OR Spectrum (2012) 34:371–390
DOI 10.1007/s00291-010-0228-x

REGULAR ARTICLE

Efficiency evaluation for pooling resources in health care

Peter T. Vanberkel · Richard J. Boucherie · Erwin W. Hans · Johann L. Hurink · Nelly Litvak

Published online: 26 September 2010

Abstract Hospitals traditionally segregate resources into centralized functional departments such as diagnostic departments, ambulatory care centers, and nursing wards. In recent years this organizational model has been challenged by the idea that higher quality of care and efficiency in service delivery can be achieved when services are organized around patient groups. Examples include specialized clinics for breast cancer patients and clinical pathways for diabetes patients. Hospitals are struggling with the question of whether to become more centralized to achieve economies of scale or more decentralized to achieve economies of focus. In this paper we examine service and patient group characteristics to study the conditions where a centralized model is more efficient, and conversely, where a decentralized model is more efficient. This relationship is examined analytically with a queuing model to determine the most influential factors and then with simulation to fine-tune the results. The trade-offs between economies of scale and economies of focus measured by these models are used to derive general management guidelines.

Keywords Slotted queueing model · Simulation · Resource pooling · Focused factories · Health care modeling

P. T. Vanberkel (✉) · E. W. Hans
Operational Methods for Production and Logistics, School of Management and Governance,
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
e-mail: p.t.vanberkel@utwente.nl

P. T. Vanberkel · R. J. Boucherie · J. L. Hurink · N. Litvak
Department of Applied Mathematics, Faculty of Electrical Engineering,
Mathematics and Computer Science, University of Twente,
P.O. Box 217, 7500 AE Enschede, The Netherlands

1 Introduction

Health care facilities are under mounting pressure to both improve the quality of care and decrease costs by becoming more efficient. Efficiently organizing the delivery of care is one way to decrease cost and improve performance. At the national level this is achieved by aggregating services into large general hospitals in major urban centers, thereby gaining efficiencies through economies of scale (EOS). At the same time, some hospitals are becoming more specialized and offer a limited range of services aiming to breed competence and improve service rates (Leung 2000). Such strategies aim to improve performance through focus.

At the hospital level, similar strategies to exploit focus are being considered (Tiwari and Heese 2009; Schneider et al. 2008). Rather than organizing departments around function (e.g., radiology, phlebotomy, etc.), departments dedicated to treating a particular patient population are being created. Examples include focused departments for back patients (Wickramasinghe 2005), cancer patients (Vanberkel et al. 2010; Langabeer and Ozcan 2009), outpatients (McLaughlin et al. 1995), trauma patients (Hyer et al. 2009) and inpatients (Wolstenholme 1999; Huckman and Zinner 2008). In these studies the benefits of increased focus have shown mixed results, leading to confusion over whether to become more centralized to achieve EOS or more decentralized to achieve economies of focus (EOF). In this paper we formulate a model to measure and compare the performance of both settings. More specifically we examine service and patient population characteristics to determine under which circumstances the functional department, and conversely the patient focused department provides better patient access times.

The paper is organized as follows. Section 2 introduces the principles of pooling and focus and frames the debate between centralized and decentralized departments. Using this background information, the motivation and focus of the paper is further clarified in Sect. 3. Section 4 introduces the model used to measure the EOS lost in an unpooled system. Section 5 describes a rough analytical approximation used to identify the main factors influencing these losses. In Sect. 6, results from simulation experiments are used to provide further perspective on these factors, to fine-tune the results and to evaluate the accuracy of the approximation. Section 7 summarizes the results and provides guidelines for hospital managers. Section 8 briefly discusses potential future research.

2 The principles of pooling and focus

The pooling principle as described in Cattani and Schmidt (2005), is the, "pooling of customer demands, along with pooling of the resources used to fill those demands" in order to "yield operational improvements." This implies that a centralized (pooled) clinic that serves all customer types may achieve shorter waiting times than a number of decentralized (unpooled) clinics focusing on a more limited range of customer types. The intuition for this principle is as follows. Consider the situation in the unpooled setting, when a customer is waiting in one queue while a server for a different queue is free. Had the system been pooled in this situation, the waiting customer

could have been served by the idle server, and thus experience a shorter waiting time. The gain in efficiency is a form of EOS.

Statistically, the advantage of pooling is credited to the reduction in variability due to the portfolio effect (Hopp and Spearman 2001). This is easily demonstrated for cases where the characteristics of the unpooled services are identical. For this discussion see Joustra et al. (2010), van Dijk (2000), van Dijk and van der Sluis (2009), Ata and van Mieghem (2009). However, pooling is not always of benefit. There may be situations where the pooling of customers actually adds variability to the system thus offsetting any efficiency gains, see van Dijk and van der Sluis (2004). Furthermore when the target performances of customer types differ it may be more efficient to use dedicated capacity (i.e. unpooled capacity), see Joustra et al. (2010), Blake et al. (1996). And finally, in the pooled case all servers must be able to accommodate all demand. This flexibility may be expensive and, as is more directly related to this paper, may actually cause inefficiencies as servers are no longer able to focus on a single customer type.

The principle of focus advocates for departments to limit the range of services they offer in order to reduce complexity and allow the department to concentrate on doing fewer things more efficiently. This philosophy has been the basis for operating modern manufacturing plants which are often referred to as focused factories. Skinner (1985) argues that focus, simplicity and repetition in manufacturing breeds competence. The gain in efficiency due to focus is referred to in this paper as EOF.

To exploit the principle of focus in health care, it is suggested that hospitals aggregate patients with similar diagnoses together into dedicated departments (Hyer et al. 2009). For example the principle of focus recommends that hospitals eliminate a centralized phlebotomy department and instead have phlebotomy services located in or near diagnosis based care department. By locating all the patient services in one department or area reduces the complexity of the process and allows care givers to oversee the complete care process from start to finish.

It is clear that pooling is offered as a potential method to improve a system's performance without adding additional resources. Interestingly, the principle of focus which "advocates for hospitals to abandon functional, discipline-focused departments (e.g., radiology, nursing, etc.) in favor of a design organized around patients and their diagnoses" (Hyer et al. 2009; Kremitske and West 1997; Newman 1997), implies the same. In this paper we aim to enhance understanding of these seemingly contradictory view points in health care.

Other service industries have considered whether or not (or to which extent) resources should be pooled. van Dijk and van der Sluis (2004) show that general perceptions regarding the benefits of pooling in call centers may not be in line with results from queueing theory literature. A number of practical and theoretical scenarios encountered in call centers are considered and compared numerically by van Dijk and van der Sluis (2009). Pooling of resources in the courier industry is considered by Ata and van Mieghem (2009) where the authors use Brownian approximation models to contrast approaches used by two competing firms to provide regular and express courier services. Pooling has been studied outside of the practical domain to obtain general results. Mandelbaum and Reiman (1998) considers stations in a Jackson network of queues and encourages practitioners to take care when making pooling decisions as the effect (good or bad) can be unbounded. Whitt (1999) uses approximations for $M/G/s$

queueing systems to compare various splits of pooled systems. For more detailed reviews of pooling literature see van Dijk and van der Sluis (2004), Mandelbaum and Reiman (1998), Ata and van Mieghem (2009).

