INSEAD

The Business School for the World®

SMU – 23rd Feb 2018

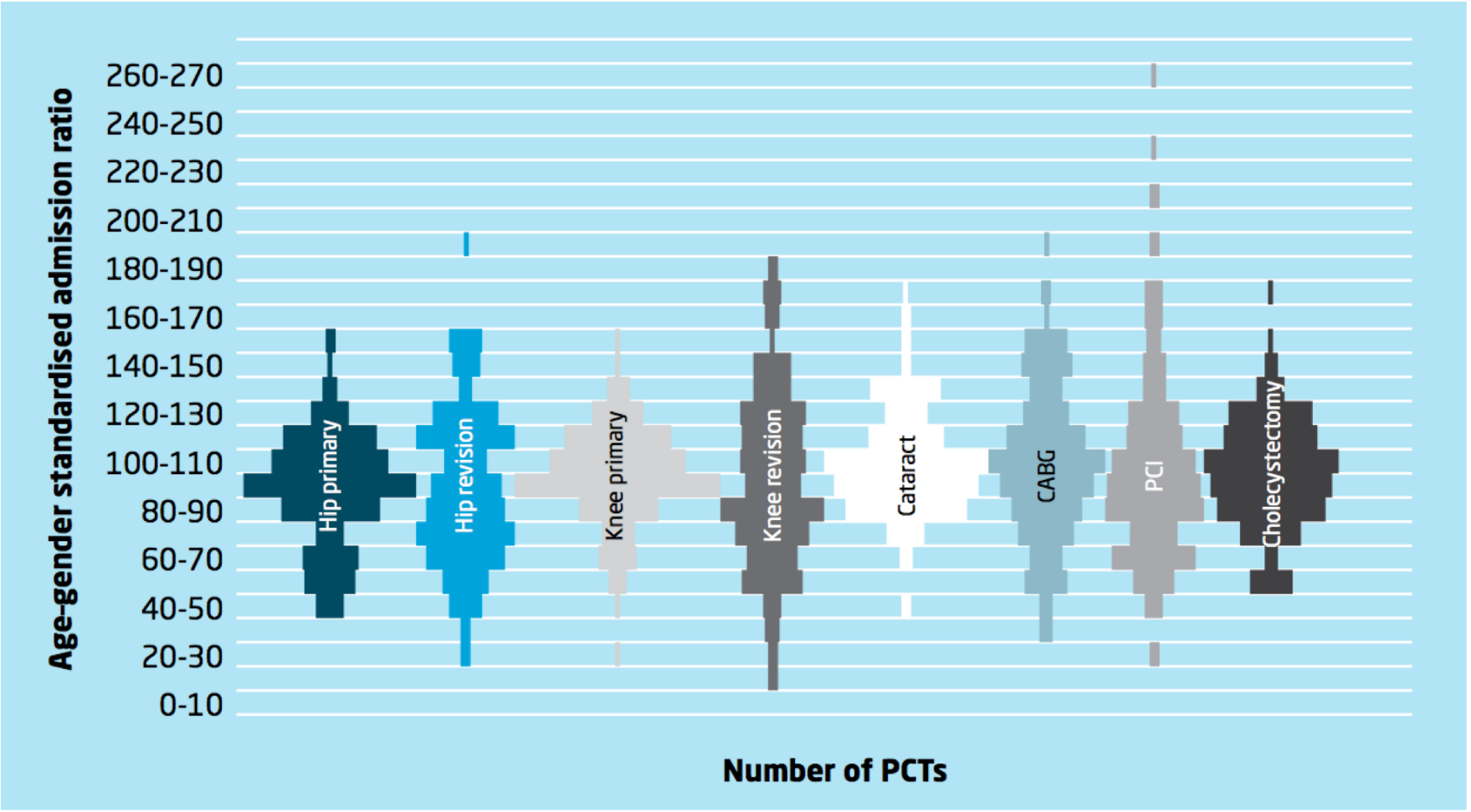
Michael Freeman
INSEAD

Gatekeeping under Congestion:
An Empirical Study of Referral Errors in the Emergency Department

Joint work with:
Susan Robinson – Cambridge University Hospitals
Stefan Scholtes – Cambridge Judge Business School

Huge variation in health care use

INSEAD
The Business School for the World®



The King's Fund. Variations in Healthcare. 2011.

Warranted variation?

INSEAD
The Business School for the World®

- Clinical need
- Medical guidelines
- Informed patient choice
- Innovations in treatment or care

Other sources of variation

INSEAD
The Business School for the World®

- Clinical decision making
- Supply-sensitive
  - e.g. availability of beds, specialists,…
- Financial incentives

"The doctor's pen is the most expensive item in a hospital"

Dr Gareth Goodier
Former CEO of Cambridge University Hospitals

Other sources of variation

INSEAD
The Business School for the World®

- Clinical decision making
- Supply-sensitive
  - e.g. availability of beds, specialists,…
- Financial incentives

"The doctor's pen is the most expensive item in a hospital"

Dr Gareth Goodier
Former CEO of Cambridge University Hospitals

Decision making in the ED

INSEAD
The Business School for the World®

- Emergency providers make disposition decisions ~350,000 times/day in US EDs
  - Option 1: admit patient to hospital
  - Option 2: discharge patient home

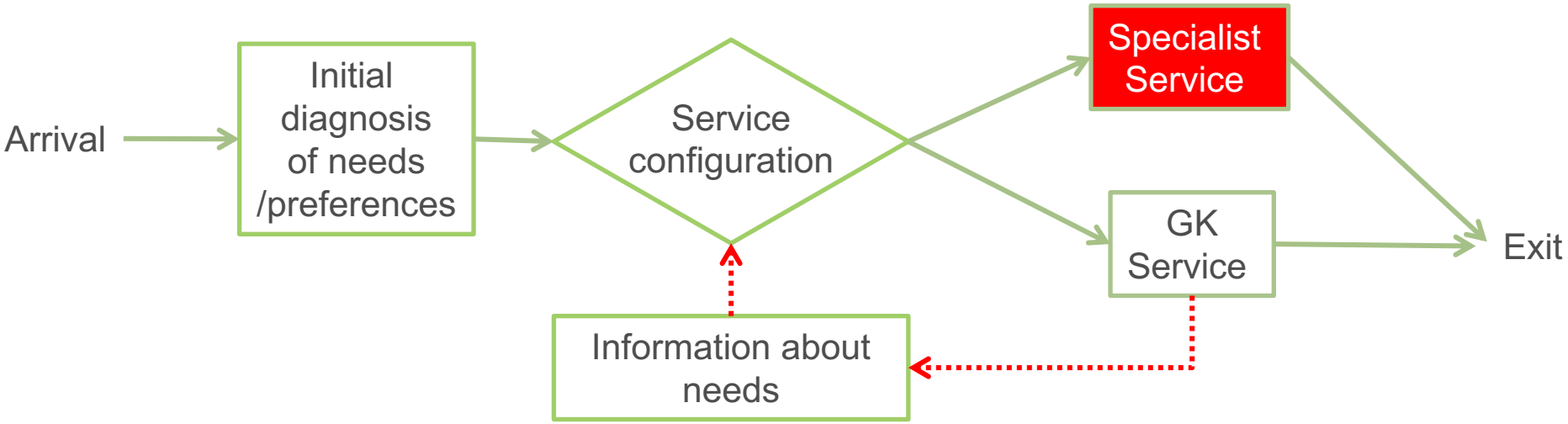
    → ED physicians act as **gatekeepers** to inpatient beds

- Significant variation in admission rates (gatekeeping referral rates) across EDs:
  - Pines at al. (2013 MCRR): US ED admission rate varied from 9.8% to 25.8% at the 10th and 90th percentiles

