GATEKEEPING UNDER TIME PRESSURE:
AN EMPIRICAL STUDY OF REFERRAL DECISIONS IN THE ED

Michael Freeman
Cambridge Judge Business School

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

DECISION MAKING IN THE ED

- Emergency providers make disposition decisions ~350,000 times a day in the US

  - Option 1: <u>admit</u> patient into the hospital (leads to ~20m admission p.a. in US, nearly 50% of all admissions)

  - Option 2: <u>discharge</u> 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

PROBLEMS WITH HOSPITAL ADMISSION

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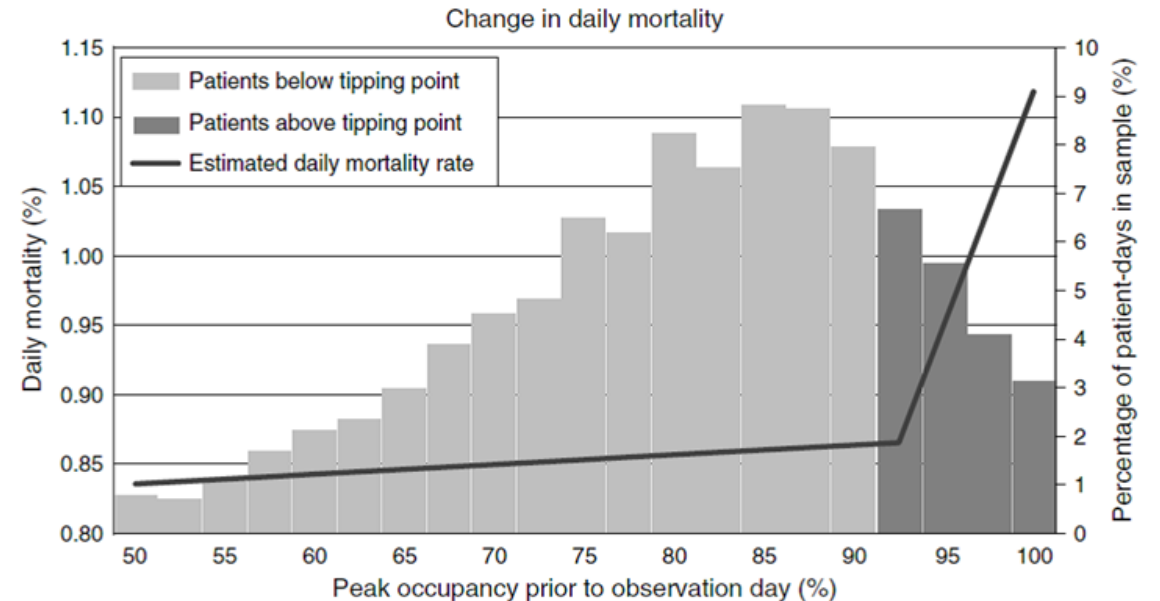
Kuntz et al. (Management Science 2015)
[~80,000 patients with STR,AMI,CHF,GIH,PNE,NOF]

THE ADMISSION TRADE-OFF

- 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

- 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

How do gatekeepers make decisions in a time-pressured environment such as the ED?

- **RQ-1:** How does increasing time pressure affect the accuracy of ED gatekeeping decisions?

    - **False negatives:** discharge patient from ED who should have been admitted to hospital

    - **False positives:** admit a patient to hospital unnecessarily

- **RQ-2:** Are there processes changes that can counter the adverse effects of time-pressure?

RELATED LITERATURE

- 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): empirical 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 and 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 congestions as discretionary components added to service

  - Debo et al. (2008) and Paç & Veeraraghavan (2015): congestion acts as a deterrent to expert overtreatment

HYPOTHESIS 1: GATEKEEPING ACCURACY

RQ-1: How does increasing time pressure affect the accuracy of ED gatekeeping decisions?

- Time pressure reduces accuracy of diagnosis (e.g. Alizamir et al. 2013) meaning that ED physicians have to make gatekeeping decisions under increased clinical uncertainty

<u>Hypothesis 1-A</u> (error-making hypothesis)

Clinical uncertainty: increases classification errors

Cognitive overloading: impairs medical decision making

- Admission errors: ↑

- Discharge errors: ↑

<u>Hypothesis 1-B</u> (over-response hypothesis)

Asymmetric risks: *"No-one has ever been sued for admitting a patient to hospital"*

Safety first principle: Minimize risk of a "catastrophe"

- Admission errors: ↑

- Discharge errors: = or ↓

HYPOTHESIS 2: TWO-STAGE GATEKEEPING

RQ-2: Are there processes changes that can counter the adverse effects of time-pressure?