Pooling resources to serve homogeneous demand is the common example used to illustrate the benefits of pooling. In practice however, demand tends to be heterogeneous, in which case, these benefits are not guaranteed. Further complicating the study of the pooling of heterogeneous demands is that they tend to be analytically intractable and therefore approximate analysis is the norm (Ata and van Mieghem 2009). Finally, most models consider continuous systems, and as discussed in Sect. 3, the clinics studied in this paper are not continuous. In this paper and in general, the terms *pooled* and *centralized* are analogous when describing the makeup of a department or clinic. In the same way, the terms *unpooled*, *decentralized* and *focused* are analogous for describing the opposite makeup.

3 Motivation and scope

An initial case study (Vanberkel et al. 2010) which provides the motivation for this paper, was completed at the Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital (NKI–AVL). The hospital is considering the use of focused factories to treat patients with similar diagnoses. From a patient satisfaction perspective this setup is preferred, however, hospital managers want to know whether additional resources are required to compensate for any losses caused by unpooling the functional departments. Using a simulation approach, the case study offered a methodology for determining resource requirements in focused factories. This allowed the hospital to compare the performance of existing functional departments with focused factory proposals.

From the case study it became apparent that numerous clinic attributes influence the losses from unpooling, such as appointment length, clinic load, number of rooms, patient demand, etc. Furthermore, many of these attributes are interrelated meaning that identifying one attribute's influence in isolation from the others was an extremely difficult task using simulation. The approach was robust but the results were specific to each problem instance. In this paper, we combine results from an analytical model and a simulation model to derive more general results.

Comparing the efficiency of the two clinic makeups requires a definition for efficiency. In this paper, to be consistent with the goals and constraints of the proposed focused factories at NKI–AVL, access time is the main measure of efficiency. Access time is influenced by two things, the arrival rate of new patients and the throughput of the clinic. Naturally, the arrival rate is assumed to be the same regardless of the clinic makeup. However the throughput of patients depends on the clinic makeup. Focused clinics are more specialized with standard practices, specialized equipment, etc., typically leading to shorter and less variable appointment durations. However, they are smaller and have less EOS than their pooled counterpart. The analytical and simulation models described in this paper evaluate the efficiency of both clinic makeups while reflecting the different throughput expected from each. Specifically, the models approximate the appointment length for the unpooled system that achieves the same access time as in the equivalent pooled system. This improved service time represents the amount of improvement due to focus (or EOF) necessary to offset the losses

of EOS. The approximation, along with simulations of typical clinic environments, provides the insight from which we develop general management guidelines.

The model and framework can represent any hospital department where the service time is less than one day and where the system empties between days. This includes outpatient clinics, diagnostic clinics and operating theaters. Since these departments empty at night, continuous time queueing models, which are typically used to study the effects of pooling, are not appropriate. In place of a continuous time model, a discrete time slotted queueing model is used. To our knowledge such a robust model for measuring the effects of pooling and unpooling has not been developed before.

4 Model

A discrete time slotted queueing model is used to evaluate the tradeoff between EOS and EOF. We describe the queueing model using language from an ambulatory clinic setting. For example, referrals for appointments are considered new arrivals, appointment length is the service time, the number of consultation rooms reflects the number of servers and finally, the time a patient must wait for a clinic appointment (often referred to as access time in health care literature) is the waiting time in the queue. In this paper we use the following notation:

λ = Average demand for appointments per day
D = Average appointment length in minutes
V = Variance of the appointment length
C = Coefficient of variation for the appointment length $\left(C = \sqrt{V/D^2}\right)$
M = Number of rooms
ρ = Utilization of the rooms
t = Working minutes per day
W = Expected waiting time in days.

A subscript "AB" corresponds to the pooled case and a subscript "A" or "B" corresponds to the unpooled case for patient groups "A" or "B" respectively. The schemes of the pooled and unpooled systems are shown in Fig. 1.

When combined, the parameters of the unpooled system must equal the parameters of the pooled system. The parameters of the two patient groups describe the patient mix. How the patient mix parameters in the unpooled system relate to the parameters in the pooled system is described below.

$$M_{AB} = M_A + M_B \tag{1}$$
$$\lambda_{AB} = \lambda_A + \lambda_B \tag{2}$$
$$D_{AB} = qD_A + (1-q)D_B \tag{3}$$
$$V_{AB} = q(V_A + D_A^2) + (1-q)(V_B + D_B^2) - D_{AB}^2 \tag{4}$$

where $q = \lambda_A/\lambda_{AB}$.

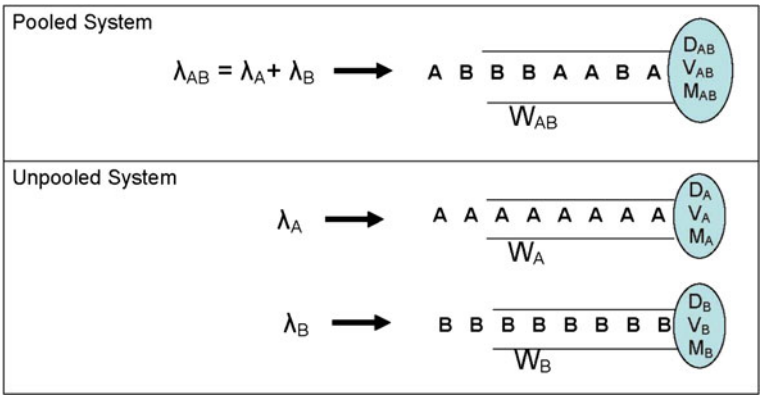

Fig. 1 Scheme of the pooled and unpooled systems

These division "rules" imply that no additional resources become available in the unpooled setting and that patients are strictly divided into one or the other group. Although we limit our analysis to splitting a department into two groups, the results are general. This is true since splitting a department into more than two groups, can be seen as splitting the original department into two groups, then splitting the resulting groups into two additional groups and so on.

Initially the waiting times in the three queueing systems depicted in Fig. 1 are evaluated separately. The structure of the three systems is the same and as such the same model is used to evaluate them (the input parameters are changed to reflect the pooled and unpooled systems). The approach used to evaluate the waiting times is described in Sects. 4.1 and 4.2, where the subscripts "A", "B" and "AB" are left out for clarity. In Sect. 4.3 we introduce a metric to compare the waiting time of the pooled and unpooled systems.