MCRR: Medical Care Research and Review

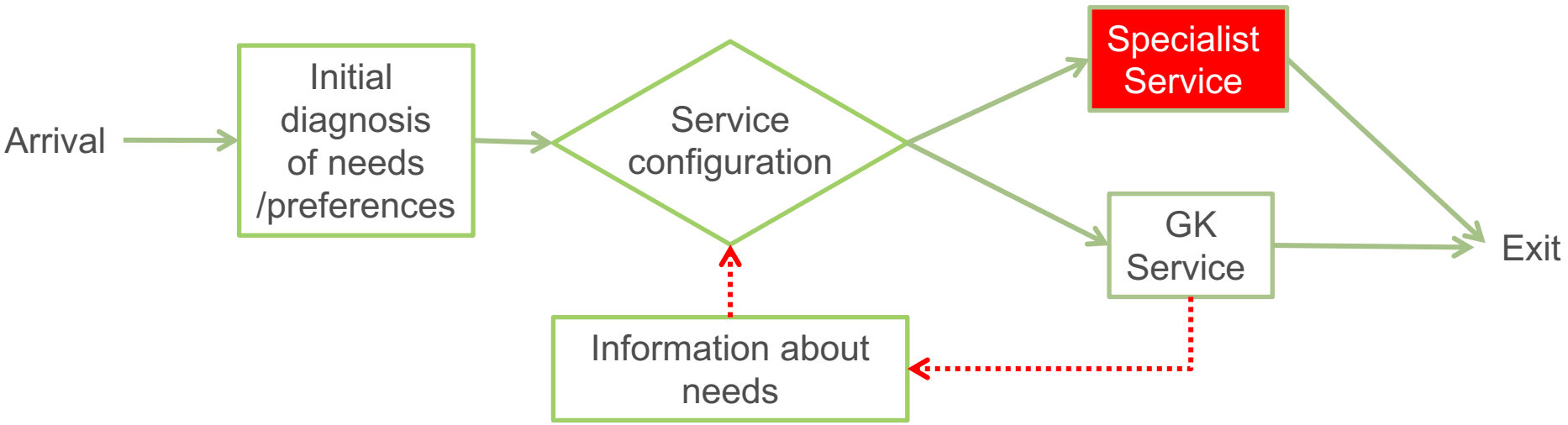
Gatekeeping (GK) Process

INSEAD
The Business School for the World®



Gatekeeping (GK) Process

INSEAD
The Business School for the World®



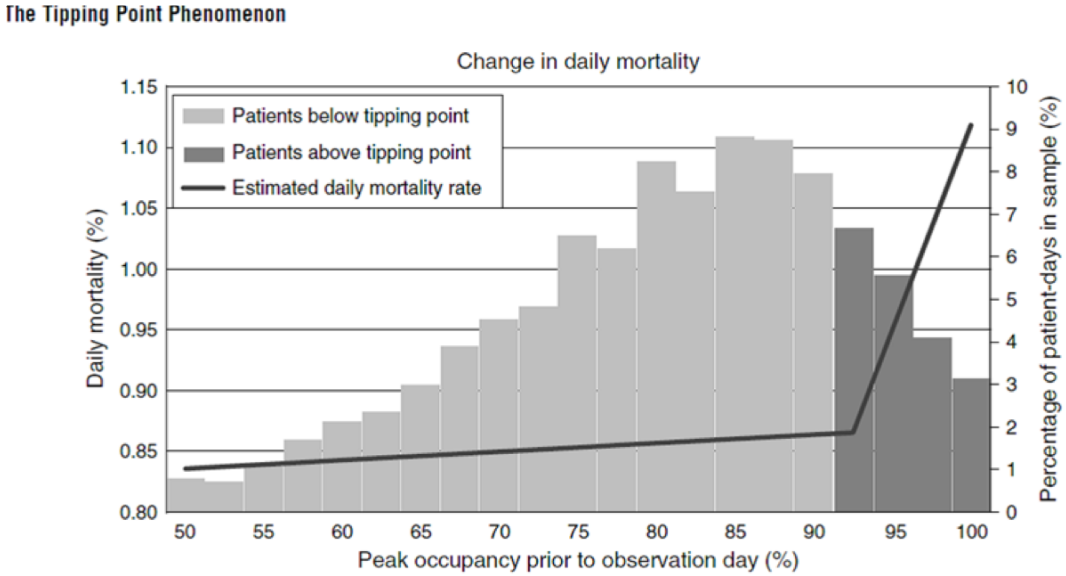
- Service configuration decision depends on customer needs & preferences
- Chance of specialist referral depends on customer complexity
- **Trade-off** (Shumsky and Pinker 2003)
  - Cost of specialist time
  - Cost of failing to resolving customer's problem

- GP systems pervasive beyond healthcare
  - Call centers, maintenance, restaurant waiting

Problems with hospital admission

INSEAD
The Business School for the World®

- Hospitals are dangerous places
  - Lack of mobility → physical and mental deterioration
  - Adverse events → infections, falls, medication errors
- Hospital admission is expensive
- Capacity (e.g. beds) is limited
- Unnecessary admission exposes other patients to risk



Kuntz et al. (Management Science 2015)
[~80,000 patients with STR,AMI,CHF,GIH,PNE,NOF]

The admission trade-off

INSEAD
The Business School for the World®

- Admission +
  - If patient unwell, increases chance of them receiving appropriate treatment
  - If problem not clear, increases chance of receiving more accurate diagnosis
- Admission –
  - Exposes patient to risk of hospital incident
  - Increases hospital crowding, exposing patients already in hospital to increased safety risk
  - Uses an expensive and constrained resource, may prevent another `more needy' patient from getting a bed

Challenges for gatekeeping in the ED

INSEAD
The Business School for the World®

- Emergency medicine: High levels of clinical uncertainty and variation in diagnostic accuracy

- Decision density high → can lead to elevated cognitive loading
  - Graber et al. (AIM 2005): cognitive factors contributed in 74% of cases of diagnostic error in the ED

- ED physicians under increasing time and workload induced pressure
  - US (1997 to 2007): ED visits grew at almost twice the rate of population growth
  - UK (1997 to 2012): ED visits grew by 47% compared to population growth of 10%

AIM: Archives of Internal Medicine

Research questions

INSEAD
The Business School for the World®

RQ1: How does congestion affect the accuracy of ED gatekeeping decisions?

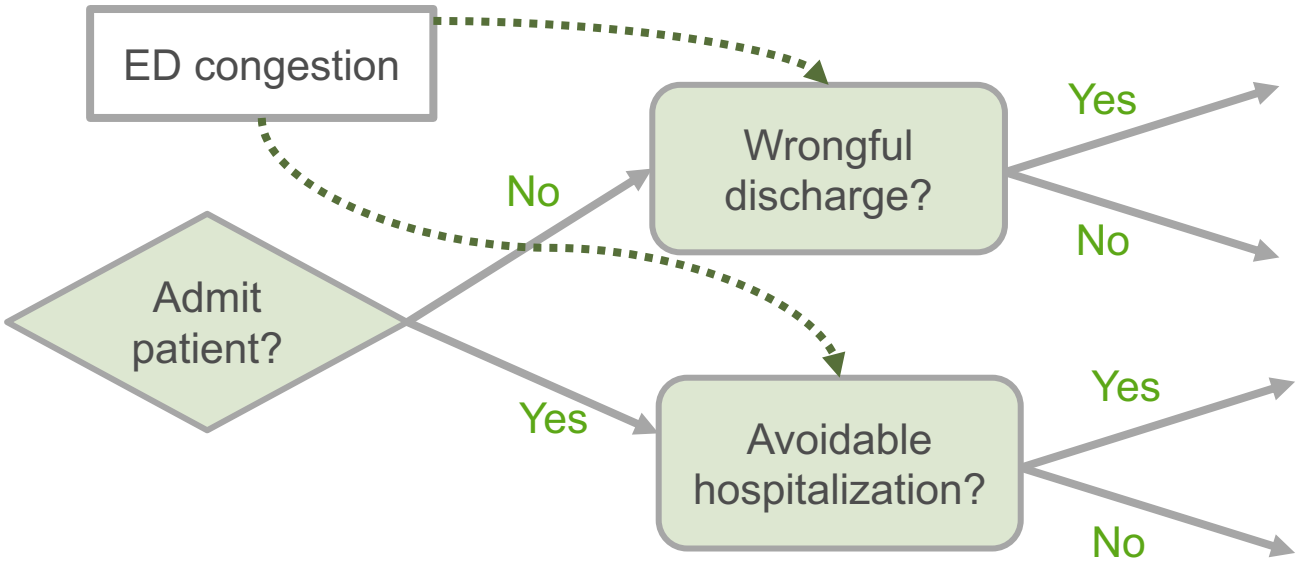
- **Wrongful discharges** (false negative)
- **Avoidable hospitalizations** (false positive)

Research questions

INSEAD
The Business School for the World®

RQ1: How does congestion affect the accuracy of ED gatekeeping decisions?