<u>Hypothesis 2</u> (uncertainty-based streaming)

Streaming patients with uncertain disposition to an intermediate "semi-specialist" referral unit can improve the accuracy of gatekeeping decisions

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

DATA

- 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 **~375k** obs. for analysis.

OUTLINE MODEL

RQ-1: How does increased time pressure affect the accuracy of ED gatekeeping decisions?

False negatives: discharge patient from ED who should have been admitted to hospital

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

False positives: admit a patient to hospital unnecessarily

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

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

DEPENDENT VARIABLES

False negative = discharge error

- 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

False positive = short-stay admission

- Patient admitted to a ward and discharged within 24hrs without procedure code

- 4.3% of ED visits and 13.7% of all admissions

- Change in rate indicative of change in likelihood of false admissions

INDEPENDENT VARIABLE

- **Time pressure**: Could use service duration, e.g. time between patient first being seen by ED physician and time of disposition decision

- **Problem 1**: This is *highly* endogenous, i.e. service duration correlated with unobservables

  - Short service duration in the ED may be a simple case or high priority/critical

  - Long service duration in the ED may be a more complex case or low priority

- **Problem 2**: Very difficult to make the case that there exists a suitable instrumental variable

→ Better to use a *proxy* for time pressure that is exogenous

- **Solution**: We use a scaled and standardized measure of <u>ED crowding</u>

TARGETS IN UK EMERGENCY DEPT.

- Waiting time targets make UK ED's an ideal context to study effects of time pressure:

  - 95% of patients in our study hospital must be out of the ED within four 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"

