IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

POOLING RESOURCES IN INFRASTRUCTURE INDUSTRIES – AN EFFICIENCY EVALUATION

Ajay Kr. Tripathi, Dr. Pankaj Kumar Shrivastava
Research Scholar, Professor & HoD
Department of Mechanical Engineering
AKS University, Satna, Madhya Pradesh (India)

Abstract

Abstract: Infrastructure industries over a period of time become specialized in execution of different infrastructure projects viz. Highways, Railways, Aviation, Power Generation, Transmissions, Tunnels, Dams, Buildings and other mega projects. Organizations traditionally segregate resources (machineries, tools and plants) into centralized functional divisions as mentioned above. This organizational model is structured based on the fact of specific nature of jobs performed with a view of higher quality of work and efficiency. Since the human resources employed to these divisions do perform their specific jobs and can contribute to higher efficiency because of job specialization, but same is not the case with mechanical resources i.e. machineries, tools and plants which are dedicated to specific divisions, since mechanical resources perform similar kind of jobs irrespective of the divisions under which they are deployed. A question of whether to become more centralized to achieve economies of scale or more decentralized to achieve economies of focus always arises. Using Queuing Theory and Simulation models, we examine the service and work load characteristics to determine the conditions where a centralized model is more efficient and conversely where a decentralized model is more efficient. The result from the model measures the trade-offs between economies of scale and economies of focus from which administrative guidelines are derived.

Index Terms **– Infrastructure Industries, Mechanical Resources, Resource Pooling, Queuing Theory, Simulation Models**

I. INTRODUCTION:

Infrastructure industries are under mounting pressure to both improve the quality of work and decrease the cost by becoming more efficient. Efficiently organizing the execution of work is one way to decrease cost and improve performance. In the department this is achieved by aggregating the different type of work into general divisions thereby gaining efficiencies through economies of scale. At the same time some divisions are becoming more specialized and offer a limited range of work aiming to breed competence and improve execution of work (Leung 2000). Such strategies aim to improve performance through economies of focus.

At the organizational level similar strategies to exploit focus are being considered (Tiwari and Heese 2009; Schneider et al. 2008). Rather than organizing mechanical resources around all the divisions at organizational level i.e. pooled mechanical resources, the mechanical resources are dedicated to individual divisions thereby becoming more focussed to work of parent division (Wickramasinghe 2005; Vanberkel et al. 2010; Langabeer and Ozcan 2009; Mc Laughlin et al. 1995; Hyer et al. 2009; Wolstenholme 1999; Huckman and Zineer 2008). In this paper we examine the service and work load characteristics of different divisions at organizational level to determine the conditions where dedicated mechanical resources are more efficient and conversely where a common mechanical resources pool is more efficient.

We derive an analytic approximation measuring economy of scale losses associated with dedicated mechanical resources. This approximate along with simulations of typical divisional environments provide the insight from which we develop general management guidelines. The model relies only on typically available data and can easily be used to analyse specific work load environment of different divisions.

II. THE POOLING PRINCIPLE:

In this section we summarize the pooling principle (Cattani and Schmidt 2005) as pooling of work load requirement along with pooling of mechanical resources used to fill those requirements in order to yield operational improvements. This implies that a centralized division that serves all the different divisions at organizational level may achieve higher percentage utilization of resources than a number of different divisions with dedicated mechanical resources employed with in that division focussing on a limited work load.

The intuition for this principle is as follows. Consider the situation in a dedicated mechanical resource division under an un-pooled setting, when work is available for a machine due to machine being busy doing other job whereas a similar machine belonging to other division in the same organization is idle due to no work load available for that machine in that division. Had the mechanical resources been pooled i.e. centralized to be able to serve work load of all the divisions, the waiting work load of first division could have been served by the idle machine of other division in the same organization and thus experience no waiting time. The gain in the efficiency is a form of Economy of scale.

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 un-pooled mechanical resources are identical (Joustra et al. 2010; van Dijk 2000; van Dijk and van der Sluis 2009; Ata and van Mieghem 2009). However pooling is not always beneficial. There may be situations where the pooling of workloads actually adds variability to the system thus offsetting any efficiency gains (van Dijk and van der Sluis 2004). Furthermore when the target performances of work type differ it may be more efficient to use dedicated capacity (Joustra et al. 2010; Blake et al. 1996). And finally in the pooled case all mechanical resources must be able to accommodate all type of work load requirements.

It is recommended by researches to limit the range of services they offer in order to reduce complexity and allow the departments to concentrate on doing fewer things more efficiently. Focus, simplicity and repetition in manufacturing breeds competence (Skinner 1985; Hyer et al. 2009).

It is clear that pooling is offered as a potential method to improve a system's performance without any additional resource (Hyer et al. 2009; Kremitske and West 1997; Newman 1997). Many researchers have considered whether to pool or not to pool the resources. Researchers considered stations in a Jackson network of queues and encourages practitioners to take care when making pooling decisions as the effect can be unbound (Mandelbaum and Reiman 1998). An approximation for M/G/s queuing systems to compare various splits of pooled system has been considered (Whitt 1999).

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

III. MODEL:

A discrete time slotted queuing model is used to evaluate the trade-offs between economy of scale and economy of focus. More specifically the access time for a centralized division serving all type of work loads of all individual division of the organization is compared to the access time of decentralized individual divisions focussing on work load of that individual division only. Generally speaking the decentralized method results in longer access time due to the loss in economy of scale. The method quantifies this loss and computes the improvement in service time required in the decentralized divisions in order to achieve the equivalent access time as in the centralized division. This improved service time represents the amount of improvement due to focus necessary to offset the losses of economy of scale.

