Efficiency evaluation for pooling resources in health care

Operational Methods for Production and Logistics
Department of Applied Mathematics
University of Twente, Enschede, The Netherlands
www.choir.utwente.nl

ORAHS July 13–17, 2009

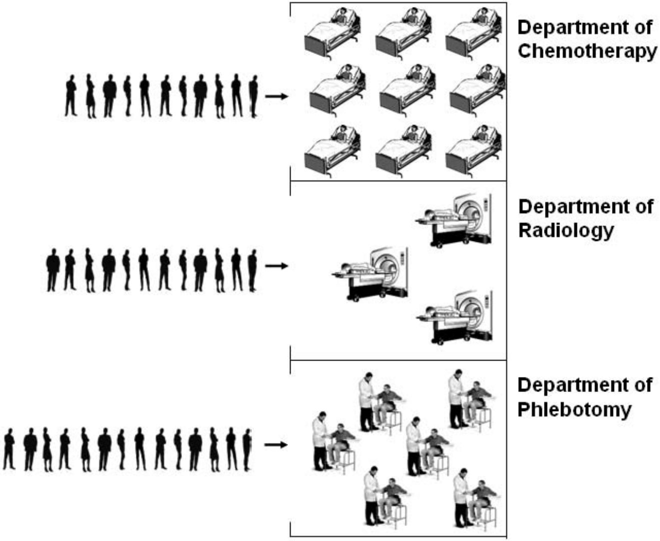
University of Twente
Enschede - The Netherlands

Presentation Outline

Specialization in Health Care

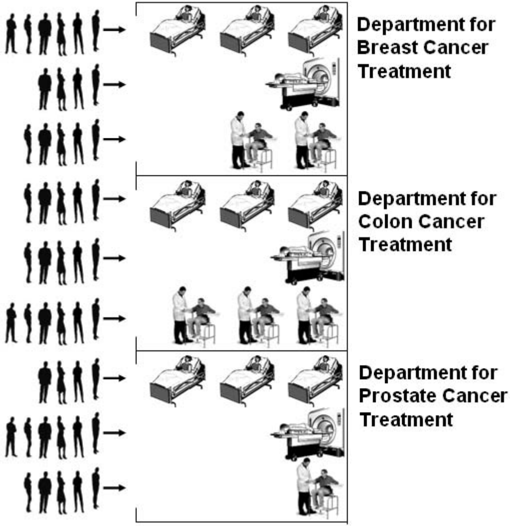
- National Level
  - Large General Hospitals
  - Specialty Hospitals
    - Cancer Hospitals
    - Pediatric Hospitals
    - Women's Hospitals
  - 'Cream Skimmers'
    - Shouldice Hernia Centre
    - Surgical Hospitals
    - Diagnostic Centres
- Hospital Level
  - Focused Clinics within Hospitals

Traditional Functional Departments



University of Twente
Enschede - The Netherlands

Diagnosis Focused Departments



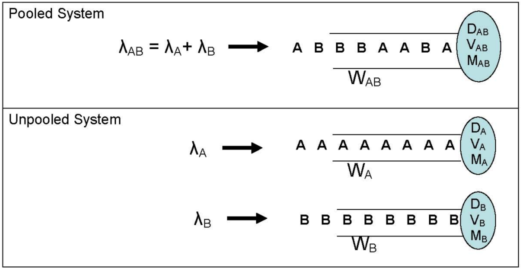
Principle of Pooling and Principle of Focus

- **The pooling principle:** advocates for departments to pool customer demands, along with the resources used to fill those demands, in order to yield economies of scale and operational improvements. This is the model typically used in hospitals.
- **The principle of focus:** advocates for departments to limit the range of services they provide in order to reduce complexity and decrease service times. To exploit this in health care, hospitals aggregate patients with similar diagnoses together into a single department.
- **Both principles** are offered as a potential method to improve a systems performance without adding additional resources.