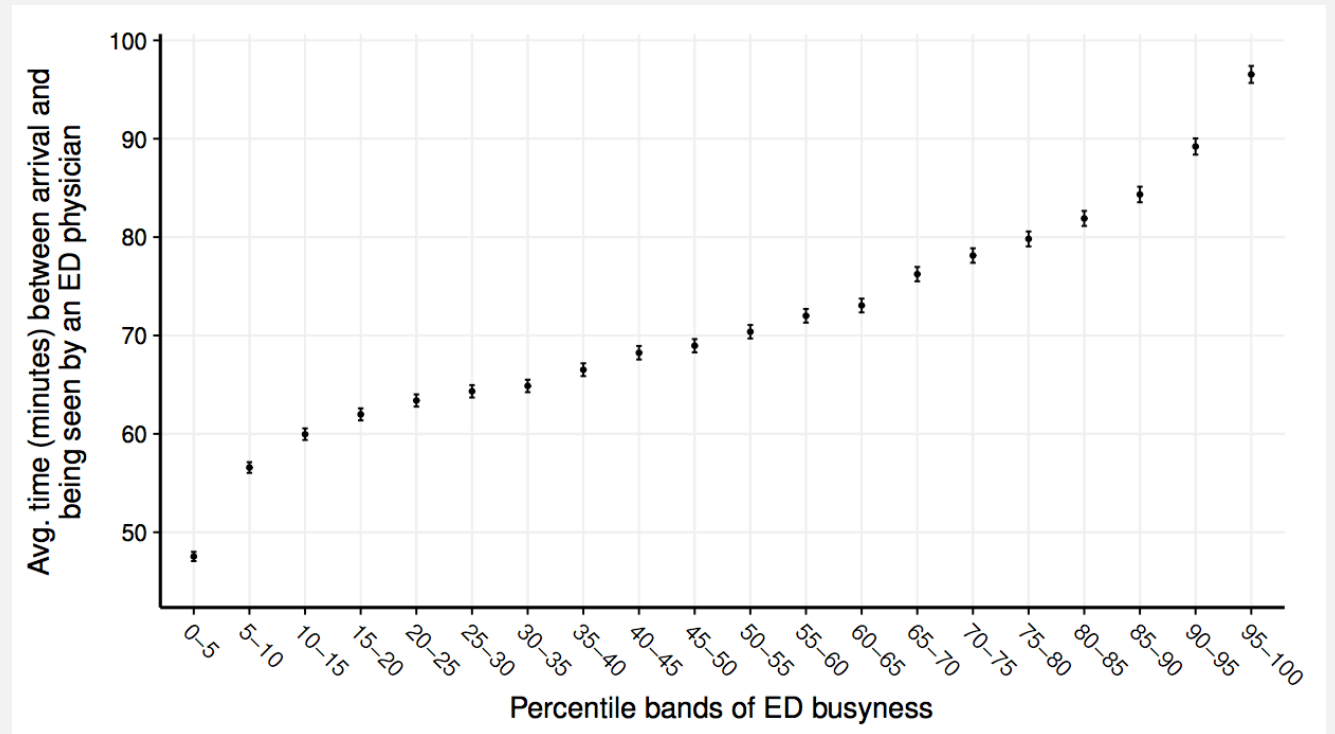Length of stay in ED
430,000 ED visits 2008-2013
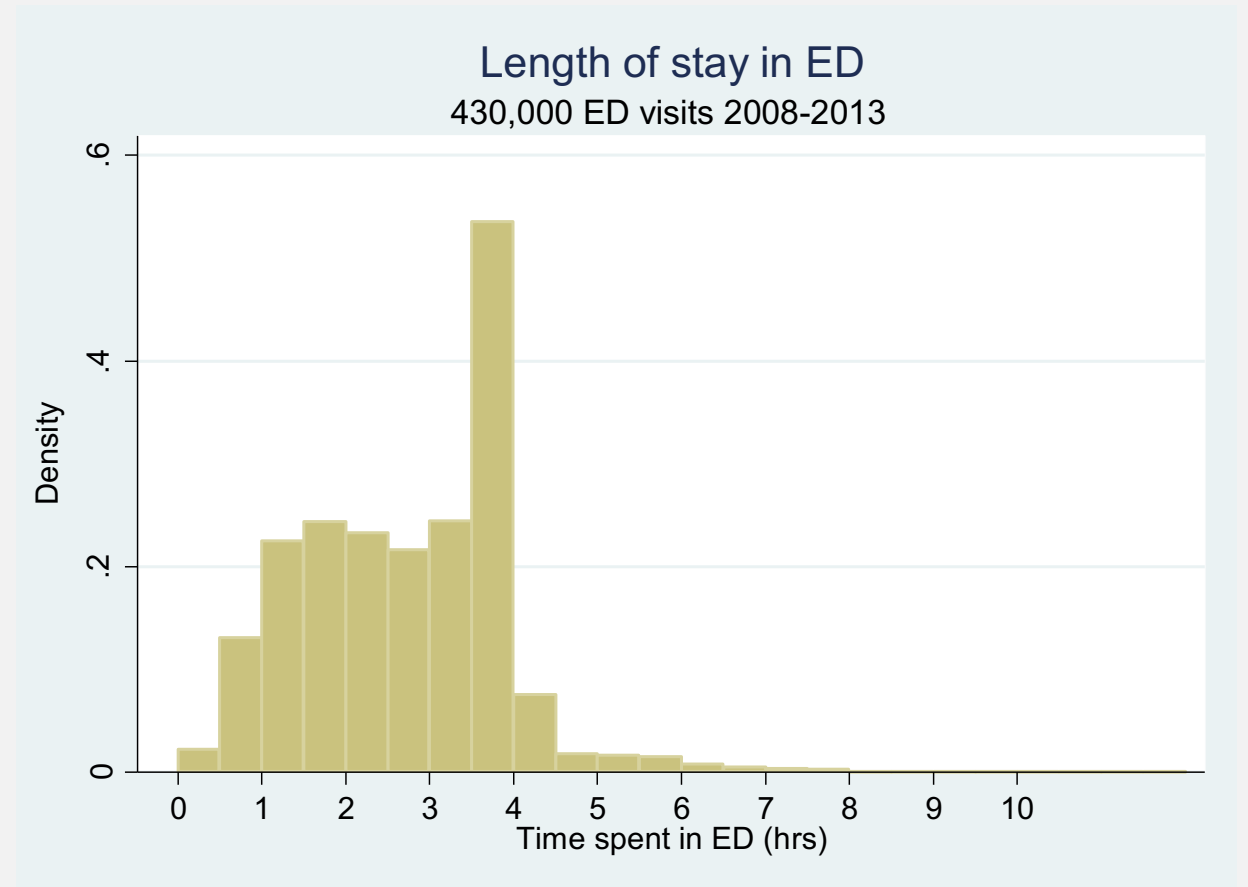
TIME PRESSURE ↑ WITH CROWDING

