Scheduling admissions and reducing variability in bed demand

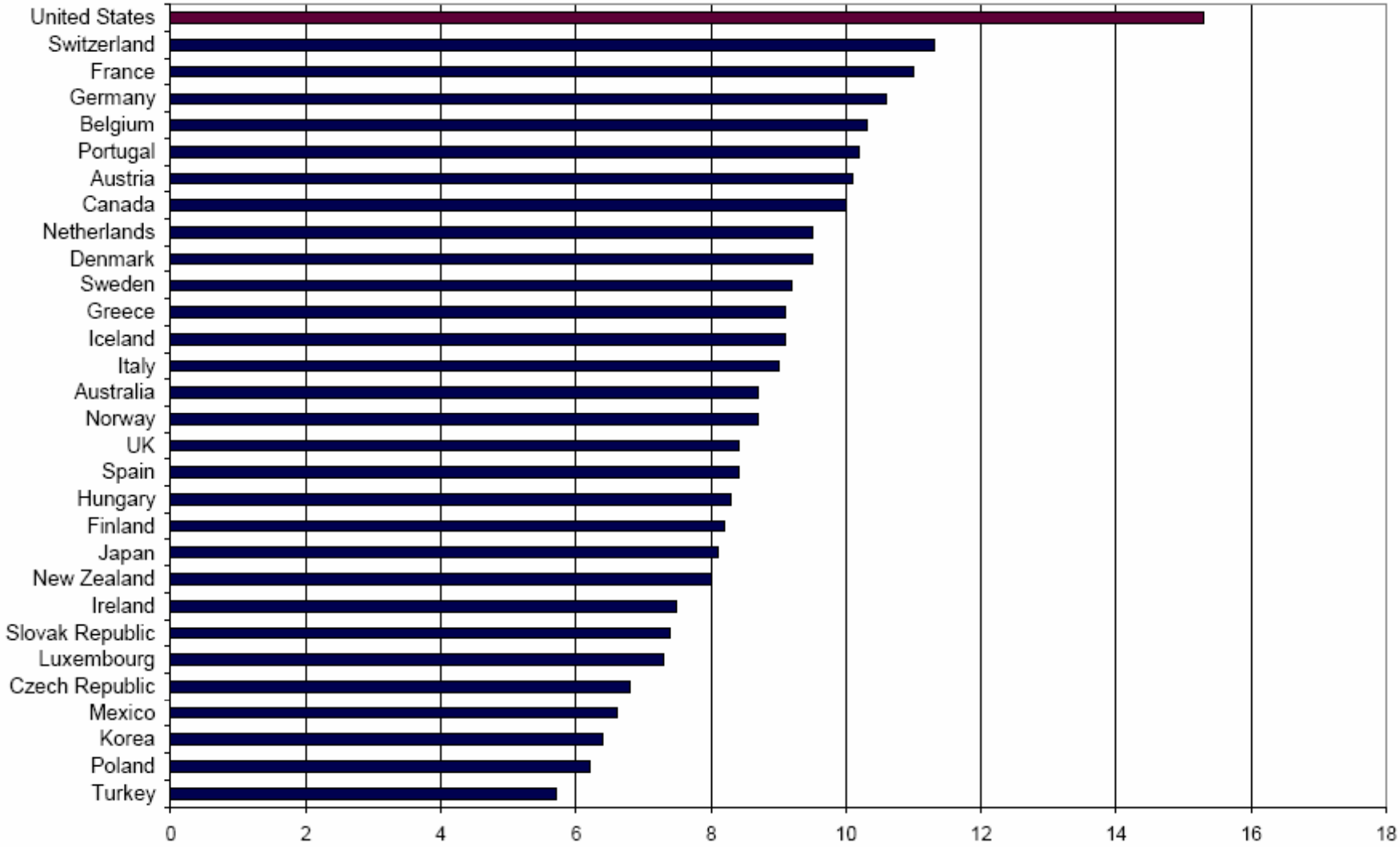
René Bekker
VU University Amsterdam

Paulien Koeleman
VU University Amsterdam
CC Zorgadviseurs

YEQT-IV, Eindhoven 2010
Optimal control in stochastic systems

VU University Amsterdam

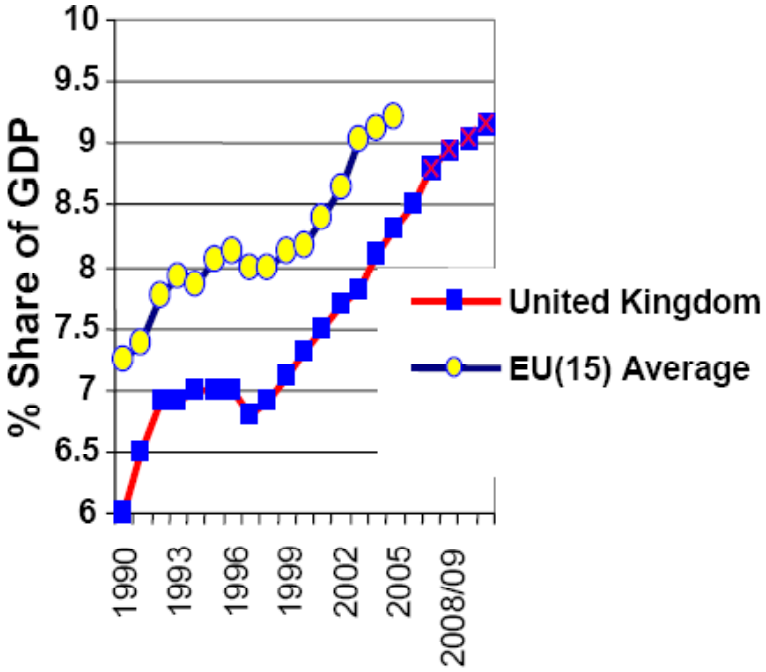
Healthcare Spending as % GDP



Source: Organization for Economic Cooperation and Development, OECD Health Data, 2008 (Paris: OECD, 2008).
Note: For countries not reporting 2006 data, data from previous years is substituted.