We describe the queuing model using language from any division of any organization involved in infrastructure projects. For example, requirements for machines for a work, placed by work in charge are considered as new arrivals, period of working by that machine is the service time, number of machines reflects the number of servers and the time a job must wait for a machine allotment, is the waiting time in the queue.

Following notations are used in this paper-

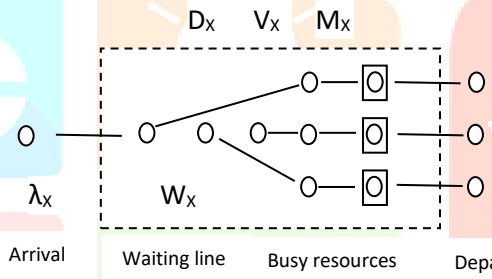
λ	=	average requirement of no. of jobs (Work load) per annum
D	=	Average time for completion of a work or work length in days
V	=	variance of the work length or period
C	=	coefficient of variance for the work length or period = $\sqrt{(V/D^2)}$
M	=	No. of machines
ρ	=	utilization of machines
t	=	working hours per day
W	=	expected waiting time in days

A subscript "XY" corresponds to the pooled case and a subscript "X" or "Y" correspond to the un-pooled case for division "X" or division "Y" respectively.

The schemes of the pooled and un-pooled system are shown in the following figures.

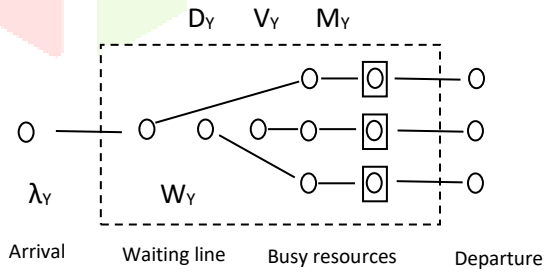
III.I Un-pooled System – X

Figure – 1 Un-pooled System - X



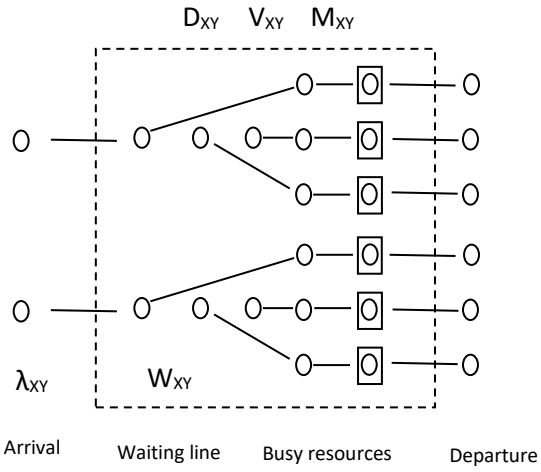
III.II Un-pooled System – Y

Figure – 2 Un-pooled System - Y



III.III Pooled System - XY

Figure – 3 Pooled System - XY



When combined, the parameters of the un-pooled system must equal the parameters of the pooled system. The parameters for the two divisions describe the division mix. How the division mix parameters in the un-pooled system relate to the parameters in the pooled system is described below. These division rules imply that no additional resources become available in the un-pooled setting and that divisions are strictly divided into one or the other group.

$$M_{XY} = M_X + M_Y \quad (1)$$

$$\lambda_{XY} = \lambda_X + \lambda_Y \quad (2)$$

$$D_{XY} = qD_X + (1-q)D_Y \quad (3)$$

$$V_{XY} = q(V_X + D^2_X) + (1-q)(V_Y + D^2_Y) - D^2_{XY} \quad (4)$$

where $q = \lambda_X / \lambda_{XY}$

Initially the waiting time in three queuing systems depicted in figure 1 are evaluated separately. The characteristics of the three systems are the same and as such the same model is used to evaluate them (the input parameters are changed to reflect the pooled and un-pooled systems). The model is described in the following subsections where the subscripts "X" and "Y" and "XY" are left out for clarity.

IV. MODELLING ARRIVALS AND SERVICES:

The mean (D) and variance (V) of work length in days is readily available in most divisions. Relying only on these data, we use renewal theory approximations to estimate the number of jobs completed during one year. We assume that D is average time for completion of a work or work length in days and that D << t. N (t) is defined as the number work or jobs completed in one division between [0, t]. Under these assumptions, from renewal theory (Tijms 2003) we find

$$E[N(t)] \approx \frac{t}{D} + \frac{1}{2}(C^2 - 1) \quad (5)$$

Let M be the number of machines, $N_i(t)$ the number of jobs completed by machine0 i=1,2….,M. We assume that $N_i(t)$s are independent and let S be the total number of completed jobs per division year given a division has M number of machines. Then

$$S = \sum_{i=1}^{M} Ni(t) \qquad E[S] \approx ME[N(t)] \approx \frac{Mt}{D} + \frac{M}{2}(C^2\text{-}1) \quad (6)$$

Renewal theory approximation implies that E[s] increases as C increases. Although perhaps counter intuitive, this means that as the variance in the division increases, so too do the number completed jobs per annum.

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

$$V_{N(t)} \approx \frac{V^2 t}{D^3} = \frac{C^2 t}{D} \quad (7)$$

$$V_S \approx MV_{N(t)} = \frac{MC^2 t}{D} \quad (8)$$

Equations (5), (6), (7) and (8) are based on the assumptions D << t.