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

  - ~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

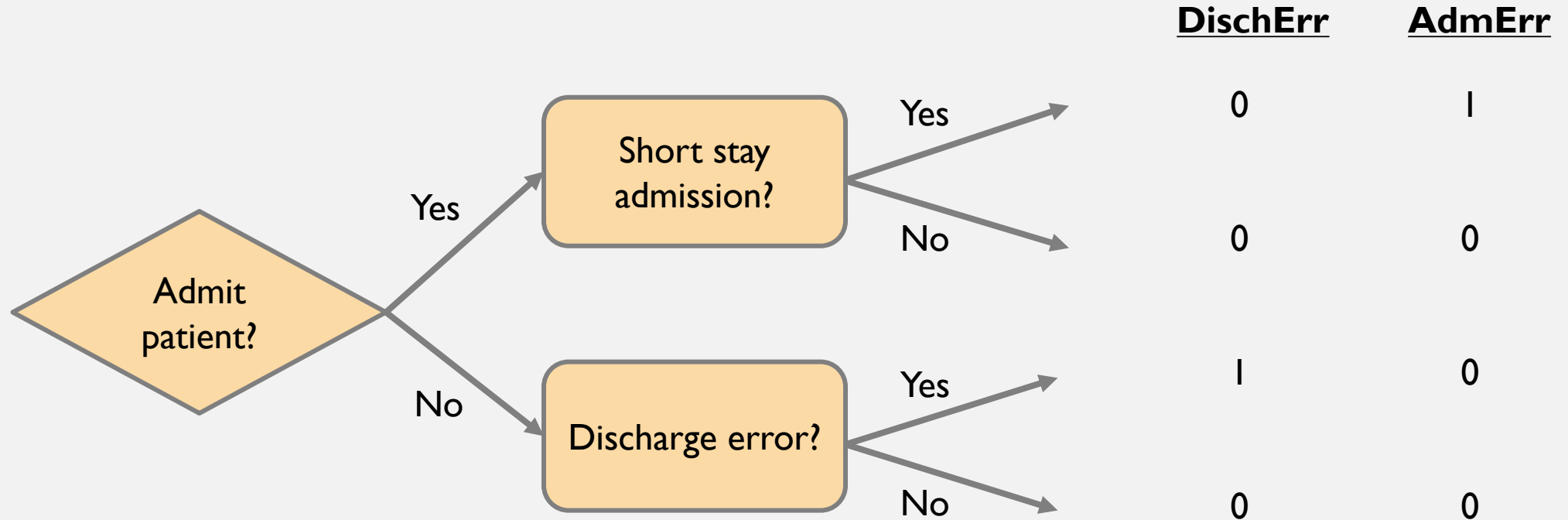 → ED crowding proxy for time-pressure

	Type	Description
Temporal (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 (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 patients condition (e.g., minor injuries, requires resuscitation, etc.)
Reason for ED visit	Categorical (32)	The reason for the ED episode (e.g., fall, burns, road traffic accident, etc.)
Contexual factors (C_i)		
Mode of arrival	Categorical (8)	The mode of transport used to get to the hospital (e.g., helicopter, private, 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 rate of hospital admissions to ED visits in the previous 12 months
Admissions per ED visit, last month	Continuous	The rate 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 (P_i)		
Admission errors	Continuous	The admission error propensity of the assigned physician, calculated as in Appendix C
Discharge errors	Continuous	The discharge error propensity of the assigned physician, calculated as in Appendix C
Admission errors x Discharge errors	Continuous	The interactions between the two variables defined above
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 (O_i)		
Length of stay in ED	Categorical(13)	30 minute time windows capturing how long the patient spent in the ED (e.g., 60-90 mins). Stays beyond 6 hours are merged into an 'over 360 minutes' category.
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 5.3