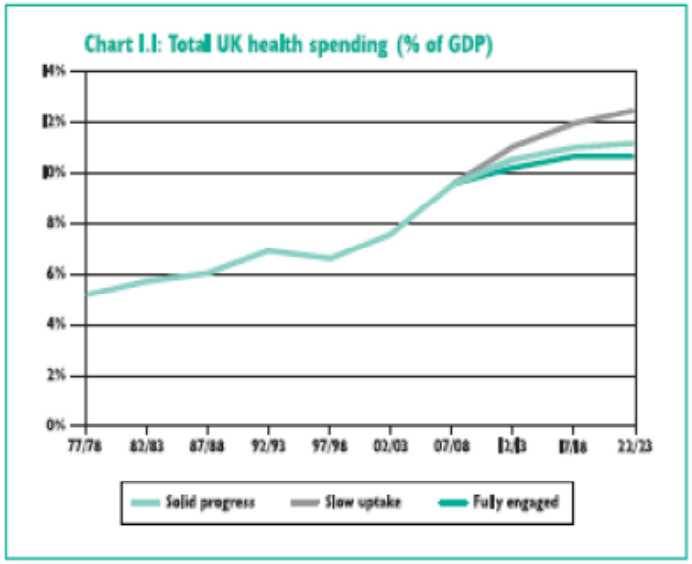
Healthcare spending UK

The proportion of UK GDP spent on healthcare has steadily risen



And is this rise is expected to continue

The Wanless report estimated that In fifteen years time between 10.5% and 13% of GDP would be required to fund the UK health system



From: Dr. Royston, Presentation OR50 Conference

Variability and dimensioning clinical wards

Key topic: variability in bed occupancy

Issues for clinical wards:

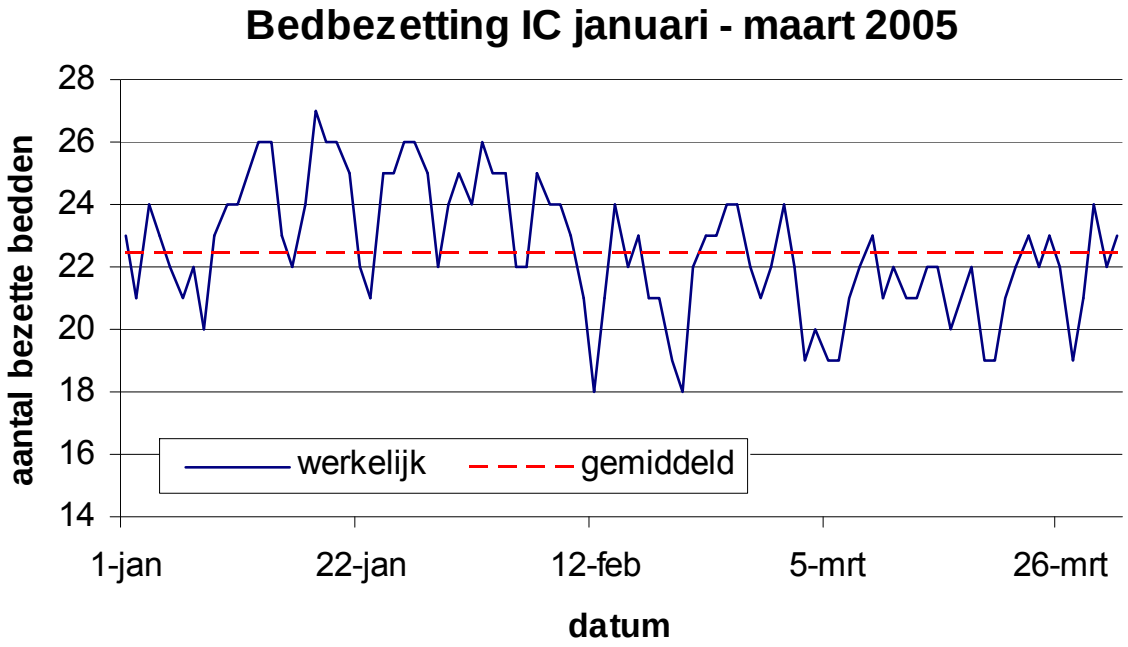
- High occupancy rate (e.g. ICU)
- Blocking of patient admissions:
  - Cancel scheduled surgeries
  - Reallocate urgent patients
- Risks for patient safity
  (Litvak et al 2005, Metcalfe et al 1997)
- High pressure medical staff
- Throughput for different specialists

Stochastic models are required!

Variability and the flaw of averages

Sources of variation:

- Admissions
  - Burstiness in admissions (variation in number per day)
  - Weekly pattern
- Length of stay



Variability and the flaw of averages

A Sobering Example of the Flaw of Averages

Consider the state of a drunk, wandering around on a busy highway. His average position is the centerline, so……..

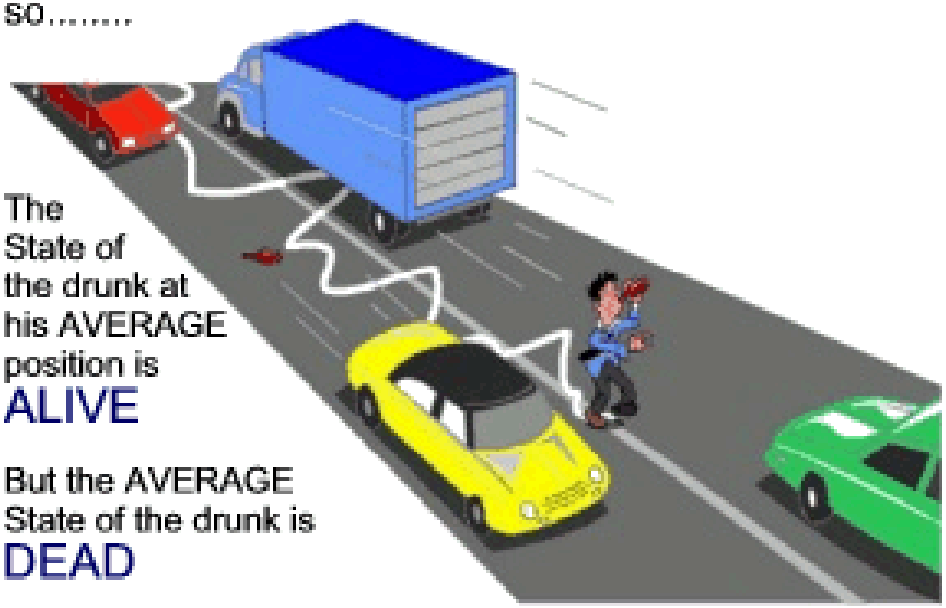


Illustration from Decision Making with Insight 2nd Edition, Sam L. Savage, ISBN 0534386393, © 2003, Reproduced with permission.

Two main research questions

Sources of variation:

- Admissions
  - Burstiness in admissions (variation in number per day)

    Q1: Impact of reduced burstiness?

  - Weekly pattern

    Q2: Schedule admissions to handle weekend effect?

