a customer is willing to wait in queue) distribution emerges as an equilibrium point. In these models, since customers only anticipate and respond to waiting times not service values which are assumed to be fixed, there is no tradeoff between responsiveness and quality.

However, when the strategic queueing model is applied to nurse lines where agents can control service value and the time/quality tradeoff exists, the equilibrium appears different from traditional results. As shown in Wang et al. (2010), the equilibrium joining probability is always 1 when the number of servers can increase at a linear staffing cost, which has been explained in Subsection 3.3.3; this is quite different from previous findings in the strategic queueing literature. A similar effect is demonstrated in Anand, Fazil Pac, and Veeraraghavan (2011). When customer choice is incorporated in a more generalised service setting with the time/quality tradeoff, they show that customer-intensity, i.e. the level of correlation between quality and speed, is a critical driver of equilibrium price, service speed, demand, and congestion in queues, and consequently of service provider revenues. These findings are again very different from those of traditional strategic queueing models (Hassin and Haviv 2003).

In addition, patients differ with respect to their symptoms, priors, associated costs, insurances and waiting costs. It might be valuable to incorporate such heterogeneity into customer models. For example, different symptoms primarily determine different call durations and possible outcomes (referral to urgent versus. non-urgent care), see our description of a protocol in Figure 1. Depending on the protocol and the age group, at most a few hundred symptoms may be relevant (Lafferty and Baird 2001). Thus it might be useful to characterise patients by symptom and correct outcome. Theoretically, a few papers model customer heterogeneity in customer-intensive services which are more general than nurse lines (Ni, Xu, and Dong 2013; Tong and Rajagopalan 2014). However, although heterogeneity is clearly present in practice, building heterogeneous customer models for nurse lines is not straightforward. Because it remains unclear clinically which types of heterogeneity are the most relevant and how different types of heterogeneity are correlated. Once more empirical supports are provided, extant models can be extended by exploiting OR/OM techniques.

3.5 *Overall benefits from providing nurse line services (Box 5)*

The OR/OM literature for call centers has traditionally focused on minimising customer waiting times and agent staffing costs (Aksin, Armony, and Mehrotra 2007). The role of engaging customers and maintaining customer satisfaction has not been emphasised historically, but it is crucial for nurse lines: Different from other call centers, at nurse lines the overall benefits are the cost savings from the averted inappropriate treatment (either under or over provision) in the entire health care delivery system. Many studies have estimated positive treatment cost-savings that can be attributed to nurse lines. The main reason is that nurse lines help direct callers to the appropriate level of health care, relieving the unnecessary traffic at the expensive/urgent options (Cariello 2003; Bogdan et al. 2004; Bunik et al. 2007). Thus, the benefit of offering this service (cost savings) both depends on the individual benefit (cost savings from each averted inappropriate treatment) and depends on the volume of calls aggregated from individual patient's decisions as to whether to use the service or not. The individual benefit is directly determined by accuracy of agent advice which is linked to operational decisions. The demand volume is linked to customer behaviours which respond to both service quality (accuracy) and responsiveness in nurse lines.

New evidence has confirmed the importance of customer population in the optimal decisions. In practice, although staffing decisions are call volume dependent, decisions regarding protocols and accuracy (i.e. service depth) typically are not. Wang et al. (2010) provide a different perspective: In order to maximise the total benefit from investing in a nurse line, the accuracy target for the nurse line should also be based on the population size. This result implies that large nurse lines having many potential patients have greater incentive to improve the appropriateness of their recommendations. In practice, one nurse line we interviewed (NurseAdvice New Mexico) reaches about one million inhabitants, half of New Mexico's population. This nurse line works with a computerised implementation of a peer-reviewed, highly regarded protocol. Smaller nurse lines, for example cooperatives between individual physicians, might use manual protocols instead of computerised systems, which some practitioners view as less accurate. Further, some small nurse lines may even do telephone triage without protocols, which is considered very dangerous. This phenomenon suggests that as nurse line size increases, error rates may decrease, as predicted by Wang et al. (2010).

Based on nurse line's unique benefit structure, more questions and insights can be explored. For example, how do other characteristics of customer population such as age, health insurance and health status impact cost savings and optimal decisions? What are the appropriate ways to estimate cost savings involving estimating what the costs would be without a nurse line and measuring the real incurred cost after using the nurse lines?

4. Conclusion and future research avenues

Based on our survey of literature and interviews, we find there to be a very rich research potential on nurse lines and diagnostic service centers in general. The models we discuss above for the integrated analysis serve as a first step towards understanding this emerging field. We summarise remaining open questions as follows.