4.1 Modeling arrivals, services, and workload

The mean (D) and variance (V) of appointment lengths is readily available in most ambulatory clinics. Relying only on these data, we use renewal theory approximations to estimate the number of appointments completed during one clinic day. Let $N(t)$ be the number of appointments that fit into the schedule of one room between $[0, t]$. In fact, $N(t)$ is a renewal process with interarrival times distributed as appointment lengths. Further, let M be the number of rooms and $N_i(t)$ the number of completed appointment in room $i = 1, \ldots, M$. We assume that $N_i(t)$s are independent and distributed according to $N(t)$. Let S be the total number of completed appointments per clinic day for a clinic with M rooms. Then:

$$S = \sum_{i=1}^{M} N_i(t). \tag{5}$$

We assume that the number of arrivals per day is Poisson distributed with parameter λ. Then $E[X] = \lambda$, $V_X = \lambda$ and $C_X^2 = 1/\lambda$, where V_X and C_X, are, respectively, the variance and the coefficient of variation of X.

Under the assumptions above the workload of the clinic (ρ) is computed by $\rho = \lambda/E(S) = \lambda/(ME[N(t)])$.

4.2 Waiting times

With the input parameters described above, our system is a *single* server system where the department as a whole is considered the server with capacity determined according to S. As such the expected queue length can be computed using Lindley's recursion (Cohen 1982). Consider subsequent days $1, 2, \ldots$, and let L_n be the queue length at the beginning of day n. Further, let X_n be the number of arrivals on day n, and S_n the number of services that can possibly be completed on day n. We assume that X_n and $S_n, n > 1$, are independent and distributed as described above. The number of appointment requests on day n is then $L_n + X_n$, and the dynamics of the queue length process is given by:

$$L_{n+1} = (L_n + X_n - S_n)^+; \quad n > 1 \tag{6}$$

where $x^+ = x$ if $x \geq 0$ and $x^+ = 0$ otherwise. When $E[X_n] < E[S_n]$ then for $n \to \infty$ the expectation of L_n converges to equilibrium, denoted by L (Cohen 1982).

To compute the expected waiting time W we use Little's Law ($W = L/\lambda$). A related model described in Vanberkel et al. (2010) explains how to compute the waiting time distribution through a similar recursion. In general, (6) is hard to solve analytically. A variety of techniques, such as Wiener–Hopf factorization, have been developed but they usually lead to explicit solutions only in special cases. In Sect. 5 we provide a rough two-moment approximation for the average waiting time (see (15)). In the experiments of Sect. 6 we compute the average waiting time with simulations.

4.3 Required change in service time

To compare the performance of the pooled and unpooled systems, we wish to determine a new appointment length (D'_{A}) required to make $W_{\mathrm{A}} = W_{\mathrm{AB}}$. As a standard measure we define Z_{A} as the proportional difference between D_{A} and D'_{A} (likewise for D'_{B} and Z_{B}). Ignoring the subscripts "A" and "B" we formally define Z as follows:

$$Z = \frac{D'}{D} - 1. \tag{7}$$

Z essentially measures the EOF needed to make the access time in the pooled and unpooled systems equal. Z can be both negative and positive. When Z is negative it represents the amount the appointment length must decrease (attributed to the increased focus on a single patient group) in order to overcome any EOS losses resulting from unpooling. When Z is positive it indicates that the appointment length can increase

and still maintain the same service level as in the pooled system. This happens when the number of rooms assigned to one of the patient classes is disproportionately large. Although practically less relevant, the positive Z value does help illustrate how the tradeoff between EOS and EOF is influenced by the distribution of rooms.

The convenience of metric Z is that the pooled and unpooled system can be compared without any additional input. Furthermore stakeholders can easily interpret its meaning and decide if it is possible to obtain the necessary EOF to justify changing to an unpooled setup. In the simulation experiments of Sect. 6, Z_{A} and Z_{B} are computed numerically. In order to identify the system parameters that affect Z_{A} most, in the next section we carry out a crude analysis to obtain a simple two-moment approximation (17) for Z_{A}.

5 Rough analytic approximation for Z_{A}

As Z_{A} depend on (6), which can only be obtained analytically in very special cases, we apply a simple two-moment approximation to get a rough idea about the influence of various system parameters on Z_{A}.

5.1 Two-moment approximation

To obtain the approximation formula for Z_{A}, we use asymptotic results from renewal theory, and thus we must assume that the appointment length is much shorter than the clinic day, i.e., $D \ll t$. Further, $N(t)$ in our model is a number of events on $[0, t]$ when times between events are independent identically distributed appointment lengths. Thus, by definition, $N(t)$ is a renewal process, and with $D \ll t$ it follows from renewal theory (Tijms 2003, p. 315) that:

$$E[N(t)] \approx \frac{t}{D} + \frac{1}{2}(C^2 - 1). \tag{8}$$

Here, obviously, t/D is the main term, and the last term is a correction which in fact, will be neglected in the approximation (15) for the waiting time.

Now, for the total possible number S of completed appointments, using (5) we obtain:

$$E[S] \approx ME[N(t)] \approx \frac{Mt}{D} + \frac{M}{2}(C^2 - 1). \tag{9}$$

Let $V_{N(t)}$ and V_S be the variance of $N(t)$ and S respectively. From Tijms (2003), the two-moment renewal theory approximation for $V_{N(t)}$ and V_S is as follows:

$$V_{N(t)} \approx \frac{V^2 t}{D^3} = \frac{C^2 t}{D} \tag{10}$$

$$V_S \approx MV_{N(t)} = \frac{MC^2 t}{D}. \tag{11}$$

We note that (8), (9), (10) and (11) are based on the assumption $D \ll t$. In a contrary situation (e.g., chemotherapy, where appointments may last half the day), the influence of D, V, C on S is not so direct and the above approximations cannot be used, but the general model is still valid (Vanberkel et al. 2010).