In our model we assume that arrival process is Poisson. Let A be the arrival of work load per annum and V_A and C_A be the variance and coefficient of variance of A respectively. Since A is distributed according to Poisson (λ) it follows that E[A] = λ, $V_A = \lambda$ and $C_A = 1/\lambda$.

V. DIVISION LOAD:

Work load in a division is measured by utilization of its machines. The standard measure of machine utilization (ρ) is computed by ρ = λ/(ME[N(t)]). Using (6) we approximate ρ as follows

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

$$= \frac{\lambda D}{Mt} + \frac{\lambda D}{Mt}\{\sum_{i=1}^{\infty}(-1)^i\,(\frac{D}{2t}(C^2-1))^i\} \qquad (9)$$

Where the last equality holds provided $|D/(2t)(C^2-1)| < 1$, which is true in our cases since D<<t. The second term in the last expression of (9) is of the order D/t and since we assume that D<<t, it follows that it is small relative to the first term. From this observation we introduce ρ_0 as an estimate of ρ and define it as follows

$$\rho_0 = \frac{\lambda D}{Mt} \qquad (10)$$

In our simulation experiments we keep ρ_0 fixed for each setup. Because of the correction term in equation (9), actual ρ changes slightly depending on the patient mix parameters. For example if λ_X/λ_{XY} changes while C_X and C_Y remains constant, then C_{XY} must change according to equation (4). This consequently causes slight changes in E[S] and in turn in ρ.

VI. WAITING TIME:

With these input parameters the expected queue length is computed using Lindley's Recursion (Cohen 1982). Consider subsequent years 1, 2,….., and let L_n be the queue length at the beginning of the year n. further let A_n be the number of arrivals of jobs in year n, and S_n the number of jobs that can possibly be completed in year n. We assume that A_n and S_n, n>1 are independent and distributed as described above. The number of appointment requests in year n is then $L_n + A_n$, and the dynamics of the queue length process is given by

$$L_{n+1} = (L_n + A_n - S_n)^+ ; n > 1 \qquad (11)$$

Where $a^+ = a$ if $a \geq 0$ and $a^+ = 0$ otherwise.

If $n \to \infty$ then the expectation of L_n converges to its equivalent value L (Cohen 1982).

To compute the expected waiting time W we use Little's Law (W = L/λ). In general, equation (11) 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 the simulation experiments we solve (11) numerically.

The average queue length (L) in our slotted queuing model is analogous to the average waiting time of a GI/GI/1 queue because both are measured by Lindley's Recursion. The waiting time of a GI/GI/1 queue can be approximated with Allen-Cunneen approximation (Allen 1990) thus leading to an approximation for L in our slotted model. Using (6) and (8) and the assumption that D<<t, we write the approximation formula as

$$L \approx \lambda \frac{\rho}{1-\rho}\,\frac{1}{2}[C_S^2 + (1/\lambda)^2] = \lambda \frac{\rho}{2(1-\rho)}[\frac{1}{\lambda} + \frac{MC^2t}{D}\,\frac{1}{M^2(\frac{t}{D}+\frac{1}{2}(C^2-1))^2}]$$

$$\approx \frac{\rho}{2(1-\rho)}\,(1+\frac{C^2}{\rho_0}) \qquad (12)$$

Using Little's law and (12) we approximate the expected waiting by

$$W \approx \frac{\rho}{2(1-\rho)\lambda}\,(1+\frac{C^2}{\rho_0}) \qquad (13)$$

VII. REQUIRED CHANGE IN SERVICE TIME:

To compare the performance of the pooled and un-pooled systems, W is computed for the three queuing systems depicted in figure 1. The objective of the model is to determine a new appointment length D'_X required to make $W_X = W_{XY}$. As a standard measure we define Z_X as the proportional difference between D_X and D'_X (likewise for D'_Y and Z_Y). Ignoring the subscripts "X" and "Y" we formally define Z as follows

$$Z = \frac{D'}{D} - 1 \qquad (14)$$

Z essentially measures the Economy of Focus needed to make the access time in the pooled and un-pooled systems equal. Z can be both negative and positive. When Z is negative it represents the amount the work length must decrease in order to overcome any economy of scale losses resulting from un-pooling. When Z is positive it indicates that the work length can increase and still maintain the same service level as in the pooled system. Although practically less relevant, the positive Z value does help illustrate how the trade-off between economy of scale and economy of focus is influenced by the distribution of rooms.

Using our estimation (13) for W, we show how the Z values can also be estimated. First we assume $\rho_0 \approx \rho$ and define ρ'_0 as the load in the un-pooled division "X" with work length D'_X.

$$\rho'_0 = \frac{\lambda_X D'_X}{M_X t}$$

next we set the waiting time approximations (13) for the pooled and un pooled system "X" equal to each other.

$$\frac{\rho'_0}{2(1-\rho'0)\lambda_X}\left(1+\frac{C_X^2}{\rho'_0}\right) = \frac{\rho_0}{2(1-\rho_0)\lambda_{XY}}\left(1+\frac{C_{XY}^2}{\rho_0}\right) \qquad (15)$$

We also assume that the machines are divided between the pooled and un-pooled divisions in such a way that the division load remains the same. From this it follows

$$\rho_0 = \frac{D_{XY}}{M_{XY}}\frac{\lambda_{XY}}{t} \approx \frac{D_X}{M_X}\frac{\lambda_X}{t}$$

finally, with algebra and by ignoring second order and higher terms of $(1-\rho_0)$ we solve (15) for D'_X/D_X to obtain

$$Z_X = \frac{D'_X}{D_X} - 1 \approx \left(1 - \frac{1+C_X^2}{1+C_{XY}^2}\frac{\lambda_{XY}}{\lambda_X}\right)(1-\rho_0) \qquad (16)$$

Similarly (16) can be re written to obtain $Z_Y = D'_Y/B_Y - 1$. From 4 it can be shown that either Z_X or Z_Y in (16) is negative.