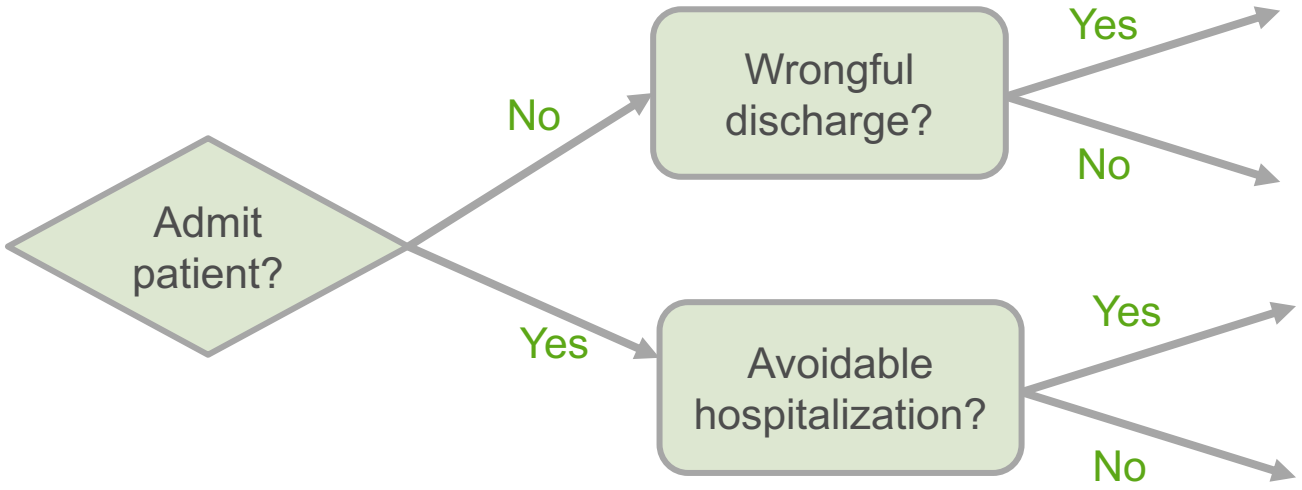
- **Wrongful discharges** (false negative)
- **Avoidable hospitalizations** (false positive)



Research questions

INSEAD
The Business School for the World®

RQ2: Are there process changes that can improve the accuracy of gatekeeping decisions?

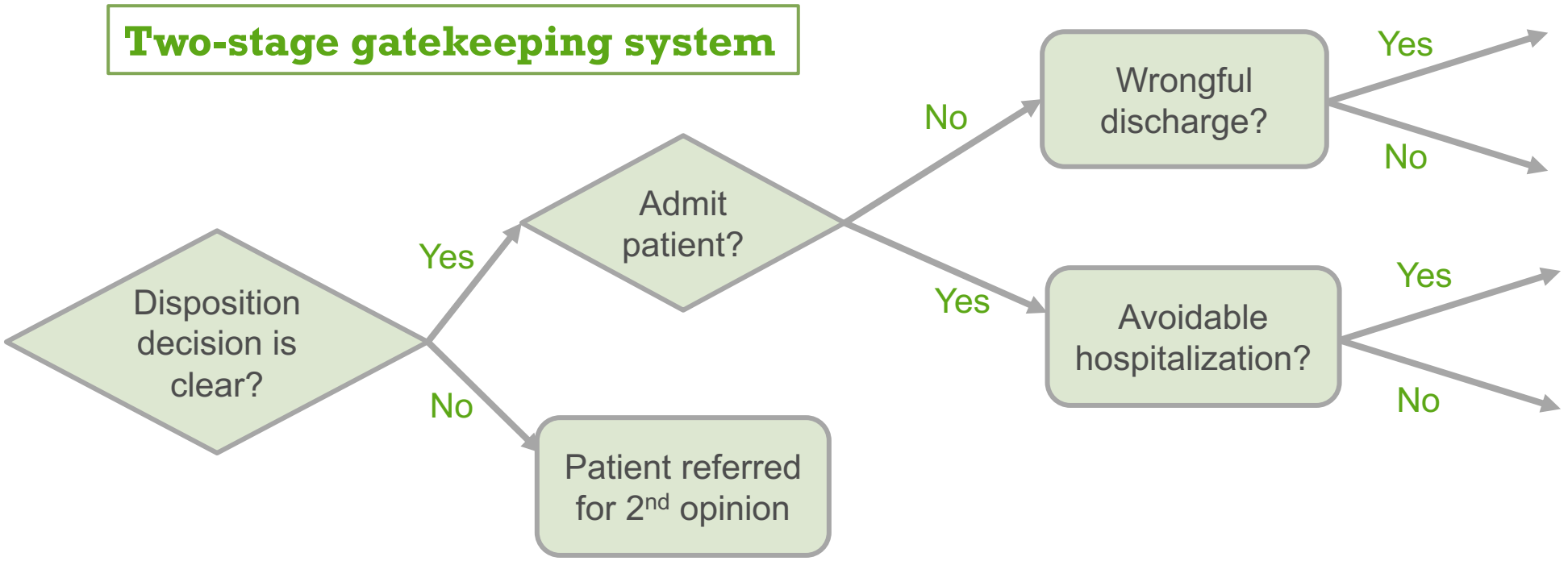


Single-stage gatekeeping system

Research questions

INSEAD
The Business School for the World®

RQ2: Are there process changes that can improve the accuracy of gatekeeping decisions?



Related literature

INSEAD
The Business School for the World®

- Gatekeeping
  - Shumsky & Pinker (2003) and Hasija et al. (2005): contracting for system optimal rate of referrals
  - Lee et al. (2012) and Zhang et al. (2011): outsourcing contracts and security-check performance
  - Freeman et al. (2016): study of workload induced changes in gatekeeping decisions in maternity care
- Type classification
  - Argon & Ziya (2009): model customer classification policies with imperfect information about customer type
  - Alizamir et al. (2013) and Wang et al. (2010): increasing service duration can improve accuracy of diagnosis, at cost of increased congestion
- Speed, quality & load
  - Anand et al. (2011) and Kostami & Rajagopalan (2013): service value increasing in duration, but cost of wait
  - Hopp et al. (2007): increasing capacity may increase congestion as discretionary service components added
  - Debo et al. (2008) and Paç & Veeraraghavan (2015): congestion acts as a deterrent to expert overtreatment
- Streaming/triage
  - Saghafian et al. (2012, 2014, 2017): triage can be augmented to stream ED patients based not only on their severity but also e.g. their complexity

Data

INSEAD
The Business School for the World®

- All visits to the ED of a large UK-based teaching hospital over a 7 year period (2007-2013)
  - Approx. 250 patient visits per day; 29% of patients admitted
  - Data set includes all inpatient info associated with those patients admitted via the ED
  - Subset to approx. 500k obs. of patients over the age of 16 who did not leave without being seen
- Exclude dates around public holidays when ED staffing and patient arrivals atypical
- Use first year of data as a run-in period to generate controls and measures of patient risk
- Leaves **~374k** obs. for analysis

Outline model

INSEAD
The Business School for the World®

*(**RQ1:** How does congestion affect the accuracy of ED gatekeeping decisions?)*

- **Wrongful discharges**

$$DischErr = \alpha_0 + \alpha_1 Congestion + \alpha_2 Controls + Error$$

- **Avoidable hospitalizations**

$$AdmErr = \beta_0 + \beta_1 Congestion + \beta_2 Controls + Error$$

- **Total errors = wrongful discharges + avoidable hospitalizations**

$$TotErr = \gamma_0 + \gamma_1 Congestion + \gamma_2 Controls + Error$$

Next few slides:

- Dependent variables (false positives and false negatives)
- Independent variable (time pressure)
- Controls
- Model error structure

Dependent variables

INSEAD
The Business School for the World®

- **Wrongful discharges**
  - Patient discharged home from ED but revisits ED within 7 days and is then admitted to the hospital
  - 1.0% of ED visits and 1.5% of all patients discharged