- Length of stay

Queueing theory

- Queueing theory for dimensioning of clinical wards
  - Often used Erlang loss (and delay) models
- Assumptions:
  - *Stationary* and *Poisson* arrivals
  - Often: exponential LOS
- In practice predictable fluctuations in number of arrivals

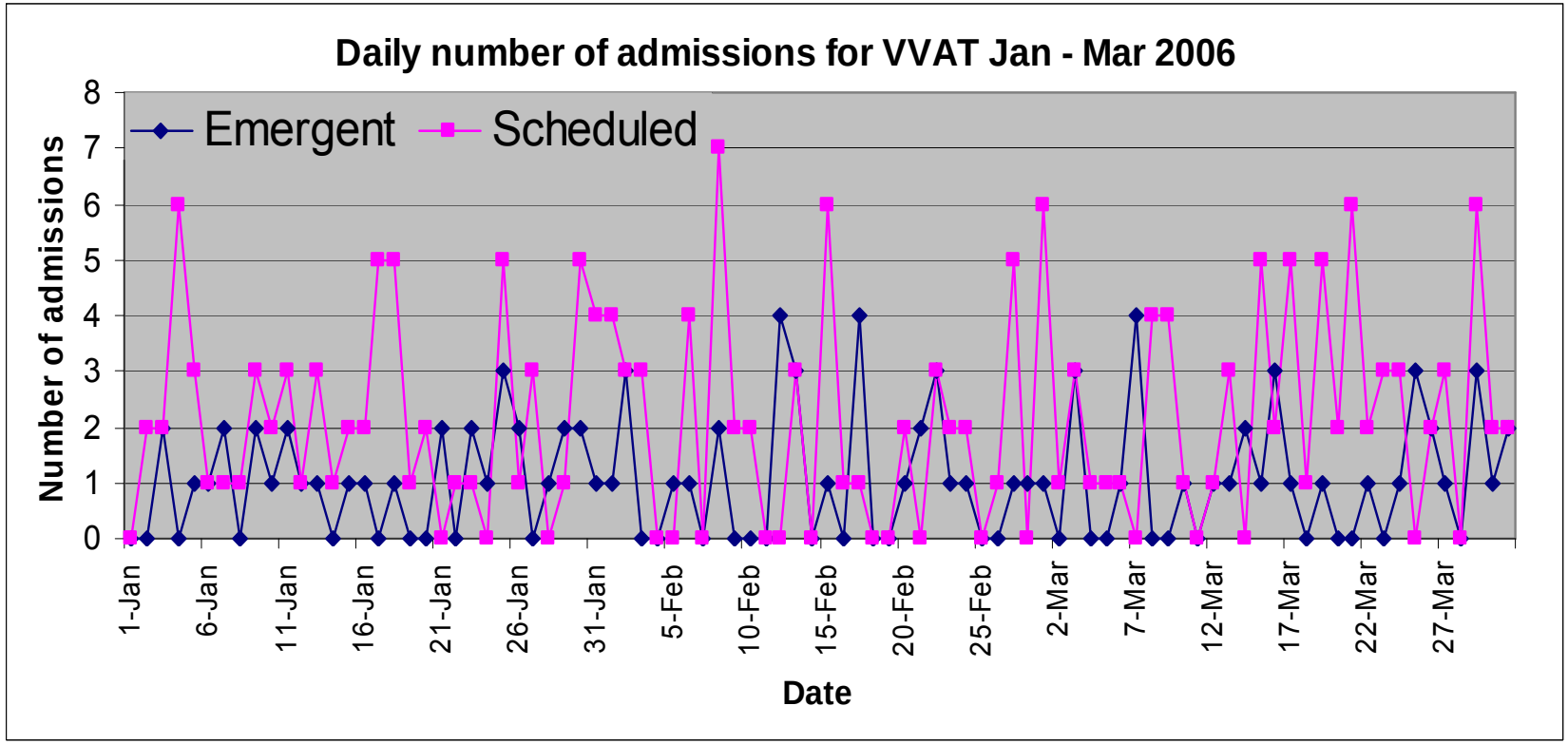
Elective admissions & Poisson arrivals?

Organization

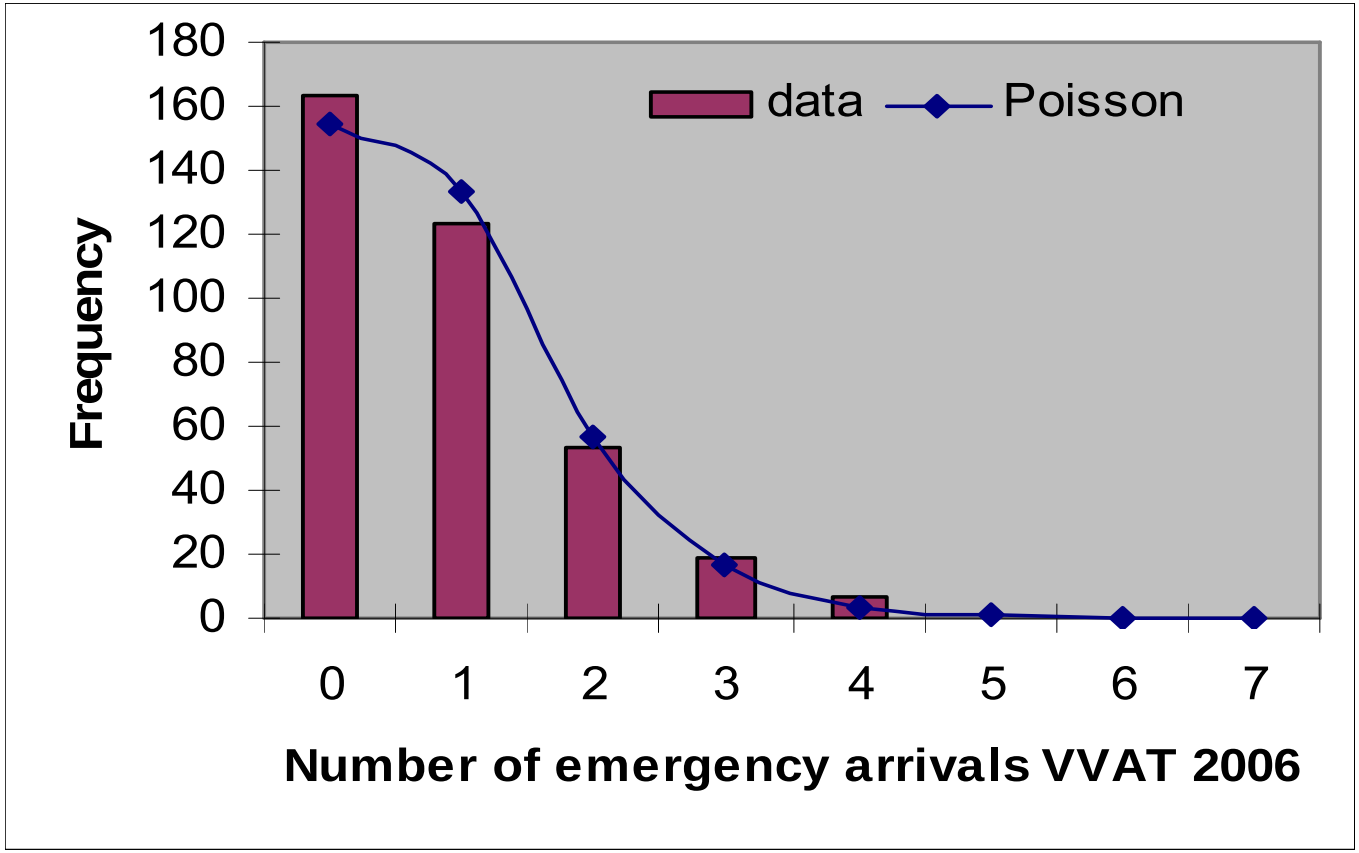
- Introduction
- Data analysis
- Q1: impact of burstiness
  - Heavy traffic limit
- Q2: scheduling elective patients
  - Performance analysis
  - Optimization
- Final thoughts…