We note that while deriving formula (16) we made a number of simplifying assumptions and ignored second order and higher terms of (1- ρ_0). Thus one can expect that (16) gives an accurate approximation for Z_X only in some special cases, e.g., when ρ_0 is close to 1. The main goal of deriving this formula however is to reveal the main parameters that influence Z_X and to identify the importance of these parameters in reasonable division settings. To this end, our calculations show that ρ_0, λ_X/λ_{XY} and $(1+C^2_X)/(1+C^2_{XY})$ are the most influential factors. Furthermore, (16) also indicates which factor can be ignored. The absences of M_{XY} and D_{XY} implies that their influence is minimal.

Table – 1 Relative importance of factors influencing Z_X according to (16)

Sr. No.	Division description	ρ_0	λ_X/λ_{XY}	$(1+C^2_X)/(1+C^2_{XY})$	Z_X
1	Heavy work-load division, $\lambda_X >> \lambda_Y$, $V_X << V_Y$	0.99	0.70	0.32	0.00
2	Heavy work-load division, $\lambda_X >> \lambda_Y$, $V_X = V_Y$	0.99	0.70	1.00	-0.01
3	Heavy work-load division, $\lambda_X >> \lambda_Y$, $V_X >> V_Y$	0.99	0.70	1.36	-0.01
4	Heavy work-load division, $\lambda_X << \lambda_Y$, $V_X << V_Y$	0.99	0.30	0.17	0.00
5	Heavy work-load division, $\lambda_X << \lambda_Y$, $V_X = V_Y$	0.99	0.30	1.00	-0.03
6	Heavy work-load division, $\lambda_X << \lambda_Y$, $V_X >> V_Y$	0.99	0.30	2.58	-0.08
7	Normal work-load division, $\lambda_X >> \lambda_Y$, $V_X << V_Y$	0.70	0.70	0.32	0.16
8	Normal work-load division, $\lambda_X >> \lambda_Y$, $V_X = V_Y$	0.70	0.70	1.00	-0.13
9	Normal work-load division, $\lambda_X >> \lambda_Y$, $V_X >> V_Y$	0.70	0.70	1.36	-0.29
10	Normal work-load division, $\lambda_X << \lambda_Y$, $V_X << V_Y$	0.70	0.30	0.17	0.13
11	Normal work-load division, $\lambda_X << \lambda_Y$, $V_X = V_Y$	0.70	0.30	1.00	-0.70
12	Normal work-load division, $\lambda_X << \lambda_Y$, $V_X >> V_Y$	0.70	0.30	2.58	-2.28

Table 2: Percentage by which Z_X is overestimated by (16)

λ_X/λ_{XY}	$\rho_0 = 0.79$	$\rho_0 = 0.88$	$\rho_0 = 0.97$
0.3	40.60 %	18.10 %	4.10 %
0.4	22.10 %	9.80 %	1.50 %
0.5	13.10 %	6.30 %	1.00 %
0.6	10.40 %	3.10 %	0.00 %
0.7	5.20 %	1.10 %	0.00 %

To illustrate the relative importance of terms ρ_0, λ_X/λ_{XY} and $(1+C^2_X)/(1+C^2_{XY})$ in (16) , consider the following typical ranges for each of them: $\rho_0 \in [0.7, 0.99]$; $\lambda_X/\lambda_{XY} \in [0.3, 0.7]$, as having values outside of this range implies a very small un pooled department which would be impractical, C^2_X , $C^2_{XY} \in [0.5, 3]$. Note also that $(1+C^2_X)/(1+C^2_{XY})$ depends on λ_X/λ_{XY} through (4). Table 1 shows twelve scenarios reflecting the border values of three influential factors. We clearly observe that when ρ_0 is large it dominates Z_X and appears to be the most influential factor. It is also observable that busier the division is, smaller the loss in economy of scale. This is consistent with (7), which states that "pooling is not so much about pooling capacity but about pooling idleness" implying that un-pooled systems with less idleness can expect less economy of scale gains when pooled. Next consider that a high value of λ_X/λ_{XY} forces $(1+C^2_X)/(1+C^2_{XY})$ close to 1 diminishing the effect of $(1+C^2_X)/(1+C^2_{XY})$ on Z_X. However, for the corresponding smaller group, this factor becomes increasingly important (see rows 9 and 10 from table 1).

Finally table 2 illustrates the accuracy of approximation (16) by showing the per cent by which (16) overestimates Z_X compared with simulated results. Here the simulation results are obtained as described below. As expected, (16) is quite accurate for larger values of ρ_0 and λ_X/λ_{XY}, while for other cases the approximation is poor. Thus in the next section we obtain an accurate approximation for Z_X in a wide range of realistic scenarios, using computer simulations.

VIII. SIMULATION EXPERIMENTS:

To gain further perspective on the factors that influence the loss in economy of scale and to validate the inference drawn from (16) a number of numeric experiments are completed.