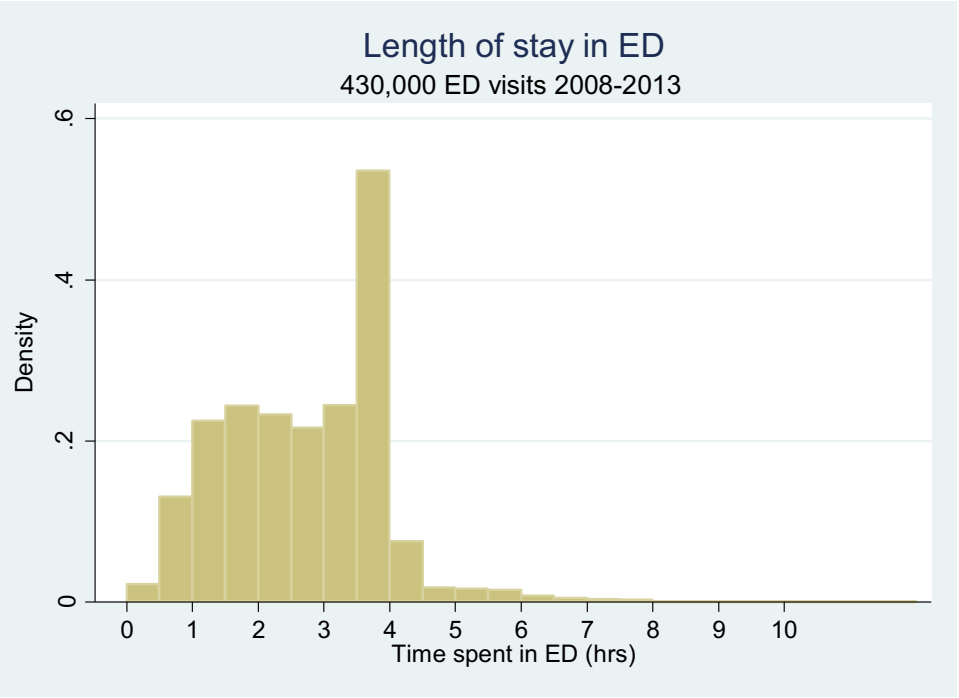
- **Avoidable hospitalizations**
  - Patient admitted to an inpatient unit and discharged within 24hrs without treatment
  - 4.3% of ED visits and 13.7% of all admissions
  - Change in rate indicative of change in likelihood of false admissions

Independent variable

INSEAD
The Business School for the World®

- The IV of interest in this study is congestion. Why?
- Waiting time targets in UK ED's mean congestion and service time highly correlated:
    - 95% of patients must be out of ED within 4 hours of arrival
    - Failure to achieve this in any month attracts a fine of £200 per breach

"When we are crowded we have two competing problems - we know we should not admit patients unnecessarily, yet we have to avoid breaching the ED waiting time target. At the moment avoiding a breach seems to be of higher priority"



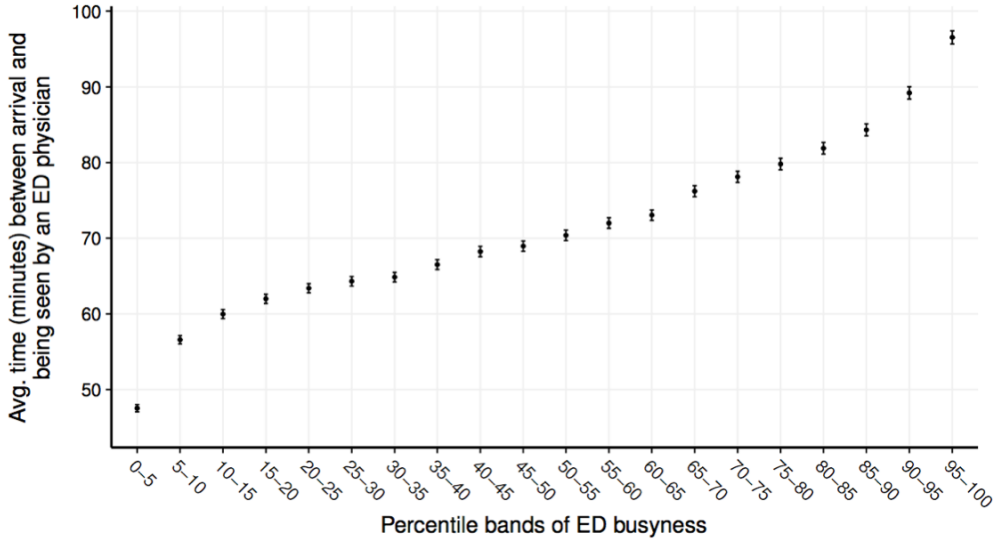
Congestion creates time pressure

INSEAD
The Business School for the World®

- Time available for service before breaching 4 hour target reduces with ED congestion.

  - ~25% reduction in avg. time available for diagnosis between first to last %ile bands

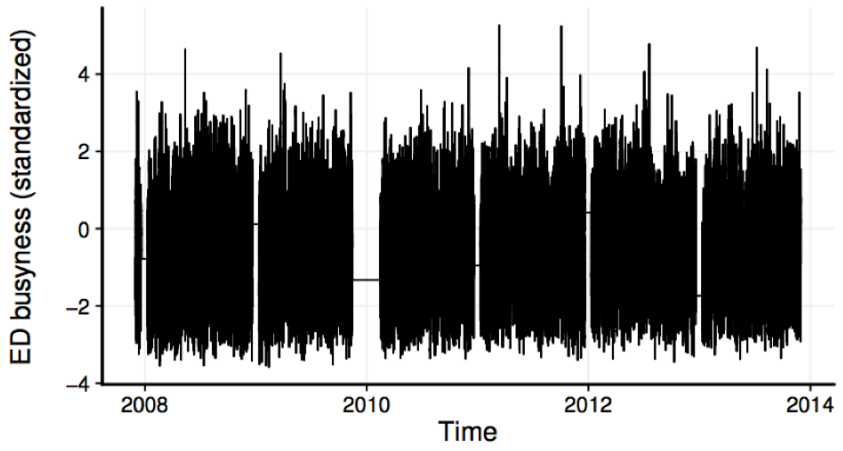
  - Consequence: ED physicians must make gatekeeping decisions under increased time pressure & uncertainty

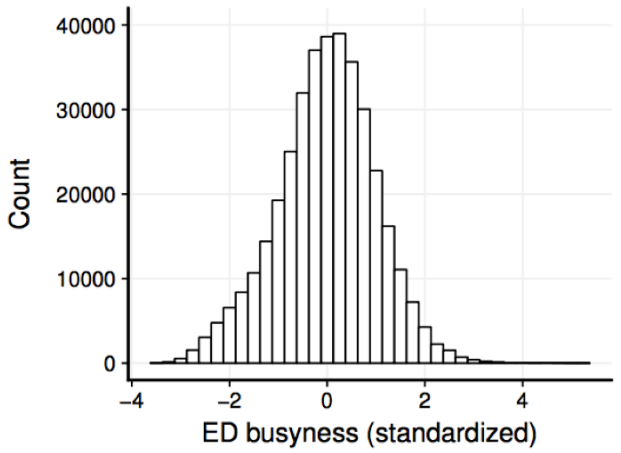


Measuring congestion

INSEAD
The Business School for the World®

- Observed occupancy in the ED at hour h: $QueueED_h$
- Proxy capacity by predicting using quantile regression the expected 95th percentile of occupancy: $QueueED_h^{95th}$
  - In QR, control for temporal effects (year, quarter, weekend, hour in 4-hour windows)
- ED congestion = $QueueED_h / QueueED_h^{95th}$
- Standardize by subtracting the mean and dividing by the SD





Controls

INSEAD
The Business School for the World®

	Type	Description
Temporal ($\mathbf{T}_i$)		
Year	Categorical (6)	Observation year (offset by one month so e.g. December '07 falls in '08), 2008 through 2013
Daily time trend	Continuous	A variable that takes value one on the first observation date and increases in value by one per day
Month	Categorical (12)	Month of the year in which the visit falls, January through December
School break	Categorical (7)	If visit occurs during a school break, equals the break type (e.g., Easter, Fall), else set to None
Day of week	Categorical (7)	Specifies the day of the week on which the visit occurred, Monday through Sunday
Window of arrival x weekend	Categorical (24)	A two-hourly arrival window (e.g., 2am to 4am) for weekdays, and a separate one for weekends
Patient and diagnosis related factors ($\mathbf{D}_i$)		
Age bands	Categorical (10)	The age of the patient, split into 10-year age bands (e.g., 10-20, 20-30, 100+)
Gender	Binary	A variable equal to one if the patient is male, else zero
Triage category	Categorical (7)	The triage level assigned to the patient on ED arrival
Initial severity assessment	Categorical (7)	The nature of the patient's condition (e.g., minor injuries, requires resuscitation, etc.)
Reason for ED visit	Categorical (32)	The reason for the ED episode (e.g., fall, burn, road traffic accident, etc.)
Contexual factors ($\mathbf{C}_i$)		
Mode of arrival	Categorical (8)	The mode of transport used to get to the hospital (e.g., helicopter, ambulance)
ED visits, last year	Continuous	The number of times the patient visited the ED in the previous 12 months
ED visits, last month	Continuous	The number of times the patient visited the ED in the previous one month
Admissions per ED visit, last year	Continuous	The proportion of hospital admissions to ED visits in the previous 12 months
Admissions per ED visit, last month	Continuous	The proportion of hospital admissions to ED visits in the previous month
Zero ED visits, last year	Binary	A variable equal to one if the patient did not attend the ED in the previous 12 months, else zero
Zero ED visits, last month	Binary	A variable equal to one if the patient did not attend the ED in the previous month, else zero
Physician related factors ($\mathbf{P}_i$)		
Historic physician error rate	Continuous	The short-stay observational admission, wrongful discharge, or total gatekeeping errors propensity of the assigned physician, calculated as in Appendix B
Physician category	Categorical (14)	Specifies the type of physician (e.g., orthopedic, plastics) for 33% of the visits where the physician name is not specified due to treatment being provided by a junior (non-consultant grade) physician
Operational/other factors ($\mathbf{O}_i$)		
Hospital congestion	Continuous	The overall busyness of the main hospital inpatient units in to which ED patients are admitted, calculated using the same method as for ED congestion in Section 5.1