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.
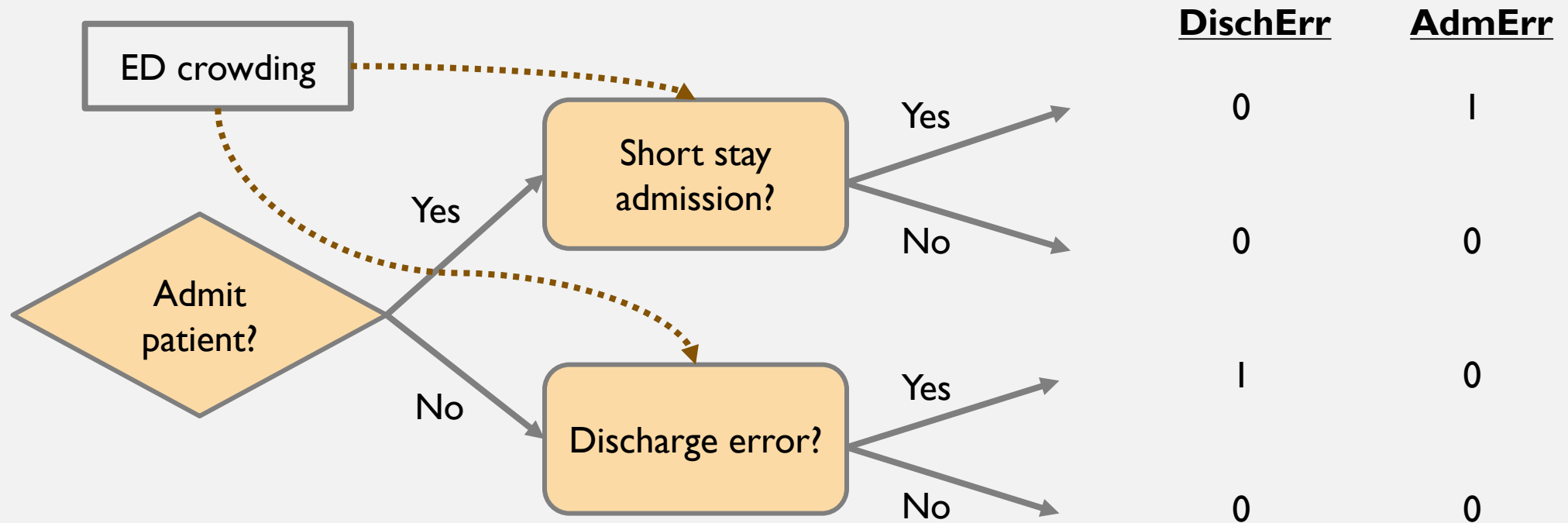
MODEL: STANDARD PROBIT

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

|	DischErr	AdmErr
Short stay admission? → Yes	0	1
Short stay admission? → No	0	0
Discharge error? → Yes	1	0
Discharge error? → No	0	0

MODEL: STANDARD PROBIT

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

	DischErr	AdmErr
Short stay admission? — Yes	0	1
Short stay admission? — No	0	0
Discharge error? — Yes	1	0
Discharge error? — No	0	0

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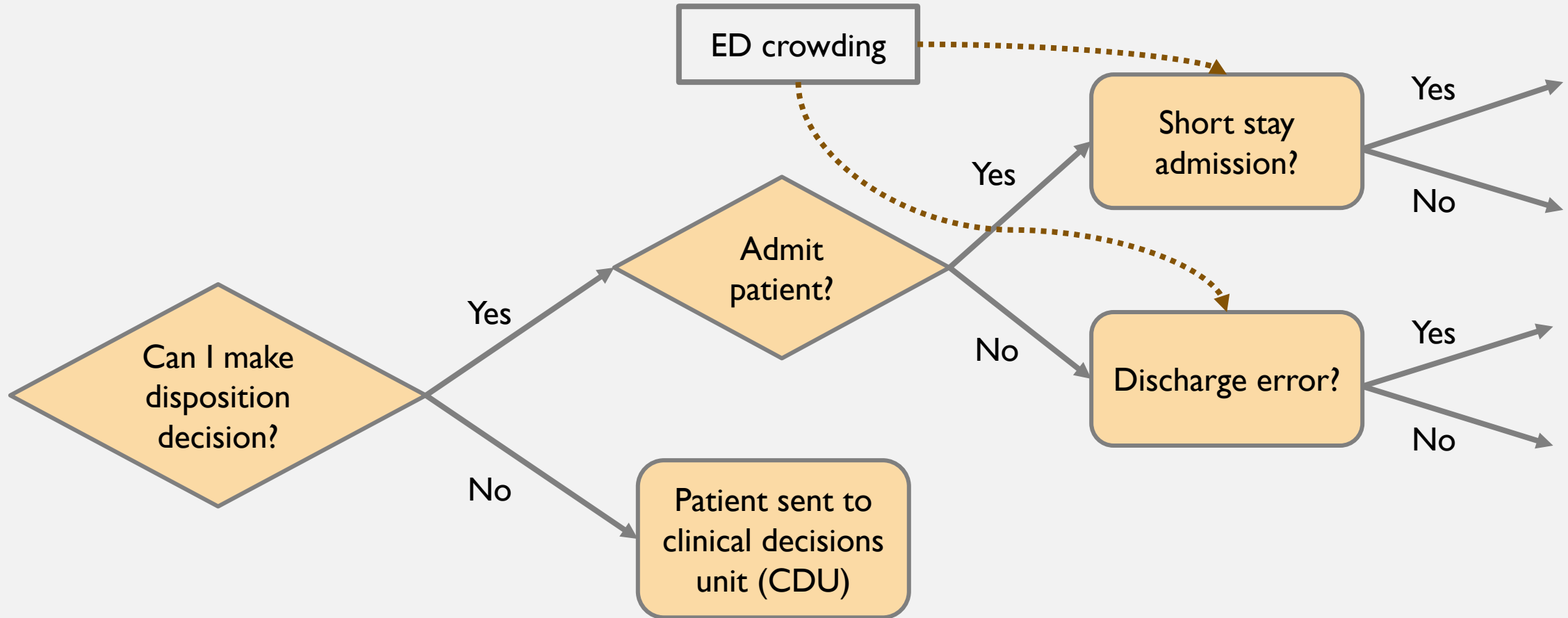$$y_i^{\text{outcome}} = 1(y_i^* > 0)$$