IX. SIMULATION DESCRIPTION:

IX.I Service rate distribution: We model the length of the job 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 job lengths are typically available. As such it is easily transferrable to other settings and the likelihood of implementation is increased.

If the job length duration has $C \leq 1$, then the job 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 jobs per annum 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) = [Poisson (μt)/k]. if C >1, the job length is assumed to follow a hyper exponential phase type distribution. The job length is distributed according to p Expo (μ_1) + (1-p) Expo (μ_2) and the total number of complete jobs per annum is computed by Monte Carlo simulation where

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

IX.II Job Mix: The job mix is described by two factors: λ_X/λ_{XY} and D_X/D_{XY}. The values for λ_X/λ_{XY} are 0.3, 0.4, 0.5, 0.6 and 0.7. This represents the range of situations where job group X is 70% , group Y is 30% of the pooled group. The values for D_X/D_{XY} are 0.5, 1, 1.5 and 2 representing situations where the job length for group X is half that of the pooled group, and up to and including the case, where it is two and half time longer. The job length of group Y can be computed easily from (3).

IX.III Machine allotment: Initially we do not impose restrictions on how to divide the machines between the two un-pooled systems as the optimal divisions follows from the model. To keep the experiments more manageable, results are limited to only "reasonable "machine allotments where $|Z_X|$ and $|Z_Y| \leq 0.25$. Practically this means we excluded situations where more than a 25% change in job length is required to make the performance of the un-pooled system equal the performance of the pooled system.

X. RESULTS:

The results in this section are organized as follows. Initially a base division is defined and analyzed for the various job mixes and machine allotments. Next the parameters for the pooled division are changed representing different division environments e.g. busy division, smaller division etc. The results for these different environments are compared to the base division. The scenarios considered in this section (as listed in table 3) are meant to encompass a wide range of typical division environments. The bold values of table 3 indicate the parameters which are changed relative to the base division.

Table 3: Parameters for different division environment scenarios

Division Environment	M_{XY}	D_{XY}	λ_{XY}	ρ_0	C_X,C_Y
Normal work-load Division	20	30	282	0.88	0.5,0.5
Heavy work-load Division	20	30	310	0.97	0.5,0.5
Smaller Division	10	30	141	0.88	0.5,0.5
Shorter Job Length	20	15	564	0.88	0.5,0.5
Higher Job Length Variability	20	30	282	0.88	2.0,2.0
Different Coefficient of Variance	20	30	282	0.88	2.0,0.5

Initial results for administrators may come from the division environment that most closely reflects their division's make up. For more specific results, the described simulation (which only requires the mean and variance data) should be used.

XI. BASE DIVISION:

The parameters and results for the initial base division environment are shown in table 4. The job mix factors λ_X/λ_{XY} and D_X/D_{XY} represent the rows and columns respectively. In each table cell multiple machine allotment (represented by the number in parenthesis) and the corresponding Z values are given. The results are in the following format: Z_X (M_X), Z_Y (M_Y). This represents the amount of change (Z_X) in D_X necessary, when the un-pooled division is allotted M_X machines (likewise for job group Y). As an example consider when $\lambda_X/\lambda_{XY} = 0.3$ and $D_X/D_{XY} = 0.5$. The values in the corresponding cell is "-10% (3), -4% (17)". The result represents the case where 3 machines are allotted to group X job and 17 to group Y job, as noted by the numbers on parenthesis. In this case for the un-pooled systems to perform equally as well as the pooled systems, group X and Y are required to change their job length by $Z_X = -10\%$ and $Z_Y = -4\%$ respectively. The blank cells in the table are the consequence of excluding machine divisions which result in a |Z| value greater than 25%. From Table 4 and as identified in (16), Z depends on the ratio λ_X/λ_{XY}. When group X is smaller than group Y (i.e. $\lambda_X/\lambda_{XY} < 0.5$), group X requires less machine but a greater decrease in service time. The counter situation (i.e. $\lambda_X/\lambda_{XY} > 0.5$) holds for group Y. It follows that larger Job groups retain economy of scale and requires less economy of focus to compensate. Furthermore the smallest total loss in economy of scale (i.e. $Z_X + Z_Y$) occurs when the two un-pooled divisions are of the same size. Practically this implies that making a small division to serve a small job population is not a good idea. This influence of λ_X/λ_{XY} is observable in all tables in this section.

Although not identified by (16), from table 4 it appears that Z depends on the ratio D_X/D_Y. This dependency is not easily characterised as it appears dependent on λ_X/λ_{XY}. Within the range of values tested, the influence of D_X/D_Y is small relative to that of λ_X/λ_{XY}. This is observable in all tables in this section except Table 5 where the factor ρ_0 dominates.