Notes: If a patient did not visit the ED in the previous 12 months (or month) then the "Admission per ED visit, last year" ("last month") variable is set equal to zero.

Hypotheses 1 & 2: Gatekeeping accuracy

INSEAD
The Business School for the World®

RQ-1: How does congestion affect the accuracy of gatekeeping decisions

- Congestion induces time pressure that reduces accuracy of diagnosis (e.g. Alizamir et al. 2013) meaning that ED physicians have to make gatekeeping decisions under increased clinical uncertainty
- ED physicians adjust their service to trade off time spent with an individual patient versus improved system throughput

Hypothesis 1 (error-making hypothesis)

As system congestion increases, ED physicians make more errors in their referral decisions

- Clinical uncertainty: increases classification errors
- Cognitive overloading: impairs medical decision making

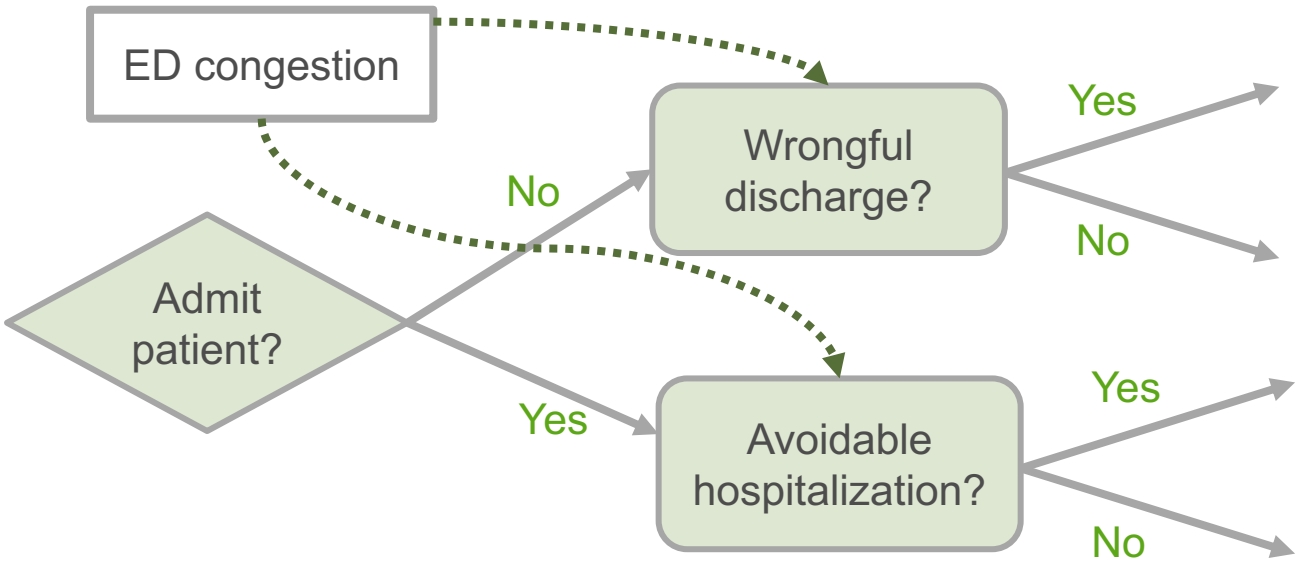
Hypothesis 2 (over-response hypothesis)

As system congestion increases, there will be an increase in the rate of avoidable hospitalizations relative to wrongful discharges, since ED physicians weight the cost of the latter more heavily

- Asymmetric disutilities: *"No-one has ever been sued for admitting a patient to hospital"*
- Safety first principle: Minimize risk of a "catastrophe"

