Inpatient Flow Management in a Singaporean Hospital

Jim Dai, Pengyi Shi
Georgia Institute of Technology, USA

Ding Ding
University of International Business & Economics, China

James Ang, Mabel Chou
NUS Business School, Singapore

Xin Jin, Joe Sim
National University Hospital, Singapore
Overview

- Empirical study
  - Inpatient flow management
  - Performance comparison after an *early discharge* policy
    - Waiting time for admission to ward
    - Stabilize hourly waiting time performance

- A stochastic network model
  - Allocation delays
  - Overflow policy
  - Endogenous service times

- What-if analysis
  - Factors help to stabilize waiting time
Patients from 4 admission sources competing for inpatient beds

- **ED-GW patients**
  - 66.9 (65%)
- **Elective patients**
  - 18.5 (18%)
- **SDA patients**
  - 9.13 (9%)
- **ICU-GW patients**
  - 9.12 (9%)

Total inpatient beds ~600
Patient distribution
Key performance measures

- Waiting time for admission to ward (Jan 08 – Jun 09)
  - Waiting time = admission time – bed request time
  - Average: 2.82 hour
  - 6.52% of ED-GW patients wait more than 6 hours to get a bed

- “x-hour service level”: Fraction of ED-GW patients waiting more than x hours
  - Ministry of Health (MOH) monitors 10-hour service level (0.80%)
  - Hospital managers also care about the 6-hour service level
Time dependency

- Waiting time depends on patient’s bed request time
  - Jan 08 – Jun 09
  - Can we stabilize?
Literature review

  - **Staffing of Time-Varying Queues to Achieve Time-Stable Performance**

  - **The relationship between inpatient discharge timing and emergency department boarding**

- Affiliations: Department of Emergency Medicine, Northwestern University; Harvard Affiliated Emergency Medicine Residency, Brigham and Women’s Hospital–Massachusetts General Hospital, …
Bed request rate and arrivals to ED

- ED-GW patient’s bed request rate (green curve) depends on arrival rate to ED (blue curve)
Mismatch between demand and supply of beds

- Jan 08 – Jun 09
Early discharge policy

• Moving the discharge time a few hours earlier in the day
  • Safe: limited effect in increasing patient’s risk
  • Costly to implement

• Recommended by many studies, policy guidelines:
  • National Health Service (NHS, UK): “planning for a reasonable proportion of patients to leave the ward before 11 am helps to manage the total loading on beds”

• Intuition: moving the discharge time earlier (by even 1 or 2 hours) can improve operations and patient flow.

Data

- The hospital implemented early discharge policy since July 2009
- Study two periods of data
  - Jan 2008 to Jun 2009 (Period 1)
    - 13% before noon
  - Jan 2010 to Dec 2010 (Period 2)
    - 26% before noon
- Is early discharge policy helpful?
  - Empirical analyses
  - Use model to evaluate
- Key performance measures in the two periods
  - Waiting time statistics (quality)
  - Overflow rate (cost)
Empirical Analysis on the two periods

- Waiting time performance

- Overflow rate
  - Period 1: 26.9%
  - Period 2: 25.0%

- BOR
Waiting time for ED-GW patients

<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt; period</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average waiting time</td>
<td>2.82 h</td>
<td>2.77 h</td>
</tr>
<tr>
<td>6-hour service level</td>
<td>6.52%</td>
<td>5.13%</td>
</tr>
</tbody>
</table>

![Graphs showing waiting time and 6-hour service level for two periods.](image)
Challenges

- Does the modest improvement come from the early discharge?

- More importantly, is any operational policy that can stabilize the waiting time?
Unstable Environment

- Both arrival volume and capacity increases during 2008 to 2010
- Bed occupancy rate (BOR) reduces in the Period 2
  - Period 1: 90.3%
  - Period 2: 87.6%

- Need a model to help evaluate the effect of early discharge
A stochastic model

- Multi-class, multi-server pool system
  - Each server pool is either dedicated to one class of customer or flexible to serve two and more classes of customers

- Periodic arrival
  - 4 types of arrival (ED-GW, Elective, ICU-GW, SDA) for each specialty

- A novel service time model

- And other key components
Simulation replicates most performance measures

- Hourly waiting time performances

(a) Hourly average waiting time

(b) Hourly 6-hour service level
Key modeling components

- Service time model
  - Determined by admission time, **LOS** and discharge distribution
  - An endogenous modeling element
  - No longer i.i.d.