Case Study

- Vanberkel et al. (2009) Reallocating Resources to Focused Factories: A Case Study in Chemotherapy
- Case study in the Dutch Cancer Institute – Antoni van Leeuwenhoek Hospital (NKI-AVL)
  - the managers are planning to organize a dedicated breast cancer department in chemotherapy;
  - an efficiency loss is expected due to unpooling;
  - our goal was to quantify this loss given the data.
- The questions arise:
  - how can we quantify the loss?
  - which parameters are determining the loss?
  - how much efficiency gain due to *focus* is needed to cover for the loss?

Slotted Queueing Model



- To compare the two systems we formulate a discrete time slotted queueing model used to evaluate "Access Time" (W) in each of the three queues

University of Twente
Enschede - The Netherlands

Slotted Queueing Model

- Queueing model Parameters
  - Arrivals / Day = referrals for appointments (assumed Poisson (λ) distributed)
  - Served / Day (S) = Function of Appointment Length (D), Coefficient of Variance (C), Number of Rooms (M) and Daily Working Hours (t)
- Evaluation
  - Simulated Case Study
  - Analytic Approach
  - Numeric Experiments

Modelling Served / Day (S)

- Consider a single room open for t minutes per day
  - From renewal theory, patients completed per day ($N(t)$) is:

$E[N(t)] \approx \frac{t}{D} + \frac{1}{2}(C^2 - 1)$ assume $D << t$ and i.i.d.

- Consider rooms $i = 1, \ldots M$ open for t minutes per day then:

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

- Associated Variances (also from renewal theory):

$V_{N(t)} \approx \frac{V^2 t}{D^3} = \frac{C^2 t}{D}$ $\qquad V_S \approx MV_{N(t)} = \frac{MC^2 t}{D}$

Evaluating Waiting Time(W)

- Queue Length Dynamics are described by Lindley's Recursion:

$$L_{n+1} = \text{Max}(L_n + X_n - S_n, 0) \quad n > 1 \qquad n \to \infty, L_n \to L$$

- And can be approximated with Allen-Cunneen approximation:

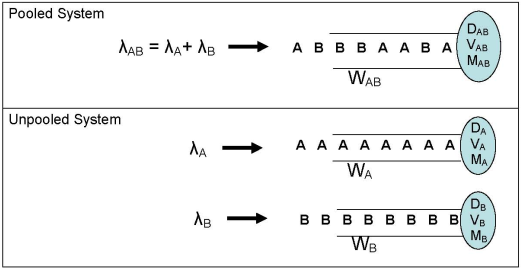
$$L \approx \lambda \frac{\rho}{1-\rho} \frac{C_S^2 + (1/\lambda)^2}{2} = \lambda \frac{\rho}{2(1-\rho)} \left(\frac{1}{\lambda} + \frac{MC^2 t}{D} \frac{1}{M^2 \left(\frac{t}{D} + \frac{1}{2}(C^2 - 1) \right)^2} \right)$$

$$\approx \frac{\rho}{2(1-\rho)} \left(1 + \frac{C^2}{\rho} \right) \qquad \text{where } \rho \approx \frac{\lambda D}{Mt}$$

- Finally with Little's Law, W can be approximated by:

$$\text{W} \approx \frac{\rho}{2(1-\rho)\lambda} \left(1 + \frac{C^2}{\rho} \right)$$

Comparing Performance



- Assume the unpooled clinic can treat patients faster, due to the principle of focus. How much shorter does the service time need to be in order to achieve the same performance as the pooled clinic?