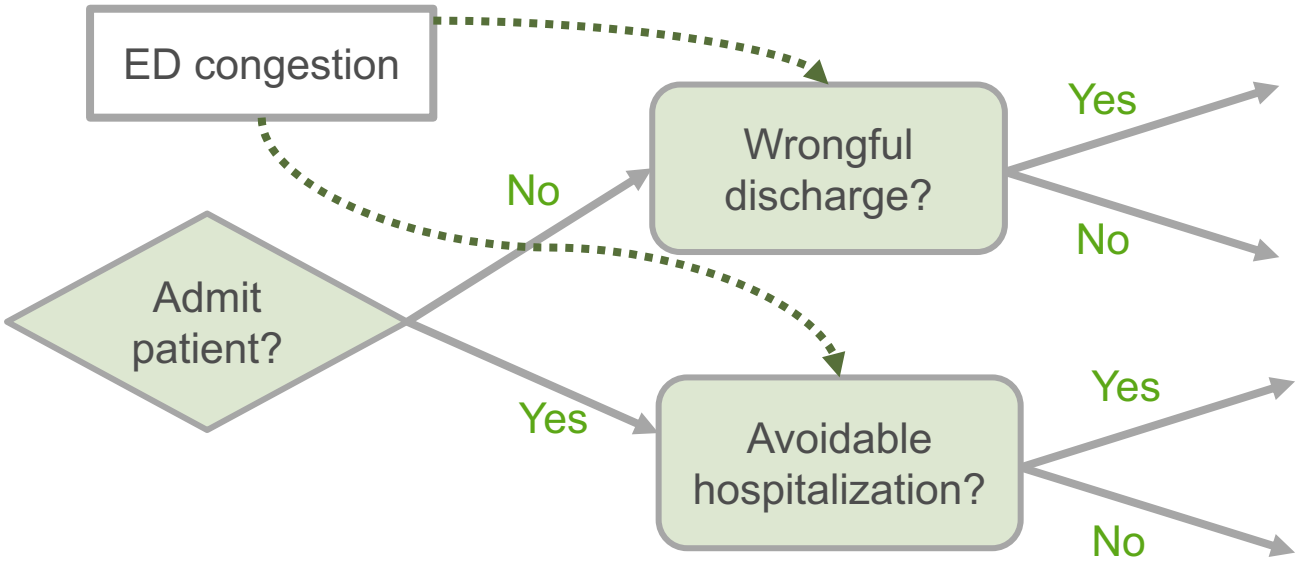
Model: Standard probit

INSEAD
The Business School for the World®



Model: Standard probit

INSEAD
The Business School for the World®



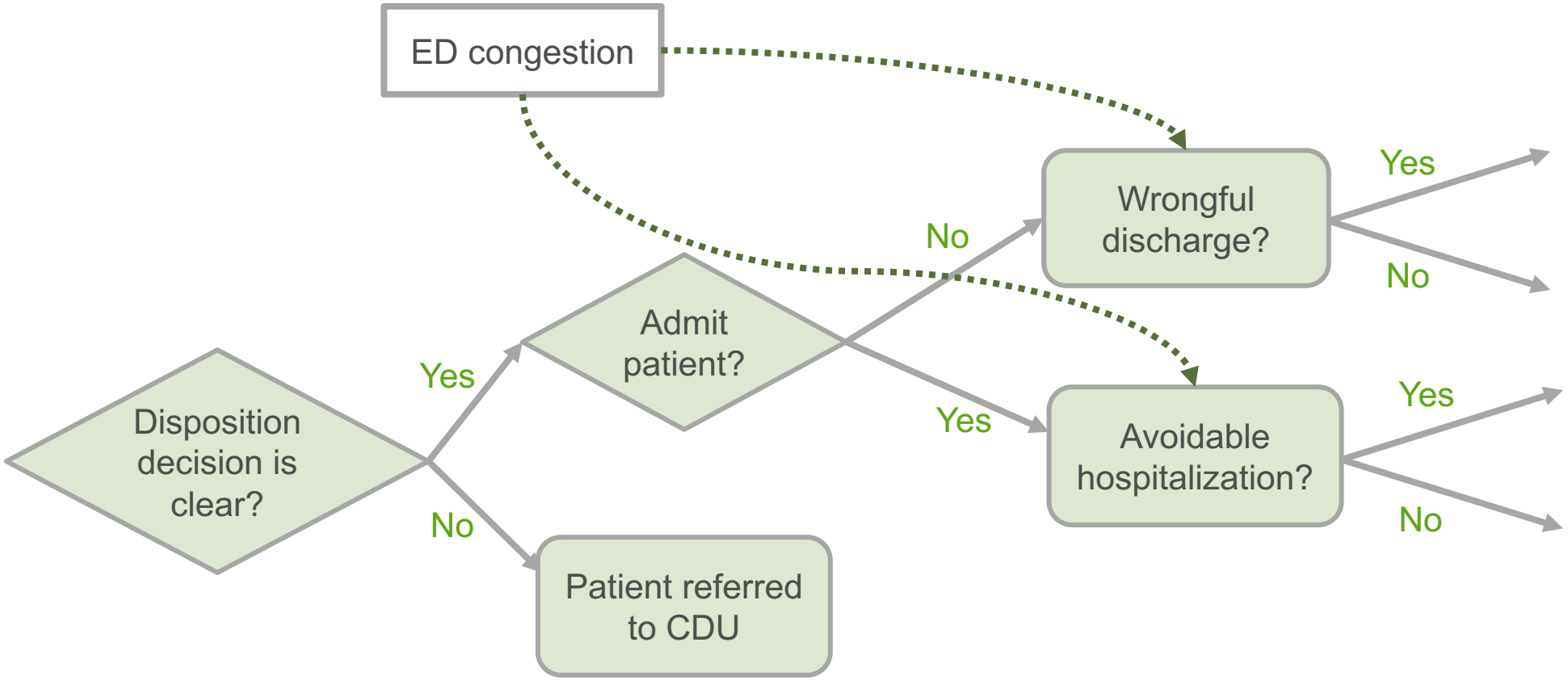
Tot. Error	Disch. Error	Adm. Error
1	1	0
0	0	0
1	0	1
0	0	0

$$y_i^* = \alpha + u_i\beta + x_i\gamma + \varepsilon_i, \quad \varepsilon_i \sim N(0,1)$$
$$y_i^{\text{outcome}} = 1(y_i^* > 0)$$

(where u_i is ED crowding)

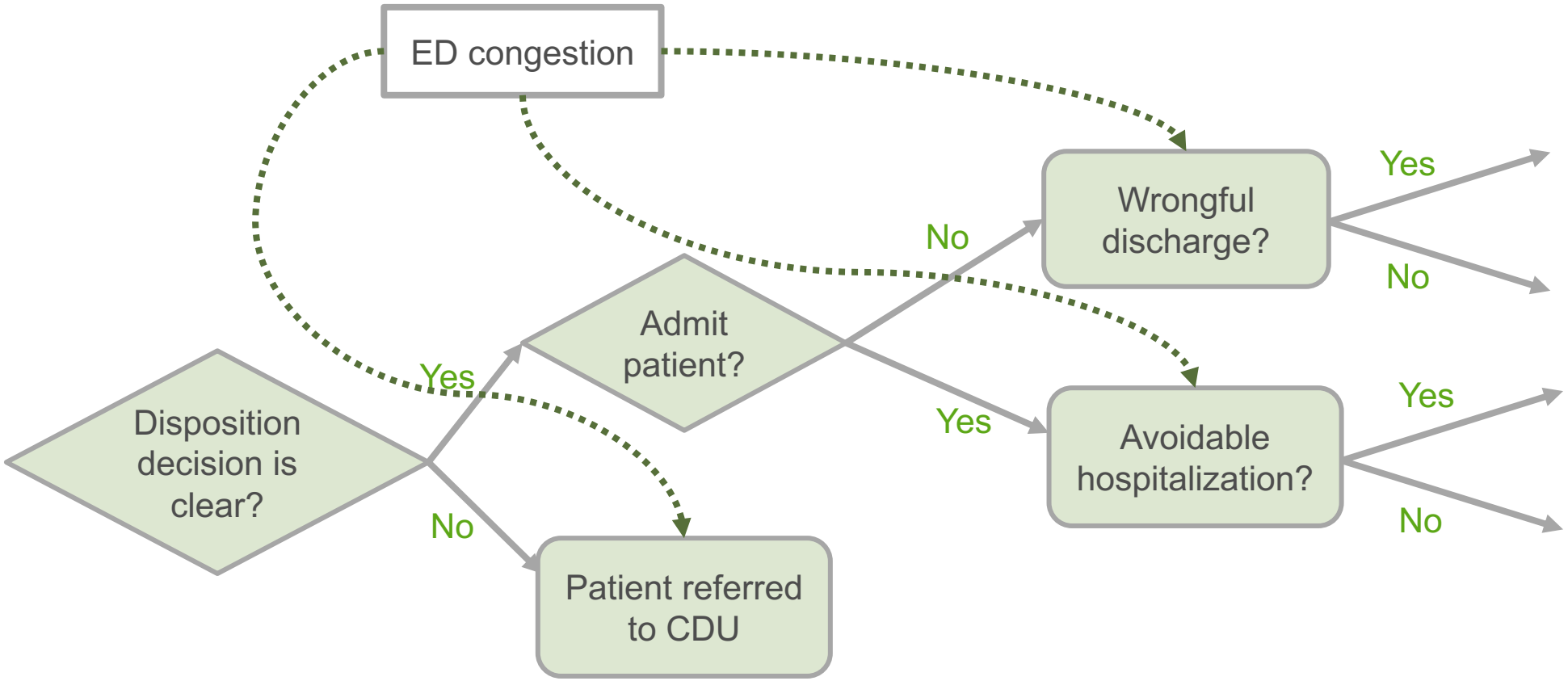
Modeling challenge

INSEAD
The Business School for the World®



Modeling challenge

INSEAD
The Business School for the World®



Heckman probit model

INSEAD
The Business School for the World®

$$y_i^{\text{outcome}} = 1(\alpha + u_i\beta + x_i\gamma + \varepsilon_i > 0)$$
$$y_i^{\text{select}} = 1(\alpha' + u_i\beta' + x_i\gamma' + z_i\delta' + \varepsilon_i' > 0)$$
$$(\varepsilon_i, \varepsilon_i') \sim N\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

(where u_i is ED crowding)

- $y_i^{\text{select}} = 1$ if the patient is **not** referred to the CDU
- Censoring assumption: y_i^{outcome} is not observed when $y_i^{\text{select}} = 0$
- Consistent and asymptotically efficient coefficient estimates when $\rho \neq 0$ (Van de Ven / Van Praag 1981)

Instrumental variables (I)

INSEAD
The Business School for the World®

Model estimation improved and coefficients more reliable when IVs provided

IV 1: Propensity of assigned ED physician to refer patients into CDU over previous year

Relevance:

- If a patient is assigned to a physician who has a history of referring more patients to the CDU, they are more likely to be referred there themselves

Validity:

- Include controls for the physician's historic error rates over the same period
- After controlling for physician-specific historic error rates, the CDU referral rate of the assigned physician is uncorrelated with the residuals

Instrumental variables (II)

INSEAD
The Business School for the World®

Model estimation improved and coefficients more reliable when IVs provided

IV 2: Congestion level in the CDU

Relevance:

- If the CDU is congested then it becomes less available to ED physicians as an option

Validity:

- For patients who are not admitted to the CDU, the busyness of the CDU should have no direct effect on their likelihood of being admitted or discharged in error

(Instrument validity can be tested with several instruments → all tests pass)

Congestion effect

INSEAD
The Business School for the World®

	Decision made by ED physicians		
	(1e) TotErr	(2e) ObsAdm	(3e) DischErr
ED congestion	0.020^{***}	0.028^{***}	$-0.014^{\dagger}$
	(0.005)	(0.005)	(0.008)
ρ	-0.220^{***}	-0.114^{*}	-0.179^{**}
	(0.040)	(0.047)	(0.057)
N	373,663	373,663	373,663
N uncensored	337,144	337,144	337,144
Log-lik	−157,985	−147,023	−116,471

- A 1σ increase in ED congestion results in a:
  - 7.2% increase in both types of error
  - 7.7% increase in avoidable hospitalizations
  - 3.3% reduction in wrongful discharges

Consistent with Hypotheses 1 and 2

Hypothesis 3: Two-stage gatekeeping

INSEAD
The Business School for the World®

RQ-2: Are there process changes that improve the accuracy of gatekeeping decisions?