Data analysis

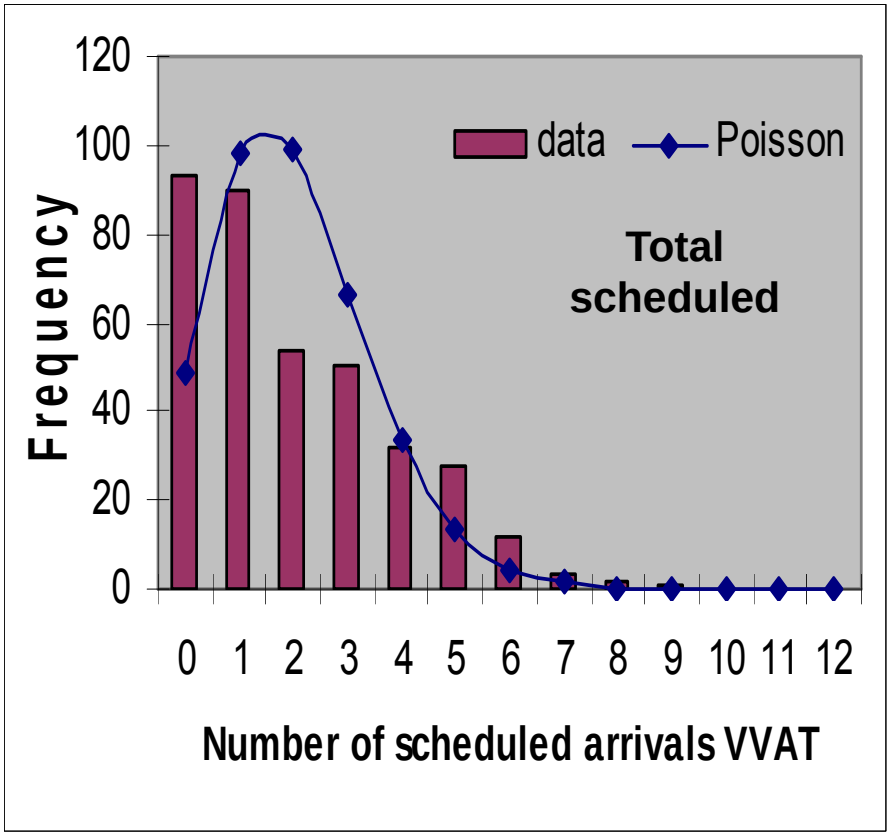
Admissions

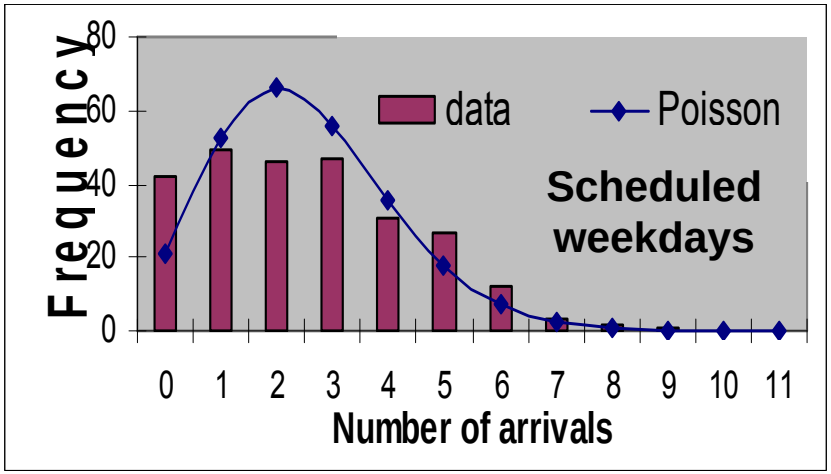

Daily number of admissions for VVAT Jan - Mar 2006
Emergent
Scheduled
Number of admissions
0
1
2
3
4
5
6
7
8
1-Jan
6-Jan
11-Jan
16-Jan
21-Jan
26-Jan
31-Jan
5-Feb
10-Feb
15-Feb
20-Feb
25-Feb
2-Mar
7-Mar
12-Mar
17-Mar
22-Mar
27-Mar
Date