Using (9) we approximate the room utilization ρ as follows:

$$\rho \approx \frac{\lambda}{\frac{Mt}{D} + \frac{M}{2}(C^2 - 1)} = \frac{\lambda D}{Mt} \frac{1}{1 + \frac{D}{2t}(C^2 - 1)}. \tag{12}$$

From (12) we observe $1/(1 + \frac{D}{2t}(C^2 - 1)) \approx 1$ when $D \ll t$, which is true in our case. From this observation we introduce ρ_0 as an estimate of ρ and define it as follows:

$$\rho_0 = \frac{\lambda D}{Mt}. \tag{13}$$

The average queue length (L) in our slotted queueing model is analogous to the average waiting time of a GI/GI/1 queue because both are measured by Lindley's Recursion. In particular (6) corresponds to a GI/GI/1 queue with Poisson distributed service times and interarrival times distributed as S in (5). The waiting time of a GI/GI/1 queue can be approximated with the Allen–Cunneen approximation (Allen 1990) thus leading to an approximation for L in our slotted model. Using (9) and (11) we obtain $C_S^2 = V_S/(E[S])^2$ and write the approximation formula for L as:

$$\begin{aligned} L &\approx \lambda \frac{\rho}{1-\rho} \frac{C_S^2 + (1/\lambda)^2}{2} = \lambda \frac{\rho}{2(1-\rho)} \left(\frac{1}{\lambda} + \frac{MC^2 t}{D} \frac{1}{M^2 \left(\frac{t}{D} + \frac{1}{2}(C^2 - 1) \right)^2} \right) \\ &\approx \frac{\rho}{2(1-\rho)} \left(1 + \frac{C^2}{\rho_0} \right). \end{aligned} \tag{14}$$

Using Little's Law and (14) we approximate the expected waiting time by:

$$W \approx \frac{\rho}{2(1-\rho)\lambda} \left(1 + \frac{C^2}{\rho_0} \right). \tag{15}$$

If λ grows and ρ remains the same then we observe a decreasing waiting time, which is credited to the EOS. Indeed, if $\lambda \to \infty$, then proportional capacity growth results in $W = 0$, see e.g. Janssen et al. (2008) for the asymptotic analysis of a similar slotted model with S equal to a constant.

Using our estimation (15) for W, we can also estimate the Z values based on (7). First we assume $\rho_0 \approx \rho$ and define ρ_0' as the load in the unpooled clinic A with appointment length D_{A}'. Formally we define ρ_0' as follows:

$$\rho_0' = \frac{\lambda_{\mathrm{A}} D_{\mathrm{A}}'}{M_{\mathrm{A}} t}.$$

Next we set the waiting time approximations (15) for the pooled and unpooled system A equal to each other:

$$\frac{\rho_0'}{2(1-\rho_0')\lambda_A}\left(1+\frac{C_A^2}{\rho_0'}\right)=\frac{\rho_0}{2(1-\rho_0)\lambda_{AB}}\left(1+\frac{C_{AB}^2}{\rho_0}\right) \tag{16}$$

We also assume the servers are divided between the pooled and unpooled clinics in such a way that the clinic load remains the same. The load in the two clinics may not be exactly equal since M_{AB} and M_A must be integers. From this it follows:

$$\rho_0=\frac{D_{AB}\lambda_{AB}}{M_{AB}t}\approx\frac{D_A\lambda_A}{M_At}.$$

Finally, with algebra and by ignoring second order and higher terms of $(1-\rho_0)$ we solve (16) for D_A'/D_A to obtain:

$$Z_A=\frac{D_A'}{D_A}-1\approx\left(1-\frac{1+C_A^2}{1+C_{AB}^2}\frac{\lambda_{AB}}{\lambda_A}\right)(1-\rho_0). \tag{17}$$

Similarly (17) can be rewritten to obtain $Z_B=D_B'/D_B-1$. Using (4) it can be shown that either Z_A or Z_B in (17) is negative. This proves that splitting a pooled clinic will negatively impact the access time of at least one of the unpooled clinics.

While deriving formula (17) we made a number of simplifying assumptions and ignored second order and higher terms of $(1-\rho_0)$ and the first order and higher terms of D/t. Thus, one can expect that (17) gives an accurate approximation for Z_A only in some special cases, e.g., when ρ_0 is close to one. However, the main goal of deriving this formula is to reveal the main parameters that influence Z_A and to identify the relative importance of these parameters in reasonable hospital settings. To this end, our calculations show that ρ_0, λ_A/λ_{AB}, and $(1+C_A^2)/(1+C_{AB}^2)$ are the most influential factors. Furthermore, the absences of M_{AB} and D_{AB} in (17) implies that their influence is minimal. This is also confirmed by simulation experiments in Sect. 6.2.3. Thus, in the rest of the paper we focus on the most influential factors appearing in (17).

5.2 Approximation results for Z_A

To illustrate the relative importance of terms ρ_0, λ_A/λ_{AB}, and $(1+C_A^2)/(1+C_{AB}^2)$ in (17), consider the following typical ranges for each of them: $\rho_0 \in [0.7, 0.99]$; $\lambda_A/\lambda_{AB} \in [0.3, 0.7]$, as having values outside of this range implies a very small unpooled department which would be impractical (Vanberkel et al. 2010); $C_A^2, C_B^2 \in [0.5, 3]$. Note also that $(1+C_A^2)/(1+C_{AB}^2)$ depends on λ_A/λ_{AB} through (4). Table 1 shows twelve scenarios reflecting the border values of these three influential factors.

We clearly observe that when ρ_0 is large it dominates Z_A and appears to be the most influential factor. It follows that the busier the clinic is, the smaller the loss in EOS.

Table 1 Relative importance of factors influencing Z_A, according to (17)

No.	Clinic description	ρ_0	$\frac{\lambda_A}{\lambda_{AB}}$	$\frac{1+C_A^2}{1+C_{AB}^2}$	Z_A
1	Busy Clinic, $\lambda_A \gg \lambda_B$, $V_A \ll V_B$	0.99	0.7	0.32	0
2	Busy Clinic, $\lambda_A \gg \lambda_B$, $V_A = V_B$	0.99	0.7	1	−0.01
3	Busy Clinic, $\lambda_A \gg \lambda_B$, $V_A \gg V_B$	0.99	0.7	1.36	−0.01
4	Busy Clinic, $\lambda_A \ll \lambda_B$, $V_A \ll V_B$	0.99	0.3	0.17	0
5	Busy Clinic, $\lambda_A \ll \lambda_B$, $V_A = V_B$	0.99	0.3	1	−0.03
6	Busy Clinic, $\lambda_A \ll \lambda_B$, $V_A \gg V_B$	0.99	0.3	2.58	−0.08
7	Quite Clinic, $\lambda_A \gg \lambda_B$, $V_A \ll V_B$	0.7	0.7	0.32	0.16
8	Quite Clinic, $\lambda_A \gg \lambda_B$, $V_A = V_B$	0.7	0.7	1	−0.13
9	Quite Clinic, $\lambda_A \gg \lambda_B$, $V_A \gg V_B$	0.7	0.7	1.36	−0.29
10	Quite Clinic, $\lambda_A \ll \lambda_B$, $V_A \ll V_B$	0.7	0.3	0.17	0.13
11	Quite Clinic, $\lambda_A \ll \lambda_B$, $V_A = V_B$	0.7	0.3	1	−0.7
12	Quite Clinic, $\lambda_A \ll \lambda_B$, $V_A \gg V_B$	0.7	0.3	2.58	−2.28

This is consistent with van Dijk and van der Sluis (2009), who states that "pooling is not so much about pooling capacity but about pooling idleness" implying that unpooled systems with less idleness can expect less EOS gains when pooled. Next consider that a high value of λ_A/λ_{AB} forces $(1 + C_A^2)/(1 + C_{AB}^2)$ close to 1 diminishing the effect of $(1 + C_A^2)/(1 + C_{AB}^2)$ on Z_A. However, for the corresponding smaller group, this factor becomes increasingly important (see rows 9 and 10 from Table 1).