(where u_i is ED crowding)
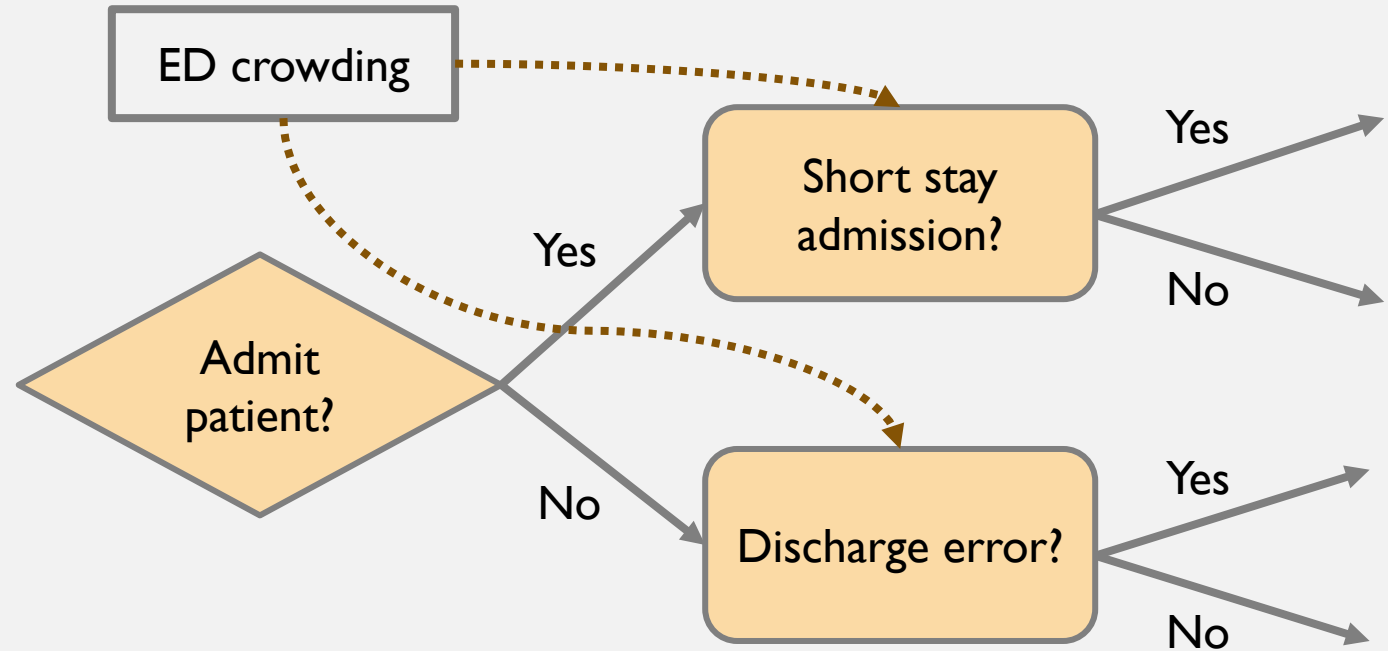
RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