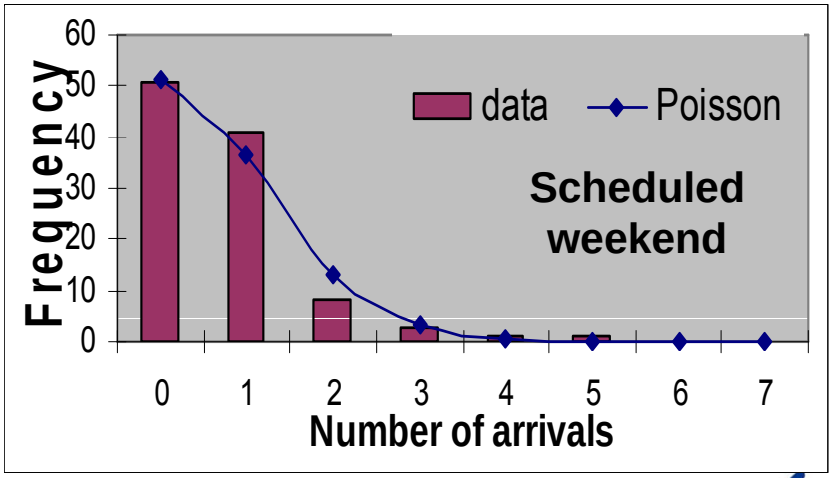
Admission emergency patients



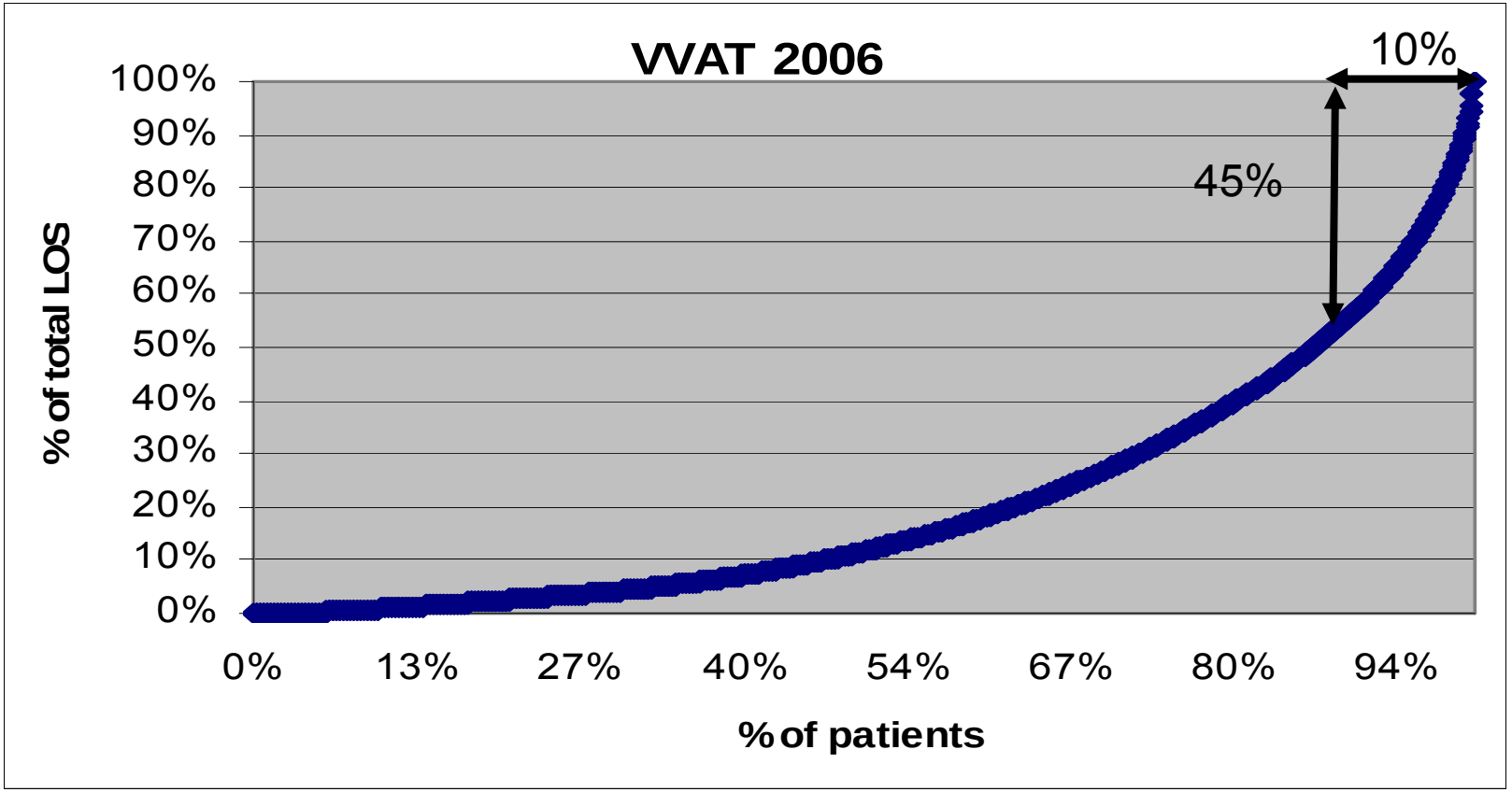
Admission scheduled patients







LOS and Lorenz curves



LOS and Gini coefficient (*G*)

- Traditionally: measure of inequality in wealth
  → measure of dispersion in bed capacity
- Calculation: area on the Lorenz curve diagram
- Piecewise differentiable service time distribution:

$$G = 1 - \frac{1}{\mathbb{E}[S]} \int_0^\infty \mathbb{P}(S > y)^2 \mathrm{d}y$$

- Example:
  - Deterministic: *G = 0*
  - Exponential: *G = 1/2*
  - Hyperexp (balanced means): $G = \frac{3}{4} - p_1 p_2$

Q1: Extending the Erlang loss model …

Heavy traffic approximation

- What are the efficiency gains for smoother elective admissions?
- Stationary arrival process with squared coefficient of variation c_a^2
- Approximations based on infinite server queue $G/G/\infty$
- X_ρ = eq. # busy servers in $G/G/\infty$
- Heavy traffic limit, Borovkov (1967),

where $$\frac{X_\rho - \rho}{\sqrt{\rho z}} \to N(0,1) \qquad \text{as } \rho \to \infty$$

$$z = 1 + (c_a^2 - 1)\frac{1}{\mathbb{E}S}\int_0^\infty \mathbb{P}(S > y)^2 \mathrm{d}y$$

From Whitt (1984)

Heavy traffic approximation

- What are the efficiency gains for smoother elective admissions?
- Stationary arrival process with squared coefficient of variation c_a^2
- Approximations based on infinite server queue $G/G/\infty$
- X_ρ = eq. # busy servers in $G/G/\infty$
- Heavy traffic limit, Borovkov (1967),

where $$\frac{X_\rho - \rho}{\sqrt{\rho z}} \to N(0,1) \qquad \text{as } \rho \to \infty$$

$$z = 1 + (c_a^2 - 1)(1 - G)$$

From Whitt (1984)

Heavy traffic peakedness

$$z = 1 + (c_a^2 - 1)(1 - G)$$

Remarks:

- Variance in offered load is ρz
- Poisson arrivals $\rightarrow$ $z = 1$ and variance offered load is ρ
- Relation to Gini coefficient …
- Variability: $c_a^2, 1 - G, z$
- z increasing in $c_a^2 \rightarrow$ arrival process as smooth as possible
- Impact z on service time depends on sign $c_a^2 - 1$