First, towards practical applications, more elaborate customer and agent decision models need to be developed. Although stylised models have been developed for 'discretionary task' or 'customer-intensive services' to show new insights in the presence of time/quality tradeoff, models for health care call centers incorporating human factors for both customers and agents are rare. It will be multi-disciplinary research based on a better understanding of customers (typically anxious patients) and agent (skilled nurses) behaviours. Due to importance of nurse lines in the health delivery system, more work needs to be done to make the time/quality relationships as accurate as possible and clearly indicate how operational decisions affect it. New models and results will also broaden the traditional view of service quality, skill, customer heterogeneity and prioritisation. For example, from a nurse line decision-maker's perspective, protocols are symptom-based, and symptoms determine different call durations and possible outcomes. Thus, a possible extension is to study the optimal service depth and prioritisation for each major symptom, with individual error costs and waiting costs. In addition, one may study how the accuracy concern interacts with other classical call center concerns including: (1) where to physically locate nurses (close to where the call originates from, or at another low-cost centralised location); (2) training of nurses (Poole 2003); (3) advertisement of nurse line services: making potential caller populations aware of the nurse line and promoting its usage (The Quality Improvement of Literacy, TeleCare, and Self Help (QUILTS) Collaborative 2006); and (4) use of skill-based routing to assign incoming calls to the most suitable nurse with the right speciality/skill sets.

Second, integration of nurse line models in the entire health care delivery system needs to be investigated. How can a nurse line be jointly managed with major health care facilities such as an emergency room, to alleviate congestion and deliver high quality care? What are the customer and provider strategies when multiple nurse lines compete on service quality and response time? In addition, outbound call centers could be modelled and analysed. Due to the increasing importance of nurse lines, outbound calls are more and more often used in nurse lines to play an active and effective role in health education and health intervention. It is relevant in practice but has not received much attention in the OR/OM literature.

Finally, nurse lines have significant potential value as an emergency response to outbreaks, attacks or disasters. For example, the danger of H1N1 makes it important to have people stay home to limit contagion instead of rushing into health facilities. In such a situation, nurse lines can play a crucial role in advising anxious people and reducing infection and congestion at key health facilities. In a related example, in the face of unpredictable natural or man-made disasters, demand for emergency supplies such as water, shelter, food and medical services is highly unpredictable. A quick response and high-quality coordinated humanitarian relief efforts are needed. Emergency service centers can be established to coordinate these efforts, allocate resources, and solve problems by processing requests from different locations, and different people, with different needs. In these contexts, several research questions need to be answered. How could temporary staffing be designed to fulfil this emergency role? How can requests be prioritised to maximise the global benefit/social welfare? How can the service center maintain high quality when facing overwhelming demand?

Note that when nurse lines serve as emergency response centers, efficient forecasting methods for non-regular demand are needed. In the traditional call center literature, methodologies have been developed toward this direction of increasing forecast accuracy under high uncertainty. For example, Shen and Huang (2008) develop methods for interday (day-to-day) forecasting and dynamic intraday (within-a-day) updating to reduce forecast error. Ibrahim and L'Ecuyer (2013) evaluate different statistical models for the call arrival process in the forecasting accuracy using different lead times, ranging from weeks to hours in advance. Wooff and Stirling (2014) develop forecasting models for call centers which deal with patient requests for medical advice outside normal working hours. They discuss how the models can be adjusted to maintain forecast quality when swine flu affects call volumes. More challengingly, at nurse lines, customer behaviours (reactions to both service quality and congestion) are important to determine the actual demand. Then, the interesting research questions are as follows. How can these demand forecasting methods be combined with complex customer choice models to make a good prediction of call volumes? In the presence of the quality-speed tradeoff, how can these combined demand models be factored into operational decisions?

Funding

Xiaofang Wang's work was supported by the National Natural Science Foundation of China [grant number 71371009] and [grant number 71422004], and Program for New Century Excellent Talents in University [grant number NCET-13-0578].

Notes

1. Some nurse lines receive messages from the doctor's answering services and return patient calls in order of perceived urgency.
2. Actuals may in practice deviate from these targets.
3. In practice, the nurse's skill level will be an important intermediate variable that influences the relationship between the protocol form and the performance characteristics.

References

Aksin, O., M. Armony, and V. Mehrotra. 2007. "The Modern Call Center: A Multi-disciplinary Perspective on Operations Management Research." *Production and Operations Management* 16 (6): 665–688.

Aksin, O., and P. T. Harker. 2003. "Capacity Sizing in the Presence of a Common Shared Resource: Dimensioning an Inbound Call Center." *European Journal of Operational Research* 147: 464–483.

Alizamir, S., F. de Véricourt, and P. Sun. 2013. "Diagnostic Accuracy Under Congestion." *Management Science* 59 (1): 157–171.