Hypothesis 3 (uncertainty-based streaming)

A two stage gatekeeping system reduces both types of gatekeeping error

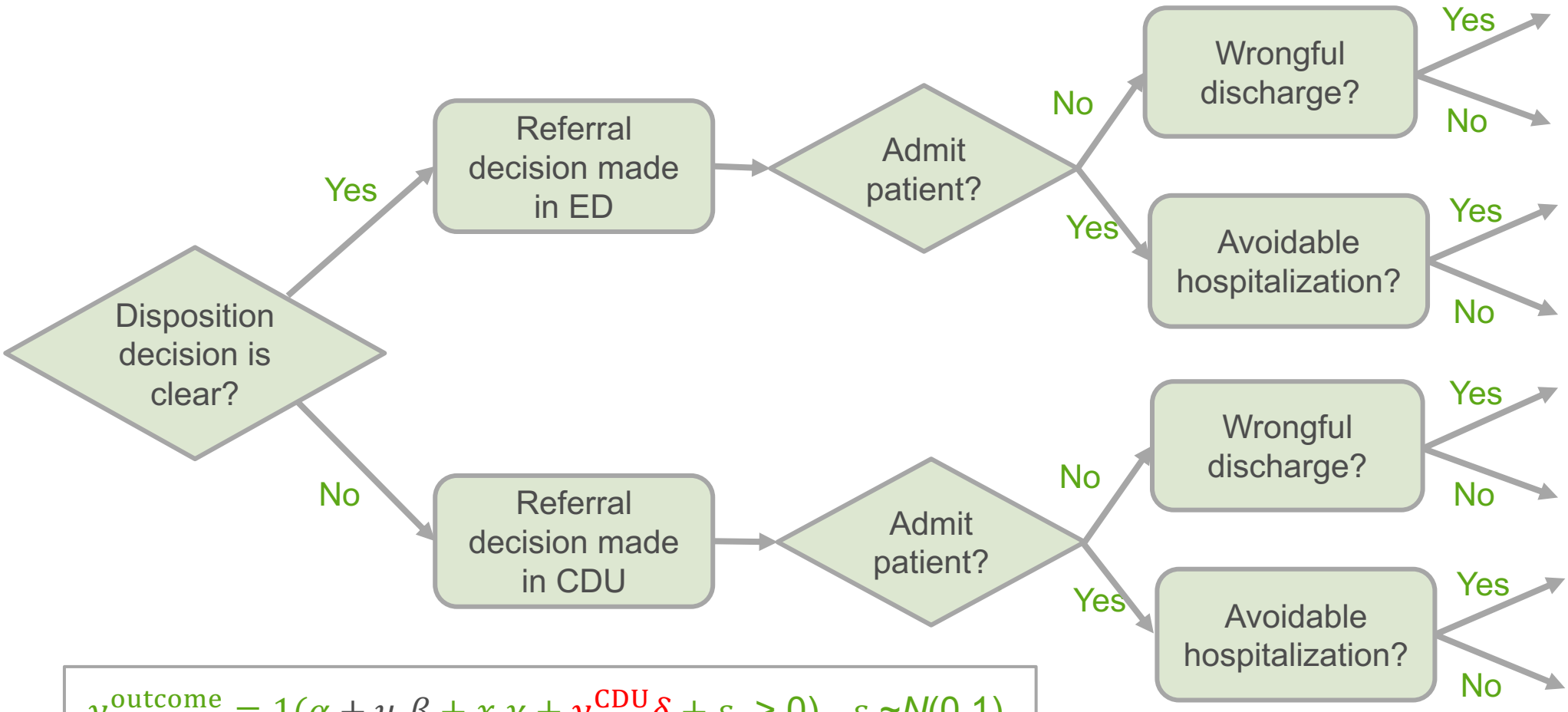
The two-stage GK system improves the match between customers with heterogeneous needs with gatekeepers with heterogeneous experience and resources

Streaming/triage

- Assigning customers different priority classes may be beneficial when they differ sufficiently in their service requirements – e.g. Mandelbaum & Reiman (1998), Dijk & Sluis (2008)
- Saghafian et al. (2012) and Saghafian et al. (2014): triage can be augmented to stream ED patients based not only on their severity but also using their (i) likelihood of being admitted and (ii) their complexity

Two-stage gatekeeping system

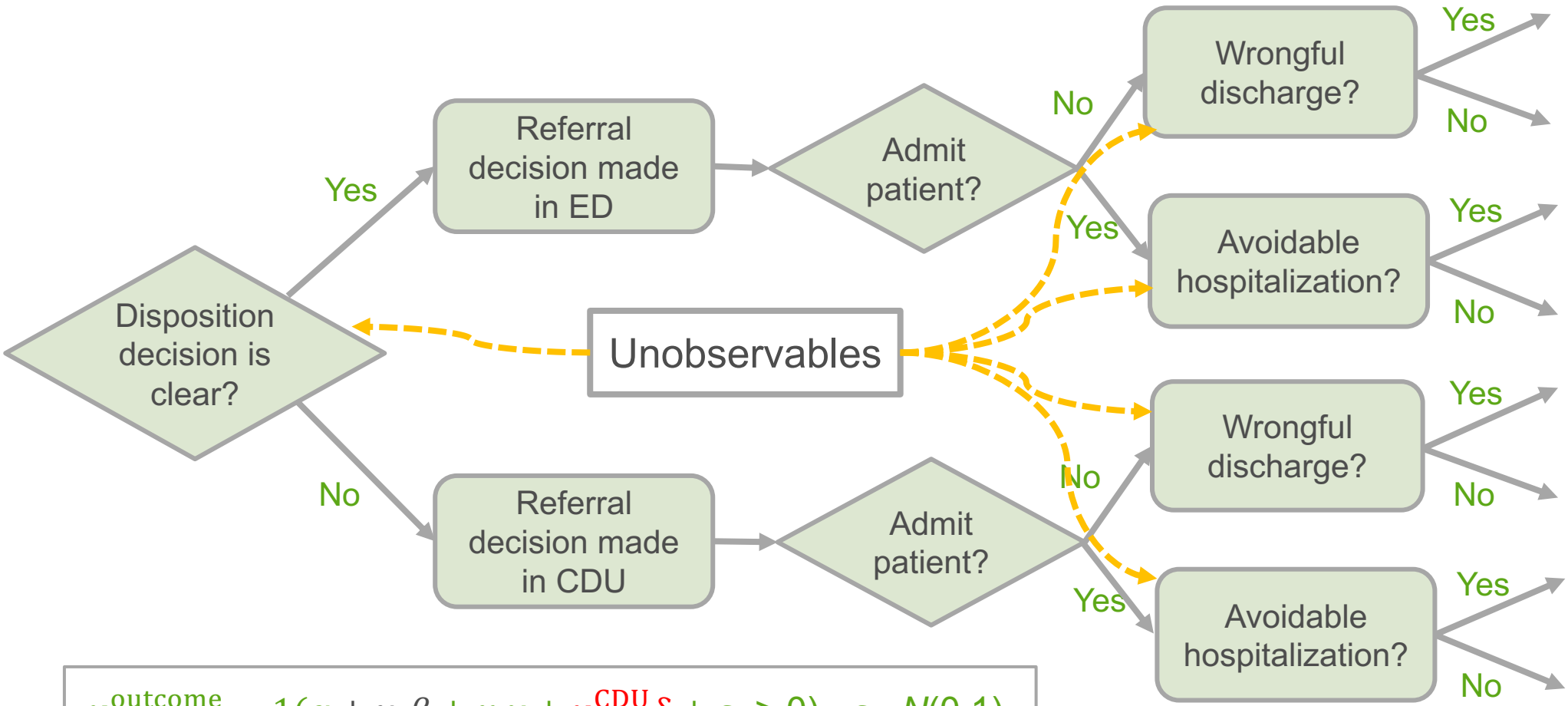
INSEAD
The Business School for the World®



$$y_i^{\text{outcome}} = 1(\alpha + u_i\beta + x_i\gamma + y_i^{\text{CDU}}\delta + \varepsilon_i > 0), \quad \varepsilon_i \sim N(0,1)$$