- Allocation delays
  - “Secondary” bottlenecks other than bed availability
    - Yankovic and Green (2011)
    - Armony et al (2011)

- Overflow policy
  - When to overflow a patient
  - Overflow to which server pool
Network structure
Service time model

- Service time model
  - Service time = Discharge time – Admission time
    \[= \text{LOS} + \text{Dis hour} - \text{Adm hour}\]

- LOS distribution
  - Average is \(\sim 5\) days
  - Depend on admission source and specialty

\[\text{Length of Stay (LOS)} = \text{Discharge day} - \text{Adm day}\]
AM PM patients (ED-GW patients)

- The admission time affects LOS
  - AM patients: average LOS = 4.24 days
  - PM patients: average LOS = 5.31 days

![Graph showing relative frequency of days for AM and PM patients]
Renal patients show a great reduction in Average LOS

- Average LOS is reduced by almost 1 day
Verify the service time model

- Service time model
  - Service time = LOS + Discharge hour – Adm hour

Matching empirical

(a) Empirical

(b) Simulation output
Pre- and post-allocation delays

- Patient experiences additional delays upon arrival and when a bed is allocated
  - Pre-allocation delay
    - BMU search/negotiate for beds
  - Post-allocation delay
    - Delays in ED discharge
    - Delays in the transportation
    - Delays in ward admission

- Must model bed turnover component
  - If not, hourly queue length does not match (right figure)
Time-dependent allocation delays

- The mean of allocation delay depends on when it is initiated
  - Use log-normal distribution
  - Pre-allocation delay
Overflow policy

- When a patient’s waiting time exceeds certain threshold, the patient can be overflowed to a “wrong” ward
- Beds are partially flexible
- Overflow wards have certain priority

<table>
<thead>
<tr>
<th>Cluster</th>
<th>1st Overflow</th>
<th>2nd Overflow</th>
<th>3rd Overflow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medicine</td>
<td>Other Med</td>
<td>Surgery/OG</td>
<td>Ortho</td>
</tr>
<tr>
<td>Surgery</td>
<td>Other Surg</td>
<td>Ortho/OG</td>
<td>Medicine</td>
</tr>
<tr>
<td>Ortho</td>
<td>Other Ortho</td>
<td>Surgery</td>
<td>Medicine</td>
</tr>
</tbody>
</table>
Dynamic overflow policy

**Fixed threshold**
- Threshold: 4.0 h

**Dynamic threshold**
- Threshold: 0.5 h for arrival between 7 pm and 7 am (next day); 5.0 h for others
Time-varying arrival rates
Simulation results

- Whether early discharge policy is beneficial or not
- What-if analysis
Simulation results

- Simulation shows the early discharge policy has little improvement
  (a) hourly avg. waiting time  
  (b) 6-hour service level
Aggressive early discharge policy
Aggressive early discharge + smooth allocation delay

- Waiting time performances can be stabilized
  
  (a) hourly avg. waiting time
  
  (b) 6-hour service level
Only use aggressive early discharge

- Cannot be stabilized

(a) hourly avg. waiting time
(b) 6-hour service level
Only smooth the allocation delays

- Assuming allocation delay has a constant mean
  
  (a) hourly avg. waiting time                          (b) 6-hour service level
Impact of capacity increase

- 10% reduction in utilization, plus assuming allocation delay has a constant mean

(a) hourly avg. waiting time  
(b) 6-hour service level
Simulation replicates most performance measures

- Hourly waiting time performances

(a) Hourly average waiting time

(b) Hourly 6-hour service level
Average waiting time for each specialty

- Renal patients have longest average waiting time
6-hour service level for each specialty

- Cardio and Oncology patients show significant improvement in the 6-hour service level
Overflow rate

- Overall overflow rate reduces in Period 2
Summary

- Conduct an empirical study of patient flow of the entire inpatient department

- Build and calibrate a stochastic model to evaluate the impact of discharge distribution on waiting for admission to ward

- Identify allocation delays as a second source of bottlenecks
  - Staffing appropriately in BMU, ED and Ward

- Achieve stable waiting time by aggressive early discharge + smooth allocation delay
Limitations

- Simulation cannot fully calibrate with the overflow rate
  - Bed class (A, B, C)
  - Gender mismatch
  - Hospital acquired infections
    - Example: a female Surg patient has to be overflowed to a Med ward, since the only available Surg beds are for males

- Day-of-week phenomenon
  - Admission and discharge both depends on the day of week
  - LOS depends on admission day
  - Performances (BOR, waiting time) varies among days
Questions?