Anand, K., M. Fazil Pac, and S. K. Veeraraghavan. 2011. "Quality-speed Conundrum: Tradeoffs in Customer-intensive Services." *Management Science* 57 (1): 40–56.

AnswerStat. 2003. "Dr. Barton Schmitt Interview: Telephone Triage Protocols." Accessed November 17, 2009. http://www.answerstat.com/articles/3/17.html.

Armony, M. 2005. "Dynamic Routing in Large-scale Service Systems with Heterogeneous Servers." *Queueing Systems: Theory and Applications* 51 (3–4): 287–329.

Armony, M., E. Plambeck, and S. Seshadri. 2009. "Sensitivity of Optimal Capacity to Customer Impatience in an Unobservable M/M/S Queue." *Manufacturing & Service Operations Management* 11 (1): 19–32.

Baron, O., and J. Milner. 2009. "Staffing to Maximize Profit for Call Centers with Alternate Service-level Agreements." *Operations Research* 57 (3): 685–700.

Bassamboo, A., J. M. Harrison, and A. Zeevi. 2006. "Design and Control of a Large Call Center: Asymptotic Analysis of an LP-based Method." *Operations Research* 54 (3): 419–435.

Bogdan, G. M., J. L. Green, D. Swanson, P. Gabow, and R. C. Dart. 2004. "Evaluating Patient Compliance with Nurse Advice Line Recommendations and the Impact on Healthcare Costs." *The American Journal Of Managed Care* 10: 534–542.

Bunik, M., J. E. Glazner, V. Chandramouli, C. B. Emsermann, T. Hegarty, and A. Kempe. 2007. "Pediatric Telephone Call Centers: How Do They Affect Health Care Use and Costs?" *Pediatrics* 119 (2): 305–313.

Cariello, F. P. 2003. "Computerized Telephone Nurse Triage: An Evaluation of Service Quality and Cost." *Journal of Ambulatory Care Management* 26 (2): 124–137.

Cezik, M. T., and P. L'Ecuyer. 2008. "Staffing Multiskill Call Centers via Linear Programming and Simulation." *Management Science* 54 (2): 310–323.

Chevalier, P., and J. C. V. den Schrieck. 2009. "Optimizing the Staffing and Routing of Small-size Hierarchical Call Centers." *Production and Operations Management* 17 (3): 306–319.

Fiorenzio, C. 2009. Interim Director of NurseAdvice New Mexico, a nurse line at NM, Personal Communication to A. Scheller-Wolf and L. Debo, September 3, 2009.

Gans, N., G. Koole, and A. Mandelbaum. 2003. "Telephone Call Centers: Tutorial, Review, and Research Prospects." *Manufacturing & Service Operations Management* 5 (2): 79–141.

Guadagnino, C. 1998. "Highmark's Blues on Call." Physician's news digest.

Harrison, J. M., and A. Zeevi. 2004. "Dynamic Scheduling of a Multiclass Queue in the Halfin and Whitt Heavy Traffic Regime." *Operations Research* 52 (2): 243–257.

Hart, M., B. Fichtner, E. Fjalestad, and S. Langley. 2006. "Contact Centre Performance. In Pursuit of First Call Resolution." *Management Dynamics* 15 (4): 17–28.

Hassin, R., and M. Haviv. 1995. "Equilibrium Strategies for Queues with Impatient Customers." *Operations Research Letters* 17 (11): 41–45.

Hassin, R., and M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Norwell, MA: Kluwer.

Hegarty, T. 2009. Clinical Nurse Manager at After Hours Telephone Care Program Affiliated with the Children's Hospital, Aurora, CO, Personal Communication to A. Scheller-Wolf and L. Debo, September 17, 2009.

Highmark. 2005. "Highmark Signs Six-year Extension of Alliance With Health Dialog; Condition Management Program to Continue Through 2011." Accessed November 17, 2009. http://www.highbeam.com/doc/1G1-131000887.html.

Hopp, W. J., S. M. Iravani, and G. Y. Yuen. 2007. "Operations Systems with Discretionary Task Completion." *Management Science* 53 (1): 61–77.

Ibrahim, R., and P. L'Ecuyer. 2013. "Forecasting Call Center Arrivals: Fixed-effects, Mixed-effects, and Bivariate Models." *Manufacturing & Service Operations Management* 15 (1): 72–85.

Kempe, A., C. Dempsey, J. Whitefield, J. Bothner, T. MacKenzie, and S. Poole. 2000. "Appropriateness of Urgent Referrals by Nurses at a Hospital-based Pediatric Call Center." *Archives of Pediatrics and Adolescent Medicine* 154: 355–360.

Kostami, V., and S. Rajagopalan. 2014. "Speed-quality Trade-offs in a Dynamic Model." *Manufacturing & Service Operations Management* 16 (1): 104–118.