Table 4: Base Division Results (M_{XY} 20, D_{XY} = 30, λ_{XY} = 282, $C_X = C_Y$ = 0.5)

λ_X/λ_{XY}	D_X/D_{XY} = 0.5	D_X/D_{XY} = 1.0	D_X/D_{XY} = 1.5	D_X/D_{XY} = 2.0
0.3	**-10%(3), -4%(17)**	20%(8), -18%(12) 5%(7), -11%(13) **-12%(6), -4%(14)**	10%(11), -21%(9) -2%(10), -12%(10) **-12%(9), -3%(11)** -22%(8), 8%(12)	-5%(13), -14%(7) **-12%(12), -2%(8)** -20%(11), 12%(9)
0.4	19%(5), -12%(15) **-7%(4), -5%(16)**	16%(10), -21%(10) 5%(9), -13%(11) **-9%(8), -5%(12)** -20%(7), 5%(13)	0%(13), -15%(7) **-9%(12), -4%(8)** -16%(11), 10%(9)	6%(17), -22%(3) **-2%(16), 6%(4)**
0.5	17%(6), -12%(14) **-4%(5), -7%(15)**	4%(11), -16%(9) **-6%(10), -6%(10)** -16%(9), 5%(11)	**-7%(15), -4%(5)** -13%(14), 16%(6)	
0.6	15%(7), -15%(13) **-3%(6), -9%(14)** -19%(5), -3%(15)	5%(13), -20%(7) **-5%(12), -8%(8)** -13%(11), 5%(9) -21%(10), 15%(10)	-5%(18), -6%(2)	
0.7	14%(8), -19%(12) **-2%(7), -13%(13)** -16%(6), -6%(14)	**-4%(14), -11%(6)** -10%(13), 5%(7) -18%(12), 19%(8)		

The machine allotment which represents the smallest loss in economy of scale occurs when the difference between ρ_{XY}, ρ_X and ρ_Y is minimized. For ease of comparison, the results for these proportional machine distributions are bold. For such allotments $\rho_{0,XY} = \rho_{0,Y}$ which implies

$$\frac{\lambda_{XY}}{t}\frac{D_{XY}}{M_{XY}} = \frac{\lambda_X}{t}\frac{D_X}{M_X}$$

$$M_X = \frac{\lambda_X}{\lambda_{XY}}\frac{D_X}{D_{XY}}M_{XY}\ ,\ M_Y = M_{XY} - M_X \qquad (17)$$

Practically speaking this division represents the most equitable way to divide the machines such that the difference in workload for staff in the two un-pooled divisions is minimized. For cases where $C_X = C_Y$, it also represents the most equitable way to divide the machines such that the difference in waiting time for both work load is minimized. The high degree by which Z depends on the machine division is observable in all the tables in this section.

XII. BUSIER DIVISION:

To determine how Z_X and Z_Y are influenced by how busy a division is, the demand for work is increased to λ_{XY} = 310. Comparing table 4 with table 5, it is clear that $|Z_X| + |Z_Y|$ is decreasing as the clinic load increases. This means, that the EOS loss of un-pooling is smaller for divisions of higher work load. This is consistent with findings from (16). In the remaining scenarios ρ_0 is kept constant with the Base Case.

Table 5: Busier Division Results (M_{XY} 20, D_{XY} = 30, λ_{XY} = 310, $C_X = C_Y$ = 0.5)

λ_X/λ_{XY}	D_X/D_{XY} = 0.5	D_X/D_{XY} = 1.0	D_X/D_{XY} = 1.5	D_X/D_{XY} = 2.0
0.3	**-4%(3), -3%(17)**	15%(7), -9%(13) **-3%(6), -2%(14)** -19%(5), 7%(15)	17%(11), -20%(9) 7%(10), -11%(10) **-6%(9), -2%(11)** -16%(8), 9%(12)	1%(13), -15%(7) **-8%(12), -3%(8)** -15%(11), 12%(9)
0.4	**-3%(4), -3%(16)**	11%(9), -10%(11) **-3%(8), -2%(12)** -15%(7), 8%(13)	5%(13), -14%(7) **-5%(12), -2%(8)** -13%(11), 12%(9)	**2%(16), 6%(4)**
0.5	18%(6), -12%(14) **-3%(5), -6%(15)**	19%(12), -22%(8) 10%(11), -12%(9) **-2%(10), -2%(10)** -12%(9), 9%(11) -22%(8), 19%(12)	**-5%(15), -3%(5)** -12%(14), 18%(6)	
0.6	16%(7), -13%(13) **-3%(6), -6%(14)** -19%(5), 2%(15)	8%(13), -15%(7) **-2%(12), -3%(8)** -10%(11), 11%(9)	**-5%(18), -3%(2)**	
0.7	14%(8), -15%(12) **-2%(7), -9%(13)** -16%(6), -2%(14)	7%(15), -19%(5) **-2%(14), -3%(6)** -9%(13), 14%(7)		

XIII. SMALLER DIVISION AND DIVISIONS WITH SHORTER JOB LENGTH:

As expected from (16), the results for the division with fewer machines showed only modest changes in Z_X and Z_Y and are therefore excluded from the text. However, it is important to note that in smaller divisions, it is more likely that (17) results in a non-integer solution, hence there is discretization effect. In (16) we assume $\rho_{0,XY} = \rho_{0,X}$ and overlook this influence. The results for a division with shorter job lengths found Z_X and Z_Y to also be insensitive to D_{XY} which is again what is expected from (16).

XIV. HIGHER JOB LENGTH VARIABILITY:

Results for a division with higher job length variability are available in table 6. Relative to the base case, C_X and C_Y were both increased from 0.5 to 2. Contrasting table 4 and table 6 it is clear that $|Z_X| + |Z_Y|$ has increased considerably with C_X and C_Y .Although an increase was expected from (16) the extent of the increase is greater than anticipated. This leads to the conclusion that changes in C_X and C_Y have a greater impact than (16) indicates. This is most usually illustrated by considering job mix when λ_X/λ_{XY} = 0.5 and D_X/D_{XY} =1 which represents the case where both job groups have equal service rate and arrival rate parameters. Furthermore, the aggregate service rate for the pooled group also have the same parameters, see (3) and (4). As such, with this job mix, C_{XY} always equals C_X and likewise C_Y. In the simulation experiment for this job mix, $|Z_X|$ increased by 4% when C_X and C_Y were increased from 0.5 to 2. Evaluating (16) for the same situations shows no change in $|Z_X|$, illustrating that (16) does not fully capture the impact of C_X on $|Z_X|$.