  $\rightarrow$ reducing variability LOS only beneficial for smooth arrival process

From Whitt (1984)

Blocking probability

- Hayward approximation (Whitt '84):

$$B_C \equiv B_C(s, \rho, z) \approx B\left(\frac{s}{z}, \frac{\rho}{z}\right)$$

- Extension of Erlang loss to non-integral *s*

- Example:
  - Arrivals: 6 patients p/d
  - ALOS = 4 days
  - Offered load ρ = 24
  - Exponential LOS

Arrivals	Nr. of beds	Blocking probability (%)
Poisson	28	6,7 %
Det	28	3,2 %
	26	6,3 %
Det + Poisson	28	5,0 %
	27	6,6 %

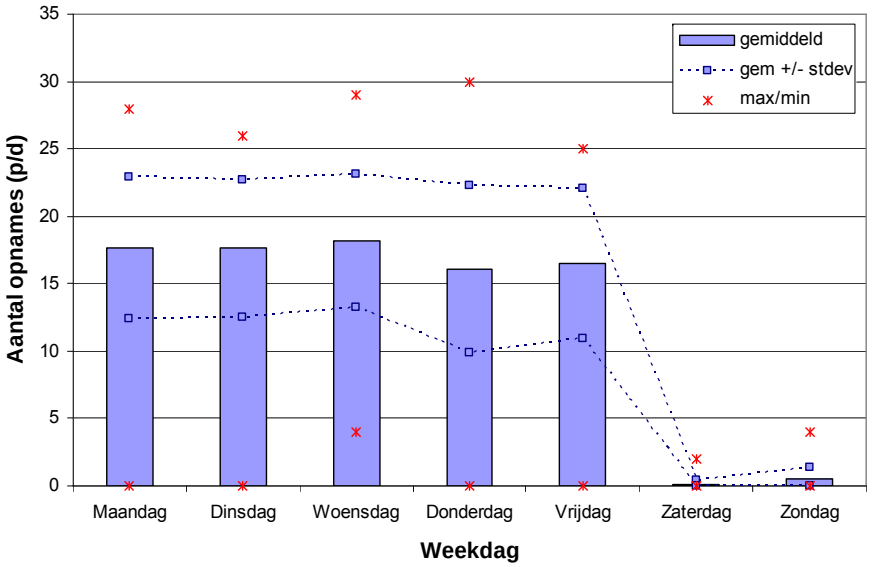
Q2: Scheduling elective patients

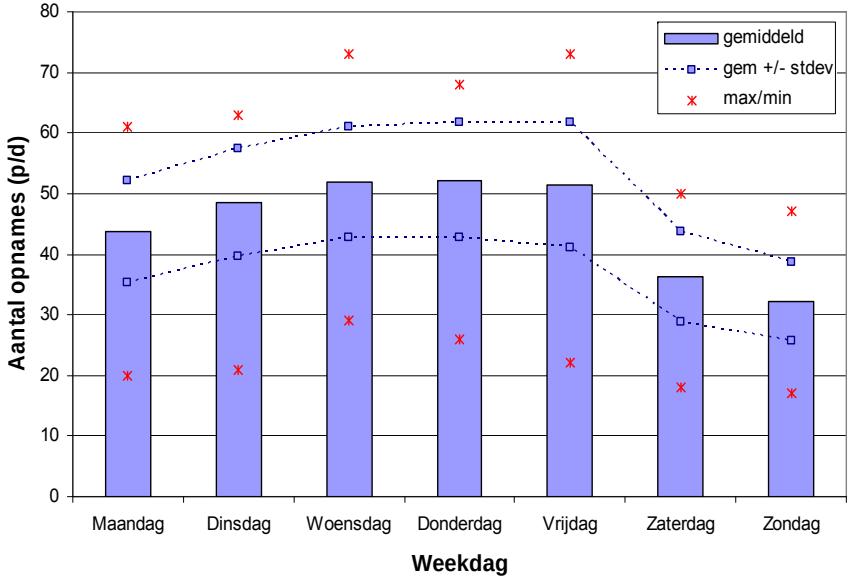
Scheduling elective patients…

Reduce variability in admissions (Q1)

Is the same number of admissions for every weekday optimal…?

Admissions → **Demand for beds**





Infinite-server queues

- Use infinite-server queues to analyze offered load
- Stationary Poisson arrivals:

$$\mathbb{P}(Q^\infty = k) = \frac{\rho^k}{k!} e^{-\rho} = \mathrm{Ps}(\rho)$$

- Predictable weekly arrival pattern
- Time-dependent Poisson arrivals:

$$\mathbb{P}(Q^\infty(t) = k) = \mathrm{Ps}(m(t))$$

with

$$\begin{aligned} m(t) &= \int_{v=0}^{\infty} \lambda(t-v)\mathbb{P}(S > v)\mathrm{d}v \\ &= \mathbb{E}\left[\lambda(t - S_e)\right]\mathbb{E}[S] \end{aligned}$$

From Eick, Massey, Whitt (1993)

Infinite-server queues

Approximation blocking probability using MOL:

$$B_t \quad = \quad \mathbb{P}(Q^s(t) = s) \;\approx\; \frac{\mathbb{P}(Q^\infty(t) = s)}{\mathbb{P}(Q^\infty(t) \leq s)}$$

Special cases:

- Exponential service times (phase type)
- Sinusoidal arrival rates
  - Eick, Massey, Whitt (1993); Davis, Massey, Whitt (1995); Green, Kolesar, Soares (2001)
- Periodic arrival pattern; *λ(t)* step function
  - B, de Bruin (2010)
  - Flexible setup

Example: two λ's per period

Example: λ_1 during $[0, a_1]$ and λ_2 during $[a_1, a_2]$ (periodic)

- Offered load 1st interval: take $0 < t < a_1$

$$
\begin{aligned}
m(t) &= \int_{v=0}^{\infty} \lambda(t-v)\mathbb{P}(S>v)\mathrm{d}v \\
&= \int_{v=0}^{t} \lambda_1 e^{-\mu v}\mathrm{d}v + \int_{v=t}^{\infty} \lambda(t-v)e^{-\mu v}\mathrm{d}v \\
&= \frac{\lambda_1}{\mu}\left(1-e^{-\mu t}\right) + e^{-\mu t}\int_{u=0}^{\infty} \lambda(a_1-u)e^{-\mu u}\mathrm{d}u \\
&= \frac{\lambda_1}{\mu}\left(1-e^{-\mu t}\right) + e^{-\mu t}m(0)
\end{aligned}
$$