Comparing Performance

- Let D' be the new faster service time of the unpooled clinic
- Let ρ' be the load of the unpooled clinic with D'

  $W'_A = W_{AB}$

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

  $\rho \approx \frac{\lambda D}{Mt}$ $\qquad C = \sqrt{\frac{V}{D^2}}$

- Relating the parameters from the pooled and unpooled clinics

  $\lambda_{AB} = \lambda_A + \lambda_B$
  $D_{AB} = qD_A + (1-q)D_B$
  $V_{AB} = q(V_A + D_A^2) + (1-q)(V_B + D_B^2) - D_{AB}^2$
  where $q = \lambda_A/\lambda_{AB}$

  $M_{AB} = M_A + M_B$

University of Twente
Enschede - The Netherlands

Comparing Performance

- Assume clinic resources are divided proportionally (Relaxed later)

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

- Finally, ignoring second order and higher terms of $(1-\rho)$:

$$D'_A \approx D_A \left(1 - \frac{1+C_A^2}{1+C_{AB}^2}\frac{\lambda_{AB}}{\lambda_A}\right)(1-\rho) + D_A$$

- Let Z_A be the Required Service Time Change to make $W'_A = W_{AB}$, then

$$Z_A = \frac{D'_A}{D_A} - 1$$

$Z_A = -5\%$ means D_A must decrease by 5%

$$Z_A = \frac{D'_A}{D_A} - 1 \approx \left(1 - \frac{1+C_A^2}{1+C_{AB}^2}\frac{\lambda_{AB}}{\lambda_A}\right)(1-\rho)$$

University of Twente
Enschede - The Netherlands

Z_A for Typical Parameter Ranges

#	Clinic Description	ρ_0	$\frac{\lambda_A}{\lambda_{AB}}$	$\frac{1+C_A^2}{1+C_{AB}^2}$	Z_A
1	Busy Clinic, $\lambda_A >> \lambda_B$, $V_A << V_B$	0.99	0.7	0.32	0
2	Busy Clinic, $\lambda_A >> \lambda_B$, $V_A = V_B$	0.99	0.7	1	-0.01
3	Busy Clinic, $\lambda_A >> \lambda_B$, $V_A >> V_B$	0.99	0.7	1.36	-0.01
4	Busy Clinic, $\lambda_A << \lambda_B$, $V_A << V_B$	0.99	0.3	0.17	0
5	Busy Clinic, $\lambda_A << \lambda_B$, $V_A = V_B$	0.99	0.3	1	-0.03
6	Busy Clinic, $\lambda_A << \lambda_B$, $V_A >> V_B$	0.99	0.3	2.58	-0.08
7	Quite Clinic, $\lambda_A >> \lambda_B$, $V_A << V_B$	0.7	0.7	0.32	0.16
8	Quite Clinic, $\lambda_A >> \lambda_B$, $V_A = V_B$	0.7	0.7	1	-0.13
9	Quite Clinic, $\lambda_A >> \lambda_B$, $V_A >> V_B$	0.7	0.7	1.36	-0.29
10	Quite Clinic, $\lambda_A << \lambda_B$, $V_A << V_B$	0.7	0.3	0.17	0.13
11	Quite Clinic, $\lambda_A << \lambda_B$, $V_A = V_B$	0.7	0.3	1	-0.7
12	Quite Clinic, $\lambda_A << \lambda_B$, $V_A >> V_B$	0.7	0.3	2.58	-2.28

- $\rho_0 \in [0.7, 0.99]$
- $\lambda_A/\lambda_{AB} \in [0.3, 0.7]$
- $C_A^2, C_B^2 \in [0.5, 3]$

University of Twente
Enschede - The Netherlands

Factors

- When ρ is large it dominates Z_A
- When $\lambda_A >> \lambda_B$ then $(1 + C_A^2)/(1 + C_{AB}^2)$ goes to 1 and has no affect on Z_A
  - However $(1 + C_B^2)/(1 + C_{AB}^2)$ becomes increasingly important for Group B
- The absences of M_{AB} and D_{AB} implies that their influence on Z_A is minimal

Accuracy of Approximation for Z_A

- Percent by which the Approximation overestimates Z_A when compared to simulated results:

$\frac{\lambda_A}{\lambda_{AB}}$	$\rho = 0.79$	$\rho = 0.88$	$\rho = 0.97$
0.3	40.6%	18.1%	4.1%
0.4	22.1%	9.8%	1.5%
0.5	13.1%	6.3%	1.0%
0.6	10.4%	3.1%	0.0%
0.7	5.2%	1.1%	0.0%