Two-stage gatekeeping system

INSEAD
The Business School for the World®



$$y_i^{\text{outcome}} = 1(\alpha + u_i\beta + x_i\gamma + y_i^{\text{CDU}}\delta + \varepsilon_i > 0), \quad \varepsilon_i \sim N(0,1)$$

Recursive bivariate probit

INSEAD
The Business School for the World®

Heckprobit

$$y_i^{\text{outcome}} = 1(\alpha + u_i\beta + x_i\gamma + \varepsilon_i > 0)$$
$$y_i^{\text{select}} = 1(\alpha' + u_i\beta' + x_i\gamma' + z_i\delta' + \varepsilon_i' > 0)$$
$$(\varepsilon_i, \varepsilon_i') \sim N\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}.$$

Effect of interest: impact of congestion on referral decisions of the first-stage GK

Coefficient of interest: β

Recursive bivariate probit

$$y_i^{\text{outcome}} = 1\left(\alpha + u_i\beta + x_i\gamma + y_i^{\text{CDU}}\delta + \varepsilon_i > 0\right)$$
$$y_i^{\text{CDU}} = 1(\alpha' + u_i\beta' + x_i\gamma' + z_i\delta' + \varepsilon_i' > 0)$$
$$(\varepsilon_i, \varepsilon_i') \sim N\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}.$$

Effect of interest: whether disposition decisions improve when patients referred through second-stage (CDU) GK

Coefficient of interest: δ

Benefits of two-stage gatekeeping system

INSEAD
The Business School for the World®

	No patient routed through CDU	All patients routed through CDU
Avoidable hospitalizations	5.23%	1.76%
Wrongful discharges	1.16%	0.67%

Avg. treatment effect (**ATE**)

- Avoidable hosp. = -3.5%
- Wrongful discharges = -0.49%

→ When patients routed through the two-stage system *both* errors go down

Benefits of two-stage gatekeeping system

INSEAD
The Business School for the World®

	No patient routed through CDU	All patients routed through CDU
Avoidable hospitalizations	5.23%	1.76%
Wrongful discharges	1.16%	0.67%

Avg. treatment effect (**ATE**)

- Avoidable hosp. = -3.5%
- Wrongful discharges = -0.49%

→ When patients routed through the two-stage system *both* errors go down

Avg. treatment effect on the treated (**ATT**)

- Avoidable hosp. = -9.3%
- Wrongful discharges = -1.2%

→ ED physicians especially good at identifying patients who would benefit most from CDU second opinion

More ED or CDU capacity?

INSEAD
The Business School for the World®

- **Counterfactual**: redeploy CDU capacity to the ED to reduce crowding
  - 1.5m patient hours in ED
  - 326,000 patient hours in CDU (=20% of ED hours)
- If there is no CDU, then adding 20% more capacity to ED reduces ED busyness by ~0.6σ:
  - leads to 0.14% reduction in avoidable hospitalizations (and slight increase in wrongful discharges)
- Keeping ED at its observed capacity and retain CDU:
  - leads to 1% reduction of avoidable hospitalizations
- **Why?**
  - Extra capacity in ED is only useful during busy periods; CDU is useful all the time
  - Extended service time in the CDU is provided only to those patients who benefit from it the most

- **Summary:** the net effect of the CDU in the study hospital, after accounting for the opportunity cost of its resources, is a relative reduction of the avoidable hospitalization rate by 16.5%

Summary

INSEAD
The Business School for the World®

- Gatekeepers frequently must take referral decisions under congestion induced time pressure, e.g.
  - PCP to specialist (e.g. 10 min appointment rule)
  - ED to inpatient (e.g. 4 hour waiting time target)
- Time targets are important because people tend to fill any amount of time they have with low-value-adding activities (see e.g. Hopp et al. 2007)

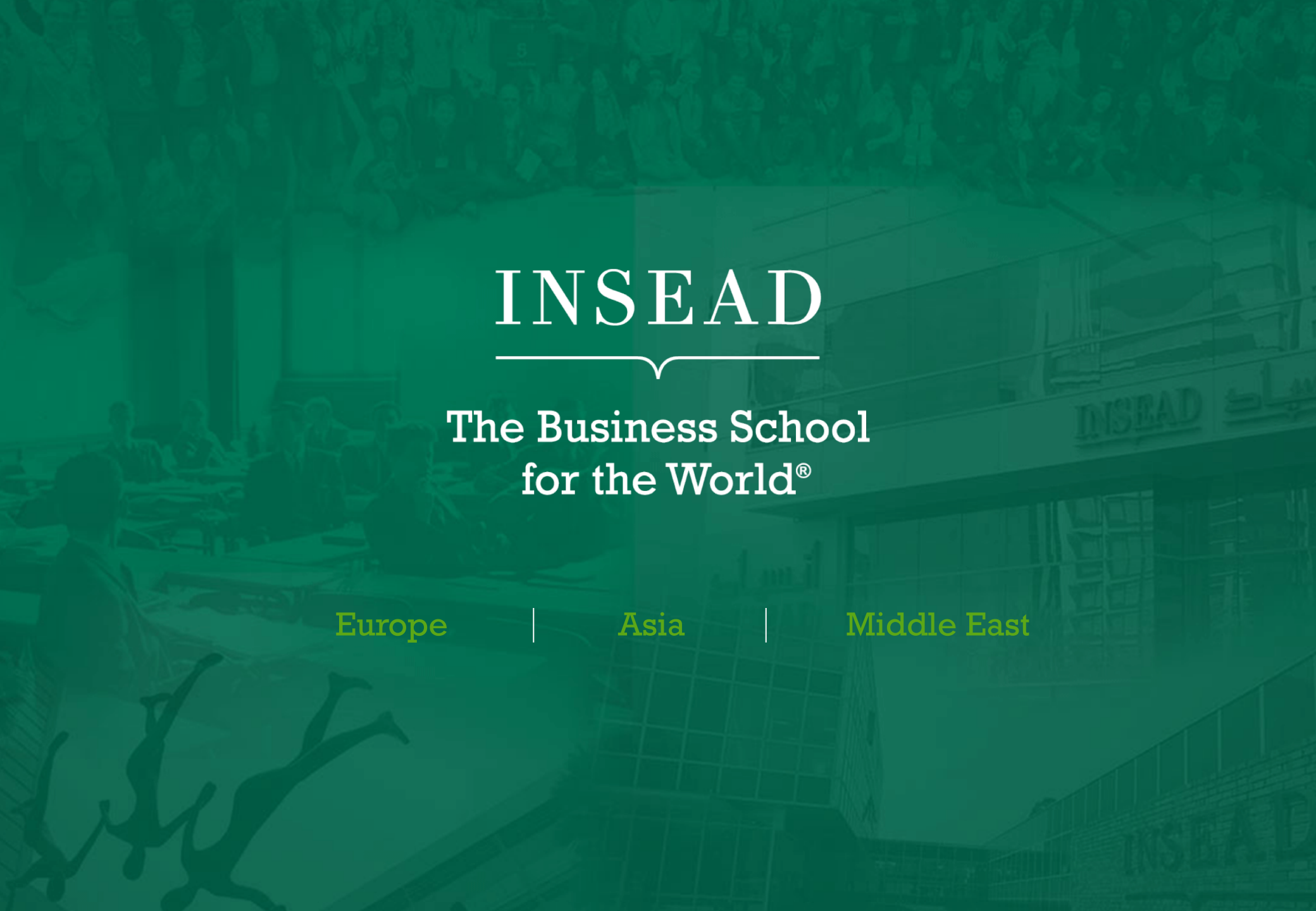
  → straight-forward decisions should be taken quickly
- However, time targets can be detrimental if they prevent people from taking time when needed
  - e.g. we show that ED physicians act cautiously when under pressure, increasing admission errors
  - the response of the gatekeeper can run counter to the system optimal response (e.g. ED bullwhip)
- A two-stage gatekeeping system coupled with a waiting time target provides incentives to triage out straight-forward cases in time but allows for "active delay" of more uncertain cases
  - e.g. it allows hospitals to keep a time target in place *and* reduce the discrepancy between the gatekeeper and the system optimal response

INSEAD

The Business School for the World®

Thank you!!

Questions?


INSEAD
The Business School
for the World®
Europe | Asia | Middle East