The main goal of deriving formula (17) is to reveal the main parameters that influence Z and their relative importance. In the next section we use simulation to fine-tune the results for Z in a wide range of realistic scenarios. Furthermore, in Sect. 6.3 we evaluate the accuracy of approximation (17), as compared to the simulated results, for the same range of scenarios.

6 Simulation experiments

To gain further perspective on the factors that influence the loss in EOS and to validate the inferences drawn from (17) a number of numeric experiments are conducted. Section 6.1 describes the Monte Carlo simulation and the range of the experiments. Section 6.2 provides and discusses the results of the experiments. Section 6.3 compares results of the simulation experiments with (17).

6.1 Simulation description

We model the appointment length as random variables with phase-type distributions (Tijms 2003; Fackrell 2009) where expectation and variance are fitted in the data. We opt for a two moment approximation, instead of a more involved distribution fit (e.g., empirical distribution), because mean and variance data for appointment lengths are

typically available. As such it is easily transferable to other settings and the likelihood of implementation is increased (Vanberkel et al. 2010).

If the appointment length duration has $C \leq 1$ then the appointment length is assumed to follow an Erlang(k,μ) distribution where $\mu = k/D$ and k is the best integer solution to $k = D^2/V$. The completed patients per day (S) is computed by considering that an Erlang(k,μ) distribution is equal to a sum of k independent exponential random variables (phases) with parameter μ and the number of such phases completed in t time units is Poisson with mean μt. It follows that $N(t) = \lfloor \text{Poisson}(\mu t)/k \rfloor$. If $C > 1$ the appointment length is assumed to follow a hyperexponential phase type distribution. The appointment length is distributed according to $p\text{Expo}(\mu_1)+(1-p)\text{Expo}(\mu_2)$ and the total number of complete patients per day (S) is computed by Monte Carlo Simulation where:

$$p = \frac{1}{2}\left(1 + \sqrt{\frac{C^2-1}{C^2+1}}\right), \quad \mu_1 = \frac{2p}{D}, \quad \mu_2 = \frac{2(1-p)}{D}.$$

With this service rate distribution and under the assumption that the arrival rate is Poisson distributed, the waiting time and Z values (as described in Sect. 4) are obtained by simulation. The average queue length, described by Lindley's Recursion, is determined by simulating 10,000 clinic days of which 100 are used as a warm up. Little's Law is used to compute the average waiting time. To compute the Z values, the input to the simulation is systematically changed and the output compared. More specifically, Z_A is computed by incrementally decreasing [or increasing] D_A by a small amount, until $W_A \leq W_{AB}$ [$W_A \geq W_{AB}$]. The percentage change (Z_B) for patient group B is computed in the same manner. All computations are automated with Microsoft Visual Basic. Each of the simulated scenarios is described by the patient mix and clinic environment as introduced below.

Patient mix: The patient mix is described by two factors: λ_A/λ_{AB}, and D_A/D_{AB}. The chosen values for λ_A/λ_{AB} are 0.3, 0.4, 0.5, 0.6, and 0.7. This represents the range of situations where patient group A is 30% [group B is 70%] of the pooled group up to the situation where group A is 70% [group B is 30%] of the pooled group. The chosen values for D_A/D_{AB} are 0.5, 1, 1.5 and 2 representing situations where the appointment length for Group A is half that of the pooled group, and up to and including the case, where it is two times longer. The appointment length of Group B can be computed easily from (3).

Clinic environments: To represent different clinic environments, the parameters for the pooled clinic are changed to represent busier clinics, smaller clinics, more variable clinics, etc. Specifically we change the values of parameters M_{AB}, D_{AB}, λ_{AB}, ρ_0, C_A and C_B. The scenarios considered are listed in Table 2 and are meant to encompass a wide range of typical clinic environments. The italicized values of Table 2 indicate the parameters which are changed relative to the Base Clinic, which is described in row 1 of Table 2.

Table 2 Parameters for different clinic environment scenarios

Clinic environments	M_{AB}	D_{AB}	λ_{AB}	ρ_0	C_A, C_B
1 Base Clinic	20	30	282	0.88	0.5, 0.5
2 Busier Clinic	20	30	*310*	*0.97*	0.5, 0.5
3 Smaller Clinic	*10*	30	*141*	0.88	0.5, 0.5
4 Shorter appointment lengths	20	*15*	*564*	0.88	0.5, 0.5
5 Higher appointment length variability	20	30	282	0.88	*2.0, 2.0*
6 Different coefficient of variance	20	30	282	0.88	*2.0*, 0.5

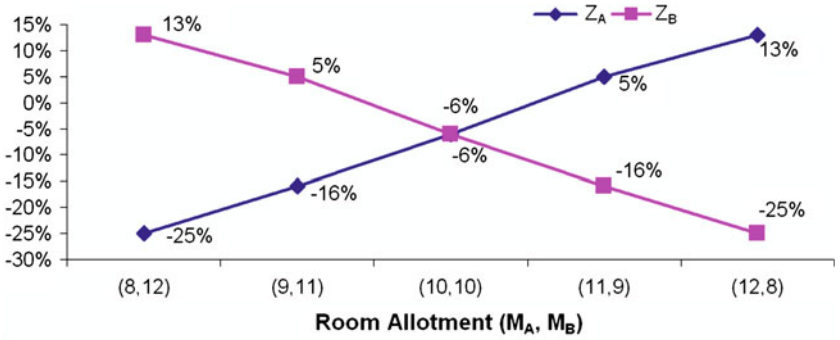

Fig. 2 Z values for various room allotments for the Base Clinic environment where $\lambda_A/\lambda_{AB} = 0.5$, and $D_A/D_{AB} = 1$

Server allotment: As discussed in Sect. 4 we wish to have the same total number of servers (rooms) in the unpooled system as in the initial pooled system. The number of rooms to allot to each of the unpooled clinics needs to be decided. To illustrate how this decision impacts Z_A and Z_B consider the results in Fig. 2 where the clinic environment is consistent with the Base Clinic and the patient mix parameters are $\lambda_A/\lambda_{AB} = 0.5$, and $D_A/D_{AB} = 1$.