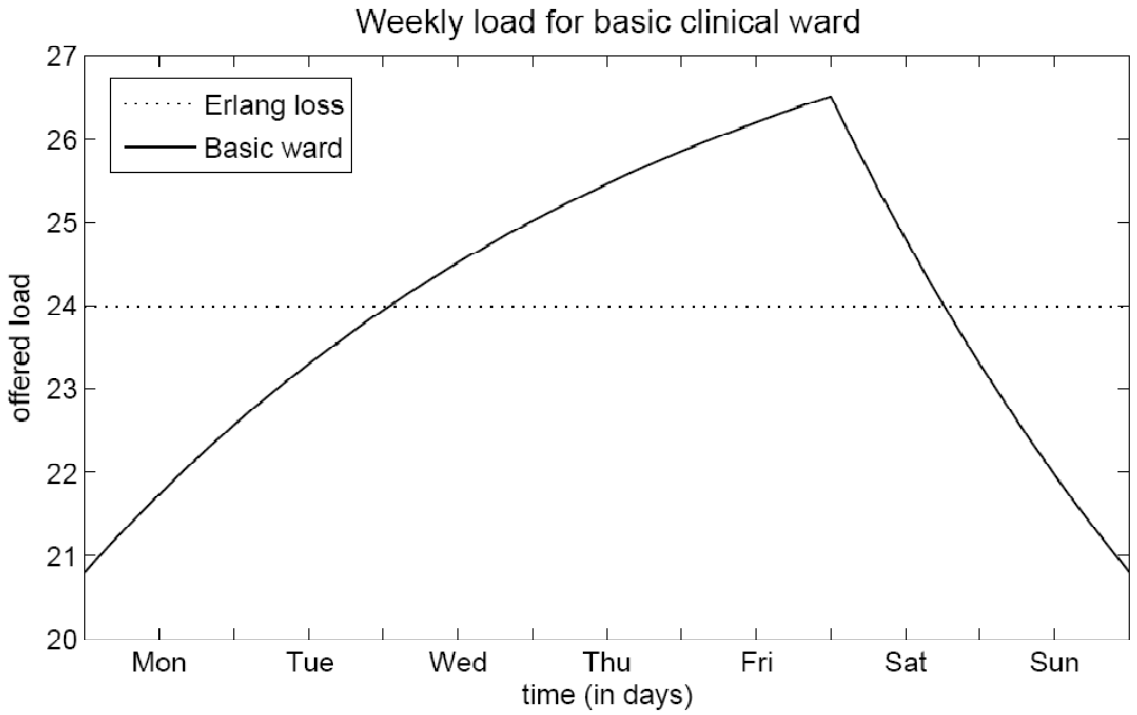
Linear in λ_1 ...

- Use $m(a_2) = m(0)$

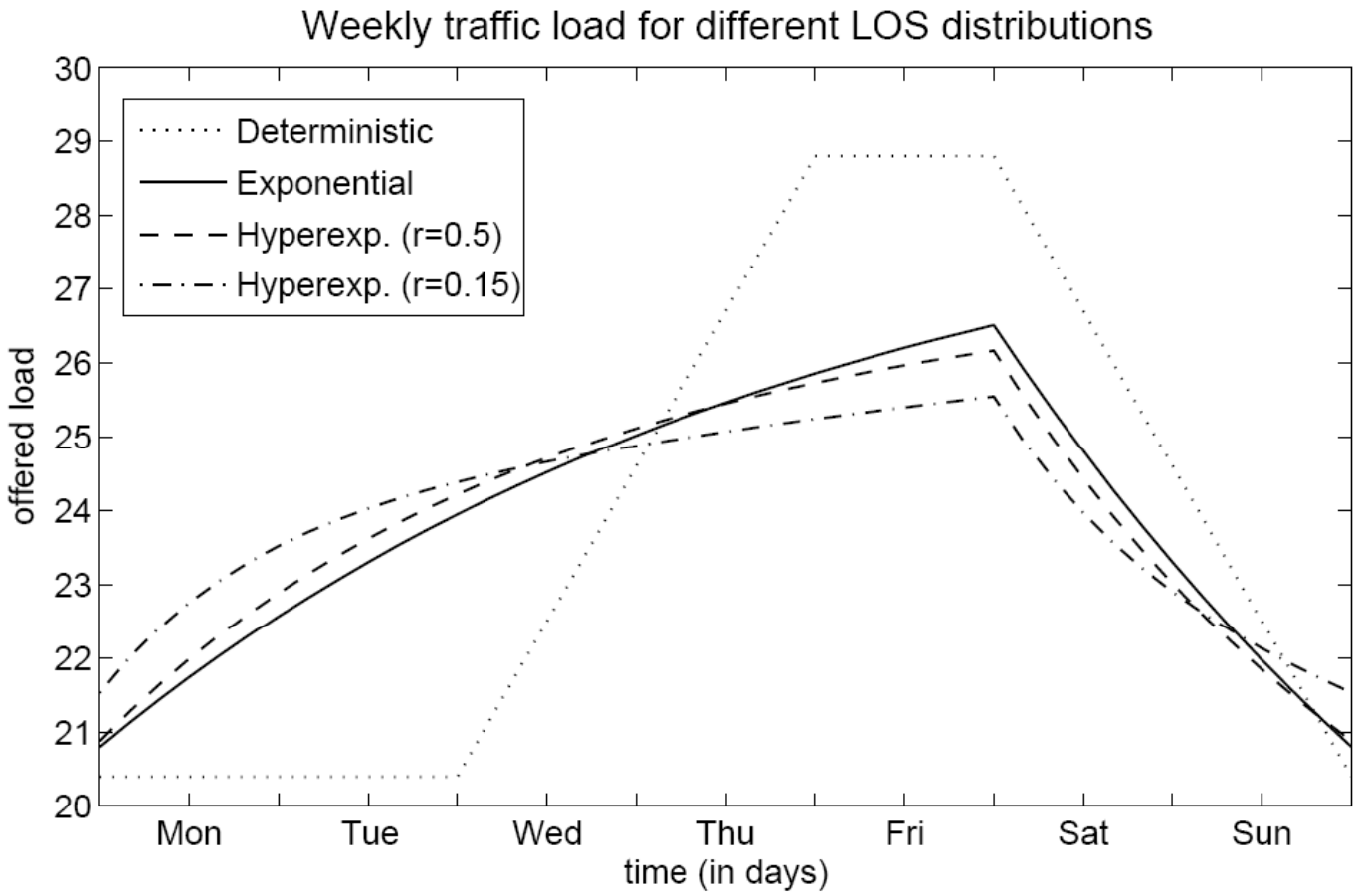
Weekend effect

Example (revisited)