Table 6: Higher Job length Variability Results (M_{XY} 20, D_{XY} = 30, λ_{XY} = 282, $C_X = C_Y$ = 2)

λ_X/λ_{XY}	D_X/D_{XY} = 0.5	D_X/D_{XY} = 1.0	D_X/D_{XY} = 1.5	D_X/D_{XY} = 2.0
0.3	8%(4), -11%(16) **-22%(3), -5%(17)**	14%(8), -20%(12) -4%(7), -13%(13) **-19%(6), -6%(14)**	6%(10), -17%(10) **-17%(9), -7%(11)**	**-18%(12), -12%(8)**
0.4	5%(5), -14%(15) **-18%(4), -8%(16)**	-2%(9), -16%(11) **-14%(8), -8%(12)**	**-13%(12), -11%(8)** -21%(11), 3%(9)	**16%(16), -17%(4)** -23%(15), 6%(5)
0.5	5%(6), -17%(14) **-15%(5), -11%(15)**	1%(11), -20%(9) **-10%(10), -10%(10)** -20%(9), 2%(11)	**-11%(15), -15%(5)** -16%(14), 5%(6)	
0.6	2%(7), -20%(13) **-14%(6), -14%(14)**	**-8%(12), -14%(8)** -16%(11), -3%(9)	**-9%(18), -22%(2)**	
0.7	**-13%(7), -19%(13)**	**-5%(14), -18%(6)** -13%(13), -5%(7) -20%(12), 13%(8)		

XV. DIFFERENT COEFFICIENT OF VARIANCE:

Results for the scenario when $C_X = 0.5$ and $C_Y = 2$ are shown in Table 7. Relative to the Base Case, Z_X decreased and with few exceptions, Z_Y faces almost no changes.

Table 7: Different Coefficient of Variance Results (M_{XY} 20, $D_{XY} = 30$, $\lambda_{XY} = 282$, $C_X = C_Y = 2$)

λ_X/λ_{XY}	$D_X/D_{XY} = 0.5$	$D_X/D_{XY} = 1.0$	$D_X/D_{XY} = 1.5$	$D_X/D_{XY} = 2.0$
0.3	**-5%(3), -5%(17)**	14%(7), -10%(13) **-4%(6), -3%(14)** -20%(5), 6%(15)	19%(11), -21%(9) 8%(10), -11%(10) **-4%(9), -2%(11)** -14%(8), 9%(12)	5%(13), -15%(7) **-5%(12), -2%(8)** -13%(11), 12%(9)
0.4	**-4%(4), -7%(16)**	12%(9), -13%(11) **-2%(8), -4%(12)** -14%(7), 6%(13)	9%(13), -16%(7) **1%(12), -3%(8)** -10%(11), 11%(9)	**-3%(16), -5%(4)** -9%(15), 20%(5)
0.5	20%(6), -16%(14) **-2%(5), -9%(15)** -21%(4), -3%(16)	12%(11), -16%(9) **-2%(10), -6%(10)** -10%(9), 6%(11) -20%(8), -16%(12)	**3%(15), -5%(5)** -6%(14), 17%(6)	
0.6	17%(7), -20%(13) **-1%(6), -13%(14)** -18%(5), -6%(15)	12%(13), -20%(7) **3%(12), -8%(8)** -7%(11), 7%(9) -15%(10), 19%(10)	-11%(17), -17%(3)	
0.7	**1%(7), -19%(13)** -15%(6), -12%(14)	**5%(14), -12%(6)** -5%(13), -6%(7)		

XVI. CONCLUSION:

From the analytic approximation of Z we conclude that when contemplating dividing a pooled department, administrators should consider ρ, λ_X/λ_{XY}, and $(1+C^2_X)/(1+C^2_{XY})$. The importance of all three of these factors is confirmed by the simulation experiments which also identified further factors for consideration. In the simulation experiments we find that Z_X and Z_Y values are influenced by C_X and C_Y. Z_X and Z_Y values also appear slightly sensitive to the ratio D_X/D_Y, although characterizing this influence is not observable from the results. Furthermore, with the simulation we identified how the division of machines between the un-pooled departments is also an important decision factor. Finally the simulation also illustrated the discretization effect that occurs in smaller divisions. Both approaches used to quantify the factors impacting the un-pooling decisions illustrated that there are numerous considerations necessary and many cannot be considered in isolation. In table 8 we summarize these factors.