Lafferty, S., and M. Baird. 2001. *Tele-nurse: Telephone Triage Protocols*. Clifton Park, NY: Delmar Cengage Learning.

Machlin, S. R. 2006. "Expenses for a Hospital Emergency Room Visit, 2003." Agency for Healthcare Research and Quality. Accessed July 23, 2007. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st111/stat111.pdf .

Mandelbaum, A., and N. Shimkin. 2000. "A Model for Rational Abandonments from Invisible Queues." *Queueing Systems* 36: 141–173.

Massaro, K. 2009. Triage Nurse and QI Coordinator for The Children's Hospital of Denver, Aurora, CO, Personal Communication to A. Scheller-Wolf and L. Debo, September 17, 2009.

McCaig, L. F., and C. W. Burt. 2003. *National Hospital Ambulatory Medical Care Survey: 2003 Emergency Department Summary*. Tech. Rep. 358. Division of Health Care Statistics, Centers for Disease Control and Prevention, Department of Health and Human Services.

Mehrotra, V., K. Ross, G. Ryder, and Y. Zhou. 2012. "Routing to Manage Resolution and Waiting Time in Call Centers with Heterogeneous Servers." *Manufacturing & Service Operations Management* 14 (1): 66–81.

Ni, G., Y. Xu, and Y. Dong. 2013. "Price and Speed Decisions in Customer-intensive Services with Two Classes of Customers." *European Journal of Operational Research* 228 (2): 427–436.

Poole, S. R. 2003. *Developing a Telephone Triage and Advice System for a Pediatric Office Practice during Office Hours and/or after Hours: The Complete Guide*. Elk Grove Village, IL: American Academy of Pediatrics.

Pot, A., S. Bhulai, and G. Koole. 2008. "A Simple Staffing Method for Multiskill Call Centers." *Manufacturing & Service Operations Management* 10 (3): 421–428.

Reisman, Anna B., and David L. Stevens. 2002. *Telephone Medicine: A Guide for the Practicing Physician*. Philadelphia, PA: American College of Physicians – American Society of Internal Medicine.

Shen, H., and J. Z. Huang. 2008. "Interday Forecasting and Intraday Updating of Call Center Arrivals." *Manufacturing & Service Operations Management* 10 (3): 391–410.

The Quality Improvement of Literacy, TeleCare, and Self Help (QUILTS) Collaborative. 2006. *BC HealthGuide Program Evaluation: Final Report*. *Evaluation of Findings to June 2005*. Tech. rep.

Tong, C., and S. Rajagopalan. 2014. “Pricing and Operational Performance in Discretionary Services.” *Production and Operations Management* 23 (4): 689–703.

Valanis, B., C. Tanner, S. R. Moscato, S. Shapiro, S. Izumi, M. David, C. Keyes, and A. Mayo. 2003. “A Model for Examining Predictors of Outcomes of Telephone Nursing Advice.” *Journal of Nursing Administration* 33 (2): 91–95.

de Véricourt, F., and Y. P. Zhou. 2005. “Managing Response Time in a Call-routing Problem with Service Failure.” *Operations Research* 53 (6): 968–981.

Wallace, R. B., and W. Whitt. 2005. “A Staffing Algorithm for Call Centers with Skill-based Routing.” *Manufacturing & Service Operations Management* 7 (4): 276–294.

Wang, X., L. G. Debo, A. Scheller-Wolf, and S. F. Smith. 2010. “Design and Analysis of Diagnosistic Service Centers.” *Management Science* 56 (11): 1873–1890.

Wang, X., L. G. Debo, A. Scheller-Wolf, and S. F. Smith. 2012. “Service Design at Diagnostic Service Centers.” *Naval Research Logistics* 59 (8): 613–628.

Wheeler, S. Q. 2003. *Telephone Triage Protocols: for Adult and School-age Populations with Women’s Health and Infant/Child Protocols*. Gaithersburg, MD: Aspen.

Wheeler, S., and B. Siebelt. 1997. “Calling all Nurses – How to Perform Telephone Triage.” *Nursing* 27 (7): 37–41.

Whitt, W. 1992. “Understanding the Efficiency of Multi-server Service Systems.” *Management Science* 38 (5): 708–723.

Wooff, D. A., and S. G. Stirling. 2014. “Practical Statistical Methods for Call Centres with a Case Study Addressing Urgent Medical Care Delivery.” *Annals of Operations Research*. doi:10.1007/s10479-014-1529-2.

Zhan, D., and A. R. Ward. 2014. “Threshold Routing to Trade Off Waiting and Call Resolution in Call Centers.” *Manufacturing & Service Operations Management* 16 (2): 220–237.