As illustrated in Fig. 2, the smallest total loss in EOS corresponds with a room allotment of 10 rooms for each of the unpooled clinics. This is also the room allotment where the difference between ρ_{AB}, ρ_A and ρ_B is minimized. Let such a division be called the proportional room division, where $\rho_{AB} = \rho_A$ which implies:

$$\frac{\lambda_{AB} D_{AB}}{t M_{AB}} = \frac{\lambda_A D_A}{t M_A}$$

$$M_A = \frac{\lambda_A}{\lambda_{AB}} \frac{D_A}{D_{AB}} M_{AB}, \quad M_B = M_{AB} - M_A. \tag{18}$$

Practically speaking this division represents the most equitable way to divide the rooms such that the difference in workload for staff in the two unpooled clinics is minimized. For cases where $C_A = C_B$, it also represents the most equitable way to divide the rooms such that the difference in waiting time for both patient groups is minimized. The high degree by which Z depends on the room division is observable in all the

Table 3 Base Clinic results ($M_{AB} = 20$, $D_{AB} = 30$, $\lambda_{AB} = 282$, $C_A = C_B = 0.5$)

$\frac{\lambda_A}{\lambda_{AB}}$	$D_A/D_{AB} = 0.5$	$D_A/D_{AB} = 1.0$	$D_A/D_{AB} = 1.5$	$D_A/D_{AB} = 2.0$
0.3	−10%(3), −4%(17)	−12%(6), −4%(14)	−12%(9), −3%(11)	−12%(12), −2%(8)
0.4	−7%(4), −5%(16)	−9%(8), −5%(12)	−9%(12), −4%(8)	−2%(16), 6%(4)
0.5	−4%(5), −7%(15)	−6%(10), −6%(10)	−7%(15), −4%(5)	
0.6	−3%(6), −9%(14)	−5%(12), −8%(8)		
0.7	−2%(7), −13%(13)	−4%(14), −11%(6)		

evaluated clinic environments. For sake of brevity, in the following subsections, results are only provided for the proportional room divisions.

6.2 Experiment results

The results in this section are organized as follows. Initially the Base Clinic is analyzed for the various patient mixes. Then the clinic environment parameters are changed one-by-one and the results for each clinic environment are discussed in relation to the Base Clinic.

6.2.1 Base Clinic

The parameters and results for the initial Base Clinic environment are shown in Table 3. The patient mix factors λ_A/λ_{AB}, and D_A/D_{AB} represent the rows and columns respectively. The results in each table cell are in the following format: Z_A (M_A), Z_B (M_B). This represents the amount of change (Z_A) in D_A necessary, when the unpooled clinic is allotted M_A rooms (likewise for patient group B). As an example consider when $\lambda_A/\lambda_{AB} = 0.3$ and $D_A/D_{AB} = 0.5$. The value in the corresponding cell is "−10%(3), −4%(17)". As noted by the numbers is parentheses, this represents the case where three rooms are allotted to Group A and 17 to Group B. In this case, for the unpooled systems to perform equally as well as the pooled systems, Groups A and B are required to change their appointment length by $Z_A = -10\%$ and $Z_B = -4\%$ respectively. The blank cells in the table are a consequence of excluding room divisions which result in a $|Z|$ value greater than 25%.

From Table 3 and as identified in (17), Z depends on the ratio λ_A/λ_{AB}. When Group A is smaller than Group B (i.e. $\lambda_A/\lambda_{AB} < 0.5$), Group A requires less rooms but a greater decrease in service time. The counter situation (i.e., $\lambda_A/\lambda_{AB} > 0.5$) holds for Group B. It follows that larger patient groups retain EOS and require less EOF to compensate. Practically this implies that making a small department to serve a small patient population is not a good idea. This influence of λ_A/λ_{AB} is observable in all tables in this section.

Although not identified by (17), from Table 3 it appears that Z depends on the ratio D_A/D_B. This dependency is not easily characterized as it appears dependent on λ_A/λ_{AB}. Within the range of values tested, the influence of D_A/D_B is small relative

Table 4 Busier Clinic results ($M_{AB} = 20$, $D_{AB} = 30$, $\lambda_{AB} = 310$, $C_A = C_B = 0.5$)

$\frac{\lambda_A}{\lambda_{AB}}$	$D_A/D_{AB} = 0.5$	$D_A/D_{AB} = 1.0$	$D_A/D_{AB} = 1.5$	$D_A/D_{AB} = 2.0$
0.3	−4%(3), −3%(17)	−3%(6), −2%(14)	−6%(9), −2%(11)	−8%(12), −3%(8)
0.4	−3%(4), −3%(16)	−3%(8), −2%(12)	−5%(12), −2%(8)	2%(16), 6%(4)
0.5	−3%(5), −6%(15)	−2%(10), −2%(10)	−5%(15), −3%(5)	
0.6	−3%(6), −6%(14)	−2%(12), −3%(8)	−5%(18), −3%(2)	
0.7	−2%(7), −9%(13)	−2%(14), −3%(6)		

to that of λ_A/λ_{AB}. This is observable in all the tables in this section except Table 4 where the factor ρ_0 dominates.

6.2.2 Busier Clinic

To determine how Z_A and Z_B are influenced by how busy a clinic is, the demand for appointments is increased to $\lambda_{AB} = 310$. Comparing Table 3 with Table 4 it is clear that $|Z_A| + |Z_B|$ is decreasing as the clinic load increases. This means, that the EOS loss of unpooling is smaller for clinics of higher load. This is consistent with the findings from (17). In the remaining scenarios ρ_0 is kept constant with the Base Clinic.

6.2.3 Smaller Clinic and Clinics with shorter appointment lengths

As expected from (17), the results for the clinic with fewer rooms showed only modest changes in Z_A and Z_B and are therefore excluded from the text. However, it is important to note that in smaller pooled clinics, it is less likely that (18) will result in a near integer solution, hence there is a discretization effect. In (17) we assume $\rho_{0,AB} = \rho_{0,A}$ and overlook this influence. The tests for a clinic with shorter appointments found Z_A and Z_B to also be insensitive to D_{AB} which is again what is expected from (17).