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

(where $\textcolor{red}{u_i}$ is ED crowding)

- $y_i^{\text{select}} = 1$ if the patient is **not** referred to the CDU

- Censoring assumption: y_i^{outcome} (e.g. short-stay admission) 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 VARIABLE 1

Heckprob model estimation improved and coefficients more reliable when IVs are provided

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

- We include controls for the physician's historic admission and discharge error rates over the same period

- After controlling for the above, the CDU referral rate of the assigned physician is uncorrelated with the error

INSTRUMENTAL VARIABLE 2

Heckprob model estimation improved and coefficients more reliable when IVs are provided

IV 2: Busyness of 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 if one has several instruments → all tests are ok)
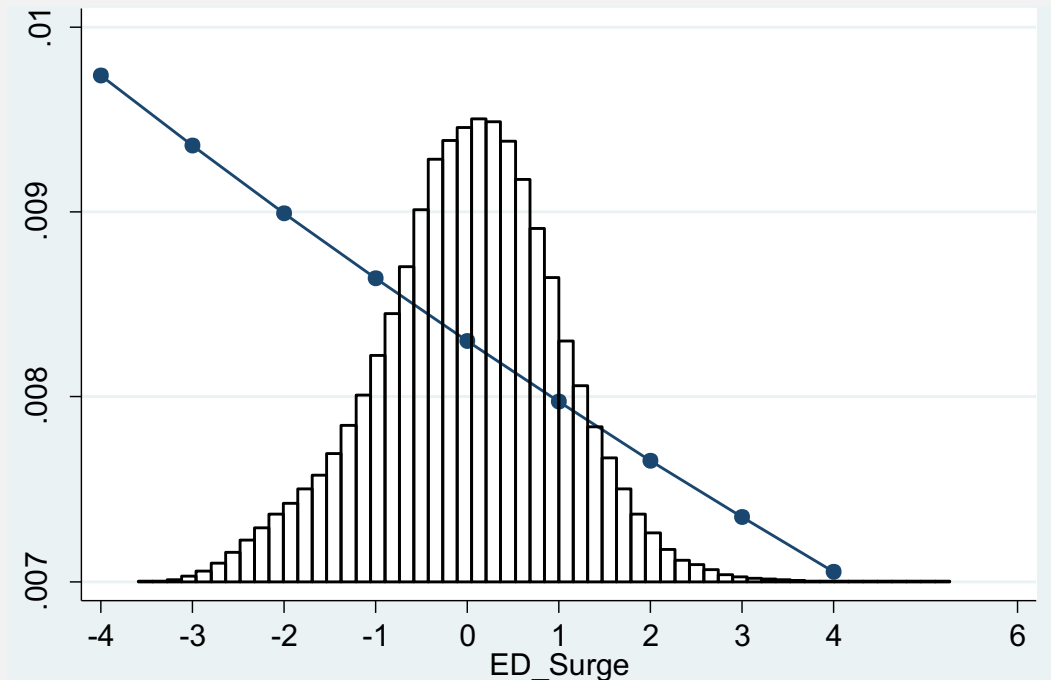
EFFECT OF TIME PRESSURE