Numeric Experiments

- Objectives
  - Further test the Approximation for Z_A
  - Gain further insight on factors influencing Z_A
  - Provide solutions for a range of clinic environments
  - Relax assumption of dividing rooms proportionately
  - Relax assumption $D << t$
- Data
  - Arrival Rate Distribution: Poisson (λ)
  - Service Rate Distribution: Phase Type Distribution, two moment fit of appointment length data
- Simulation Procedure
  - Z_A is computed by incrementally decreasing D_A until $W_A = W_{AB}$
  - Skip room divisions where $|Z_A| \geq 25\%$ and/or $|Z_B| \geq 25\%$

University of Twente
Enschede - The Netherlands

Numeric Experiments

- Experiment Range

Clinic Environments	M_{AB}	D_{AB}	λ_{AB}	ρ_0	C_A, C_B
Base Clinic	20	30	282	0.88	0.5, 0.5
Busier Clinic	20	30	**310**	**0.97**	0.5, 0.5
Smaller Clinic	**10**	30	**141**	0.88	0.5, 0.5
Shorter Appointment Lengths	20	**15**	**564**	0.88	0.5, 0.5
Higher Appointment Length Variability	20	30	282	0.88	**2.0, 2.0**
Different Coefficient of Variance	20	30	282	0.88	**2.0,** 0.5

- In total paper includes results for 80 different clinic environments

Numeric Experiments

$\frac{\lambda_A}{\lambda_{AB}}$	$D_A/D_{AB} = 0.5$	$D_A/D_{AB} = 1.0$	$D_A/D_{AB} = 1.5$	$D_A/D_{AB} = 2.0$
0.3	**-10% (3), -4% (17)**	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)		

$\rho = 0.88,\ M_{AB} = 20,\ D_{AB} = 30,\ \lambda_{AB} = 282,\ C_A = C_B = 0.5$

University of Twente
Enschede - The Netherlands

Factor Summary

Considerations	Affect on Z	General Management Guidelines When Unpooling
Clinic Load (ρ)	As ρ increases, Z_A decreases	Clinics under high load require less decrease in service time to compensate.
Room Division	Disproportionate splits increase $\lvert Z_A\rvert + \lvert Z_B\rvert$	The smallest total decrease in service time occurs when the difference between ρ_{AB}, ρ_A and ρ_B is minimized.
Clinic Size (M_{AB})	As M_{AB} increases, Z_A decreases (slightly)	The number of rooms in your clinic does not greatly influence the needed service time change. However, in smaller clinics it is more difficult to proportionally divide the rooms.
Appointment Lengths (D_{AB})	Z_A is mostly insensitive to D_{AB}	Z_A appears to be mostly insensitive to the length of the appointment.
Variability of Appointments Lengths (C_A, C_B)	As C_A, C_B increase, Z_A and Z_B increase	Unpooling patient groups with highly variable appointment lengths requires more decrease in service time to compensate.
Different Coefficient of Variance ($C_A < C_B$)	Z_A decreases when $C_A < C_B$	The patient group with the smaller C requires less decrease in service time to compensate.
Size of each group (λ_A/λ_{AB})	As λ_A/λ_{AB} decreases, Z_A increases	The smaller patient group requires more decrease in service time to compensate.
Appointment Length Ratio (D_A/D_{AB})	Mostly insensitive to D_A/D_{AB}	The required change in service time appears to be mostly insensitive to the ratio of appointment lengths.

University of Twente
Enschede - The Netherlands

ORchestra bibliography

University of Twente
Enschede - The Netherlands

ORchestra Bibliography

www.choir.utwente.nl/en/orchestra/

Visit the online, categorised resource for references in OR/MS in healthcare

Questions?

p.t.vanberkel@utwente.nl
www.choir.utwente.nl