6.2.4 Higher appointments length variability

Results for a clinic with higher appointments length variability are available in Table 5. Relative to the Base Clinic, C_A and C_B were both increased from 0.5 to 2. Contrasting Table 3 and Table 5 it is clear that $|Z_A| + |Z_B|$ has increased considerably with C_A and C_B. Although an increase was expected from (17) the extent of the increase is greater than anticipated. This leads to the conclusion that changes in C_A and C_B have a greater impact than (17) indicates. This is most easily illustrated by considering the patient mix $\lambda_A/\lambda_{AB} = 0.5$ and $D_A/D_{AB} = 1$ which represents the case where both patient groups have equal service rate and arrival rate parameters. Furthermore, the aggregate service rate for the pooled group also has the same parameters, see (3) and (4). As such, with this patient mix, C_{AB} always equals C_A and likewise C_B. In the simulation experiment for this patient mix, $|Z_A|$ increased by 4% when C_A and C_B

Table 5 Higher appointment length variability results ($M_{AB}=20$, $D_{AB}=30$, $\lambda_{AB}=282$, $C_A=C_B=2$)

$\frac{\lambda_A}{\lambda_{AB}}$	$D_A/D_{AB}=0.5$	$D_A/D_{AB}=1.0$	$D_A/D_{AB}=1.5$	$D_A/D_{AB}=2.0$
0.3	−22%(3), −5%(17)	−19%(6), −6%(14)	−17%(9), −7%(11)	−18%(12), −12%(8)
0.4	−18%(4), −8%(16)	−14%(8), −8%(12)	−13%(12), −11%(8)	−16%(16), −17%(4)
0.5	−15%(5), −11%(15)	−10%(10), −10%(10)	−11%(15), −15%(5)	
0.6	−14%(6), −14%(14)	−8%(12), −14%(8)	−9%(18), −22%(2)	
0.7	−13%(7), −19%(13)	−5%(14), −18%(6)		

Table 6 Different coefficient of variance results ($M_{AB}=20$, $D_{AB}=30$, $\lambda_{AB}=282$, $C_A=2$, $C_B=0.5$)

$\frac{\lambda_A}{\lambda_{AB}}$	$D_A/D_{AB}=0.5$	$D_A/D_{AB}=1.0$	$D_A/D_{AB}=1.5$	$D_A/D_{AB}=2.0$
0.3		−11%(6), 4%(14)	−14%(9), 3%(11)	−17%(12), 2%(8)
0.4	−23%(4), −5%(16)	−8%(8), 3%(12)	−11%(12), 2%(8)	−16%(16), −3%(4)
0.5	−5%(5), 2%(15)	−6%(10), 2%(10)	−9%(15), −2%(5)	
0.6	−4%(6), −2%(14)	−4%(12), −2%(8)	−5%(18), −24%(2)	
0.7	−4%(7), −5%(13)	−3%(14), −4%(6)		

were increased from 0.5 to 2. Evaluating (17) for the same situations shows no change in $|Z_A|$, illustrating that (17) does not fully capture the impact of C_A on $|Z_A|$.

6.2.5 Different coefficient of variance

Results for the scenario when $C_A = 2$ and $C_B = 0.5$ are shown in Table 6. Relative to the Base Clinic Z_A decreased and, with few exceptions, Z_B increases.

6.3 Comparison with analytic approximation

To evaluate the accuracy of approximation (17) and to determine in which situations it would provide accurate estimations for Z, we compare simulated results from this section with results computed according to (17). To this end, Table 7 lists the Z_A values for the six clinic environments as computed by simulation and by the approximation (the simulated Z_A values appear in parentheses). Since both the simulation and (17) found Z to be mostly insensitive to D_A/D_{AB}, we set $D_A/D_{AB} = 1$. Furthermore, since the purpose of this subsection is to compare the two approaches we only show the Z values for Group A. Due to the symmetry however, the Z_B values can also be derived from Table 7.

In the derivation of (17) we ignored second order and higher terms of $(1 - \rho_0)$ and therefore, as expected, (17) is quite accurate for larger values of ρ_0 and λ_A/λ_{AB}. This corresponds with the reasonably accurate results observed in Table 7 for the Busy Clinic environment and cases where the group size is proportionally large. In other

Table 7 Comparison of analytic approximation of Z_A with simulation experiments (simulated Z_A values appear in parentheses)

	$\frac{\lambda_A}{\lambda_{AB}} = 0.3$	$\frac{\lambda_A}{\lambda_{AB}} = 0.4$	$\frac{\lambda_A}{\lambda_{AB}} = 0.5$	$\frac{\lambda_A}{\lambda_{AB}} = 0.6$	$\frac{\lambda_A}{\lambda_{AB}} = 0.7$
Clinic environments					
1	−28%(−12%)	−18%(−9%)	−12%(−6%)	−8%(−5%)	−5%(−4%)
2	−7%(−3%)	−5%(−3%)	−3%(−2%)	−2%(−2%)	−1%(−2%)
3	−28%(−12%)	−18%(−9%)	−12%(−7%)	−8%(−5%)	−5%(−4%)
4	−28%(−10%)	−18%(−8%)	−12%(−6%)	−8%(−5%)	−5%(−3%)
5	−28%(−19%)	−18%(−14%)	−12%(−10%)	−8%(−8%)	−5%(−5%)
6	−72%(−11%)	−32%(−8%)	−16%(−6%)	−9%(−4%)	−5%(−3%)

cases simulation is a more appropriate method, especially if CV is different between the two patient groups, as in clinic environment 6.

6.4 Conclusions

From the analytic approximation of Z we conclude that when contemplating dividing a pooled department, managers should consider ρ, λ_A/λ_{AB}, and $(1+C_A^2)/(1+C_{AB}^2)$. The importance of all three of these factors is confirmed by the simulation experiments. In the simulation experiments we also find that Z_A and Z_B values appear slightly sensitive to the ratio D_A/D_B, although characterizing this influence is not observable from the results. Furthermore, with the simulation we identified how the division of rooms between the unpooled departments is also an important decision factor. Finally the simulation also illustrates the discretization effect that occurs in smaller clinics. Both approaches used to quantify the factors impacting the unpooling decisions illustrated that there are numerous considerations necessary and many cannot be considered in isolation. Table 8 summarizes these factors.

Besides mean waiting times, hospitals are also interested to waiting time norms (i.e., the percentage of patients waiting less than a given target). A recursion, similar to that of Lindley's can be formulated to determine the waiting time distribution (Vanberkel et al. 2010). Using this waiting time recursion (instead of the queue length recursion), the simulation experiments of this section could be repeated to determine the effects of pooling with respects to waiting time norms.

Finally, although not considered in this paper, partial pooling of resources may be a beneficial compromise to the strict resource pooling considered in this section. Partial resources pooling would see some resources dedicated to each group and the remaining resources shared between them (see van Dijk and van der Sluis 2009; Whitt 1999).