Table 8: Summary of factors effecting economy of scale losses due to Un-Pooling

Factors	**Change in Z_X**	**General Administrative Guidelines**
Division Load(ρ_0)	Decreases as ρ_0 increases	Un Pooling divisions with high work load results in less economy of scale losses than divisions under lesser work load
Machine Division	Disproportionate splits increase $\|Z_X\| + \|Z_Y\|$	The machine allotment representing the smallest loss in economy of scale occurs when the difference between ρ_{XY}, ρ_X and ρ_Y is minimized, see (17)
Division Size (M_{XY})	Increases (slightly) as M_{XY} decreases.	Economy of scale losses appears mostly insensitive to the size of the division. In smaller division it is more difficult to proportionally split different machines (server)
Divisions with short work lengths (D_{XY})	Mostly insensitive to D_{XY}	Economy of scale losses appear to be mostly insensitive to the job length
Divisions with highly variable job lengths (C_X, C_Y)	Increases as C_X, C_Y increases	Un pooling work groups with highly variable work lengths results in larger economy of scale losses.
Divisions with different coefficient of variance for work groups ($C_X < C_Y$)	Decreases when $C_X < C_Y$	The job group with the smaller C generally experiences a smaller loss in economy of scale as a result of un pooling
Proportional size of each group (λ_X/λ_{XY})	Increases as λ_X/λ_{XY} decreases	Smaller job groups experience a greater loss in economy of scale as a result of un pooling
Job length proportion (D_X / D_{XY})	Mostly insensitive to D_X / D_{XY}	Economy of scale losses appear to be mostly insensitive to the ratio of work length

XVII. IMPLICATION FOR PRACTICE:

In general, administrators should consider the following when approaching the decision to un-pool a centralized department. Under most circumstances access time to divisions will increase unless the service time in the un-pooled department is deceased, assuming 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 division and the characteristics of the newly created un-pooled divisions. The main characteristics to consider are division load (ρ), number of machines (N_{XY}), and variability in job length. Table 8 summarizes all factors considered in this paper.

When looking at the original pooled division consider the following. Divisions under high load require less decrease in service time to compensate for un-pooling losses. The number of machines in a division does not greatly influence the needed service time change; however in smaller divisions it is more difficult to proportionally divide the machines.

When deciding how to split the pooled divisions (which consequently define the characteristics of the new un-pooled division) consider the following. The smallest required decrease in service time occurs when the difference between the division loads in the two un-pooled divisions is minimized. The smaller job group resulting from the split will require a greater decrease in service time to compensate for un-pooling losses. Finally, un-pooling job groups with highly variable job lengths also require a greater decrease in service time to compensate.

XVIII. FUTURE RESEARCH:

The analytic approximation provided initial insight into the influence of the many factors causing losses in economy of scale, 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 large number of factors and the complex relationships that exists 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 economy of scale related to un pooling. With comprehensive descriptions of these relationships, operational researchers can further improve or even optimize the mix of the functional and job focussed divisions within an organization.

REFERENCES

[1] Allen AO. 1990. Probability, statistics and queueing theory. Academic Press, London

[2] Ata B, van Mieghem JA. 2009. The value of partial resource pooling: Should a service network be integrated or product-focused? Management Science 55(1):115–131

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

[4] Cattani K, Schmidt GM. 2005. The pooling principle. INFORMS Transactions on Education 5(2):17–24

[5] 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

[6] Fackrell M. 2009. Modelling healthcare systems with phase-type distributions. Health Care Management Science 12(1):11–26

[7] Hopp WJ, Spearman ML. 2001. Factory physics: foundations of manufacturing management. McGraw-Hill, Boston

[8] Huckman RS, Zinner DE. 2008. Does focus improve operational performance? Lessons from the management of clinical trials. Strategic Management Journal 29(2):173–193

[9] Hyer N, Wemmerlöv U, Morris J. 2009. Performance analysis of a focused hospital unit: the case of an integrated trauma centre. Journal of Operational Management 27(3):203–219

[10] Janssen A, van Leeuwaarden J, Zwart B. 2008. Corrected asymptotic for a multi-server queue in the Halfin-whitt regime. Queueing Systems 58(4):261–301

[11] 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. Annals of Operation Research 178(1):77–89

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

[13] Langabeer J, Ozcan Y. 2009. The economics of cancer care: longitudinal changes in provider efficiency. Health Care Management Science 12(2):192–200

[14] Leung GM. 2000. Hospitals must become focused factories. Br Med J 320(7239):942

[15] Mandelbaum A, Reiman MI. 1998. On pooling in queueing networks. Management Science 44(7):971–981

[16] McLaughlin CP, Yang S, van Dierdonck R. 1995. Professional service organizations and focus. Management Science 41(7):1185–1193

[17] Newman K. 1997. Towards a new health care paradigm. Patient-focused care. The case of Kingston Hospital Trust. Journal of Management of Medicines 11(6):357–371

[18] 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

[19] Skinner W. 1985. Manufacturing: the formidable competitive weapon. Wiley, New York

[20] Tijms HC. 2003. A first course in stochastic models. Wiley, New York

[21] Tiwari V, Heese H. 2009. Specialization and competition in healthcare delivery networks. Health Care Management Science 12(3):306–324

[22] van Dijk NM. 2000. On hybrid combination of queueing and simulation. In: Proceedings of the 2000 Winter simulation conference, pp 147–150

[23] van Dijk N, van der Sluis E. 2004. To pool or not to pool in call centers. Production and Operations Management 17:296–305

[24] van Dijk NM, van der Sluis E. 2009. Pooling is not the answer. European Journal of Operational Research 197(1):415–421

[26] 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

[27] Whitt W. 1999. Partitioning customers into service groups. Management Science 45(11):1579–1592

[28] Wickramasinghe N, Bloemendal JW, De Bruin AK, Krabbendam JJ. 2005. Enabling innovative healthcare delivery through the use of the focused factory model: the case of the spine clinic of the future. International Journal of Innovative Learning 2(1):90–110

[29] Wolstenholme E. 1999. A patient flow perspective of UK health services: exploring the case for new "inter-mediate care" initiatives. Syst Dyn Rev 15(3):253–271