- ALOS = 4 days
- Average arrivals (p/d)
  - Average: 6
  - *Week days*: 7,2
  - *Weekends:* 3
- Average offered load: 24 patients



Impact LOS distribution



Q2: Scheduling elective admissions

Scheduling elective admissions

- Determine target load at day *d*: *m*(d), d* = 1,…,7
- For given admission schedule *λ(d),* time-dependent analysis gives offered load *m(d)*
- Optimization:

$$\begin{aligned} \min \quad & \sum_d (m(d) - m^*(d))^2 \\ \text{subject to} \quad & \text{determining } m(d) \text{ given } \lambda(d) \\ & \sum_d \lambda(d) = \text{Weekly production target} \\ & \lambda(d) \geq 0, \quad \text{for all } d \end{aligned}$$

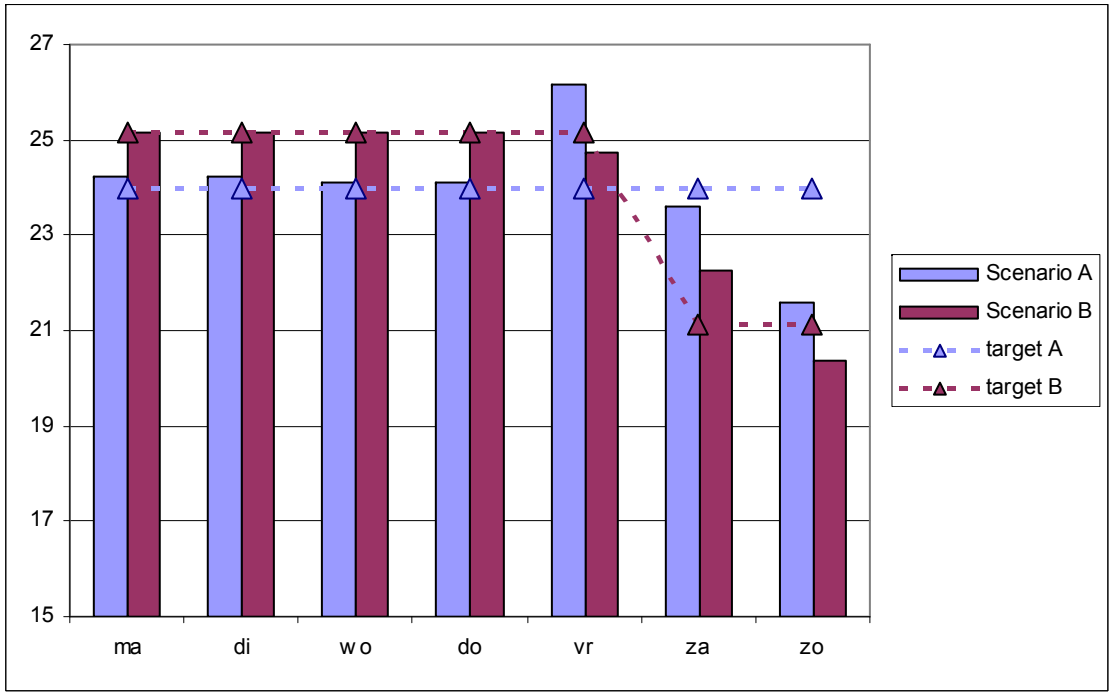
- Remark: QP with linear constraints

Example (revisited)

- ALOS = 4 days (exponential LOS)
- Arrivals
  - Emergency: 3 p/d (on average)
  - Elective: 21 per week
    - Short: ALOS = 2 days; 10,5 per week
    - Long: ALOS = 6 days; 10,5 per week
  - No elective admissions during the weekend
- Average offered load: 24 beds
- Target load:
  - Scenario A: 24 beds for every day
  - Scenario B: close 4 beds in the weekends

Example scheduling elective patients

	Admissions	Mon	Tue	Wed	Thu	Fri	Weekend
Scenario A	Short	5,4	3,8	1,3	0	0	0
	Long	0	0	2,3	3,2	5,0	0
Scenario B	Short	5,4	3,5	0,2	0,8	0,6	0
	Long	2,2	0,5	3,4	2,5	1,9	0



What about time-dependent variability? Final thoughts…

Discrete time infinite-server queues

Infinite server queue in discrete time:

- *λ(d)* deterministic number of arrivals on day *d*
- *S* generic LOS (service time)
- *Q(d)* number of patients on day

Performance characteristics:

$$m(d) := \mathbb{E}Q(d) = \sum_{k=0}^{\infty} \lambda(d-k)\mathbb{P}(S > k)$$

$$\mathbb{V}arQ(d) = \sum_{k=0}^{\infty} \lambda(d-k)\mathbb{P}(S > k)\left[1 - \mathbb{P}(S > k)\right]$$

Rm: both linear in *λ(k)*

Discrete time infinite-server queues

- Adopt heavy-traffic limit: *Q(d)* has normal distribution
- Probability of shortage of beds:

$$\begin{aligned}\mathbb{P}(Q(d) > s) &= \mathbb{P}\left(\frac{Q(d) - m(d)}{\sqrt{\mathbb{V}arQ(d)}} > \frac{s - m(d)}{\sqrt{\mathbb{V}arQ(d)}}\right) \\ &= 1 - \Phi\left(\frac{s - m(d)}{\sqrt{\mathbb{V}arQ(d)}}\right) =: \eta\end{aligned}$$

for some target η

- Time-stable staffing then requires

$$s(d) = m(d) + \beta\sqrt{\mathbb{V}arQ(d)},$$

With QoS parameter

$$\beta = \Phi^{-1}(1 - \eta)$$

Optimization…

- Possible formulation:

$$\begin{aligned} \max \min \quad & s(d) - m(d) - \beta\sqrt{\mathbb{V}arQ(d)} \\ \text{subject to} \quad & \text{determining } m(d), \mathbb{V}arQ(d) \text{ given } \lambda(d) \\ & \sum_d \lambda(d) = \text{Weekly production target} \\ & \lambda(d) \geq 0, \quad \text{for all } d \end{aligned}$$

- Can be formulated as second-order cone program (SOCP)
- Sensitivity of value β

Conclusion

- Efficiency gains required in health care
- Modification of Erlang loss model (heavy traffic limit)
- Approximation for time-dependent analysis
- Optimal admissions planning using QP…

Challenge: Implementation of results

Thanks for your attention!

Questions, comments, suggestions?

rbekker@few.vu.nl