7 Implications for practice

In general, managers should consider the following when approaching the decision to unpool a centralized department. Under most circumstances access time to clinics will increase unless the service time in the unpooled department is decreased, assuming

Table 8 Summary of factors effecting EOS losses due to unpooling

Factors	Change in Z_A	General management guidelines
Clinic load (ρ_0)	Decreases as ρ_0 increases	Unpooling clinics with high load results in less EOS losses than clinics under lesser load
Room division	Disproportionate splits increase $\lvert Z_A\rvert + \lvert Z_B\rvert$	The room allotment representing the smallest loss in EOS occurs when the difference between ρ_{AB}, ρ_A and ρ_B is minimized, see (18)
Clinic size (M_{AB})	Increases (slightly) as M_{AB} decreases	EOS losses appear mostly insensitive to the size of the clinic. In smaller clinics it is more difficult to proportionally split servers
Appointment lengths (D_{AB})	Mostly insensitive to D_{AB}	EOS losses appear to be mostly insensitive to the length of the appointment
Appointment length variability (C_A, C_B)	Increases as C_A, C_B increases	Unpooling patient groups with highly variable appointment lengths results in larger EOS losses
Different appointment length variability ($C_A < C_B$)	Decreases when $C_A < C_B$	The patient group with the smaller C generally experiences a smaller loss in EOS as a result of unpooling
Proportional size of each group (λ_A/λ_{AB})	Increases as λ_A/λ_{AB} decreases	Smaller patient groups experience a greater loss in EOS as a result of unpooling
Appointment length proportion (D_A/D_{AB})	Mostly insensitive to D_A/D_{AB}	EOS losses appear to be mostly insensitive to the ratio of appointment lengths

that no additional resources are made available. The amount of service time decrease needed to compensate for this performance loss depends on the characteristics of the original pooled clinic and the characteristics of the newly created unpooled clinics. The main characteristics to consider are clinic load (ρ), proportional size of the patient groups (λ_A/λ_{AB}), bed division and variability in appointment length. Table 8 summarizes all factors considered in this paper.

When looking at the original pooled clinic consider the following. Clinics under high load require less decrease in service time to compensate for unpooling losses. The number of rooms in a clinic does not greatly influence the needed service time change, however in smaller clinics it is more difficult to proportionally divide the rooms.

When deciding how to split the pooled clinic (which consequently defines the characteristics of the new unpooled clinics) consider the following. The smallest required decrease in service time occurs when the difference between the clinic load in the two unpooled clinics is minimized. To compute the resource allocation that corresponds

to this bed division see (18). The smaller patient group resulting from the split will require a greater decrease in service time to compensate for unpooling losses. Finally, unpooling patient groups with highly variable appointment lengths also requires a greater decrease in service time to compensate.

For more specific results refer to the tables in Sect. 6 or apply the approach described in the same section. The approach used for developing these tables is versatile in terms of the application area and practical in that it requires only typical clinical data as input.

8 Future research

The analytic approximation provided initial insight into the influence of the many factors causing losses in EOS, however since it is an approximation it does not fully account for them. The simulation provided more accurate results for a given range of circumstances, and the approach is demonstrated to be robust. However, due to the large number of factors and the complex relationships that exist between them, it proved difficult to use simulation to draw stringent general conclusions. Further research is required to determine how exactly these factors influence losses of EOS related to unpooling. With comprehensive descriptions of these relationships, operational researchers can further improve or even optimize the mix of the functional and patient focused departments within a hospital.

References

Allen AO (1990) Probability, statistics and queueing theory. Academic Press, London

Ata B, van Mieghem JA (2009) The value of partial resource pooling: Should a service network be integratedor product-focused?. Manag Sci 55(1):115–131

Blake JT, Carter MW, Richardson S (1996) An analysis of emergency room wait time issues via computer simulation. INFOR 34:263–273

Cattani K, Schmidt GM (2005) The pooling principle. INFORMS Trans Educ **5**(2):17–24

Cohen JW (1982) The single server queue. In: North-Holland series in applied mathematics and mechanics, vol 8, 2nd edn. North-Holland Publishing Co., Amsterdam

Fackrell M (2009) Modelling healthcare systems with phase-type distributions. Health Care Manag Sci 12(1):11–26

Hopp WJ, Spearman ML (2001) Factory physics: foundations of manufacturing management. McGraw-Hill, Boston

Huckman RS, Zinner DE (2008) Does focus improve operational performance? Lessons from the management of clinical trials. Strateg Manag J 29(2):173–193

Hyer N, Wemmerlöv U, Morris J (2009) Performance analysis of a focused hospital unit: the case of an integratedtrauma center. J Oper Manag 27(3):203–219

Janssen A, van Leeuwaarden J, Zwart B (2008) Corrected asymptotics for a multi-server queue in the halfin-whitt regime. Queueing Syst 58(4):261–301

Joustra P, vander Sluis E, van Dijk N (2010) To pool or not to pool in hospitals: a theoretical and practical comparison for a radiotherapy outpatient department. Ann Oper Res 178(1):77–89

Kremitske DL, West DJ (1997) Patient-focused primary care: a model. Hosp Topics 75(4):22–28

Langabeer J, Ozcan Y (2009) The economics of cancer care: longitudinal changes in provider efficiency. Health Care Manag Sci 12(2):192–200

Leung GM (2000) Hospitals must become focused factories. Br Med J 320(7239):942
Mandelbaum A, Reiman MI (1998) On pooling in queueing networks. Manag Sci 44(7):971–981
McLaughlin CP, Yang S, van Dierdonck R (1995) Professional service organizations and focus. Manag Sci 41(7):1185–1193
Newman K (1997) Towards a new health care paradigm. Patient-focused care. The case of Kingston Hospital Trust. J Manag Med 11(6):357–371
Schneider JE, Miller TR, Ohsfeldt RL, Morrisey MA, Zelner BA, Li P (2008) The economics of specialty hospitals. Med Care Res Rev 65(5):531
Skinner W (1985) Manufacturing: the formidable competitive weapon. Wiley, New York
Tijms HC (2003) A first course in stochastic models. Wiley, New York
Tiwari V, Heese H (2009) Specialization and competition in healthcare delivery networks. Health Care Manag Sci 12(3):306–324
van Dijk NM (2000) On hybrid combination of queueing and simulation. In: Proceedings of the 2000 Winter simulation conference, pp 147–150
van Dijk N, van der Sluis E (2004) To pool or not to pool in call centers. Prod Oper Manag 17:296–305
van Dijk NM, van der Sluis E (2009) Pooling is not the answer. Eur J Oper Res 197(1):415–421
Vanberkel PT, Boucherie RJ, Hans EW, Hurink J, Litvak N (2010) Reallocating resources to focused factories: a case study in chemotherapy. In: Blake J, Carter M (eds) International perspectives on operations research and health care. Proceedings of the 34th meeting of the European Working Group on operational research applied to health services, pp 152–164
Whitt W (1999) Partitioning customers into service groups. Manag Sci 45(11):1579–1592
Wickramasinghe N, Bloemendal JW, De Bruin AK, Krabbendam JJ (2005) Enabling innovative healthcare delivery through the use of the focusedfactory model: the case of the spine clinic of the future. Int J Innov Learn 2(1):90–110
Wolstenholme E (1999) A patient flow perspective of UK health services: exploring the case for new "intermediate care" initiatives. Syst Dyn Rev 15(3):253–271