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

- Time pressure (proxied by ED crowding) **increases** short-stay admissions
  - 0.23% increase per 1σ increase
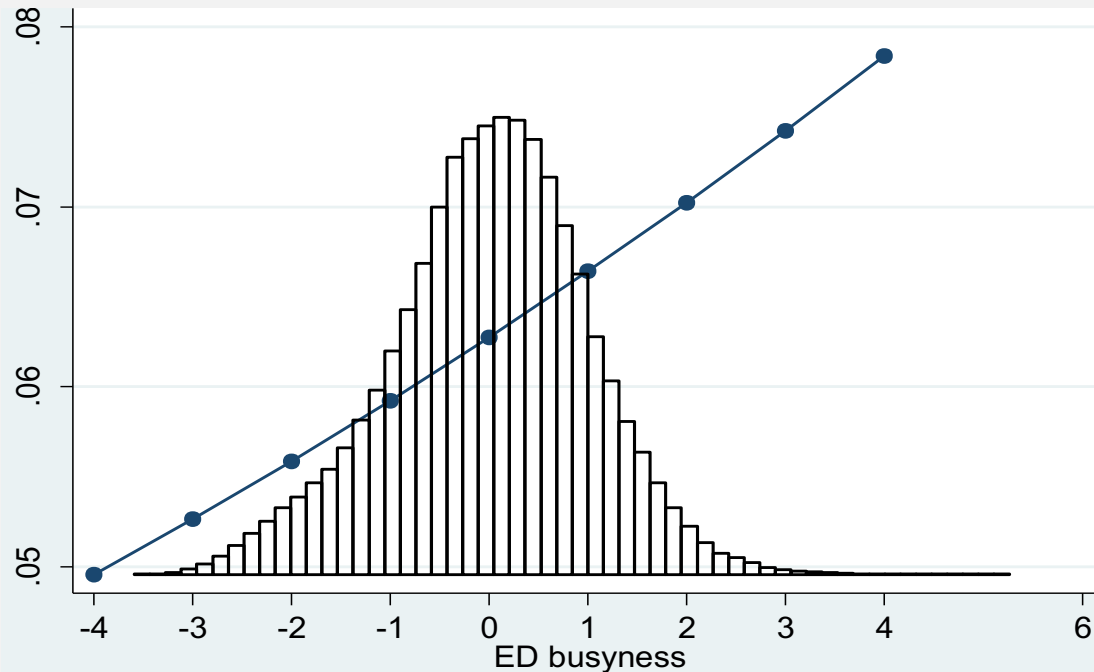
- Time pressure (proxied by ED crowding) **decreases** discharge errors
  - 0.04% reduction per 1σ increase

RQ-1: *How does increasing time pressure affect the accuracy of ED gatekeeping decisions?*

<u>Hypothesis 1-A</u> (error-making hypothesis)

- Admission errors: ↑

- Discharge errors: ↑

✗

<u>Hypothesis 1-B</u> (over-response hypothesis)

- Admission errors: ↑

- Discharge errors: = or ↓
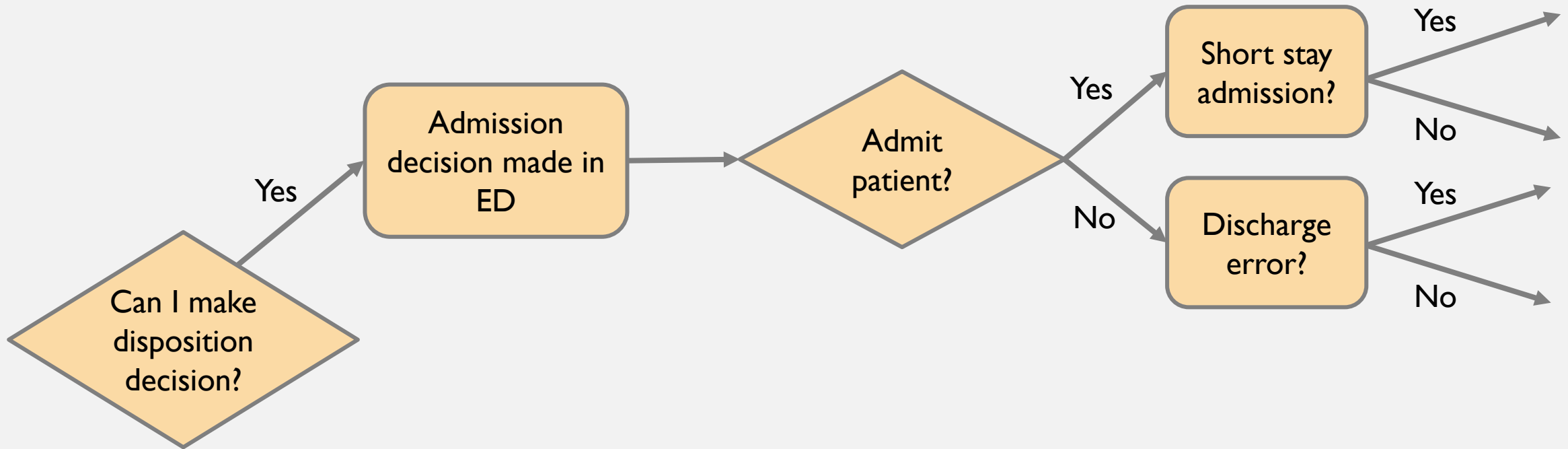
✓

Results in a bullwhip-type effect:

- Demand surges in the ED lead to relatively greater demand pressures in the hospital

Opposite of desirable system response to demand surges:

- Prefer physicians to *increase* the bar for hospital admission, as hospital congestion increases risk for the patient at hand & for the other patients in the hospital
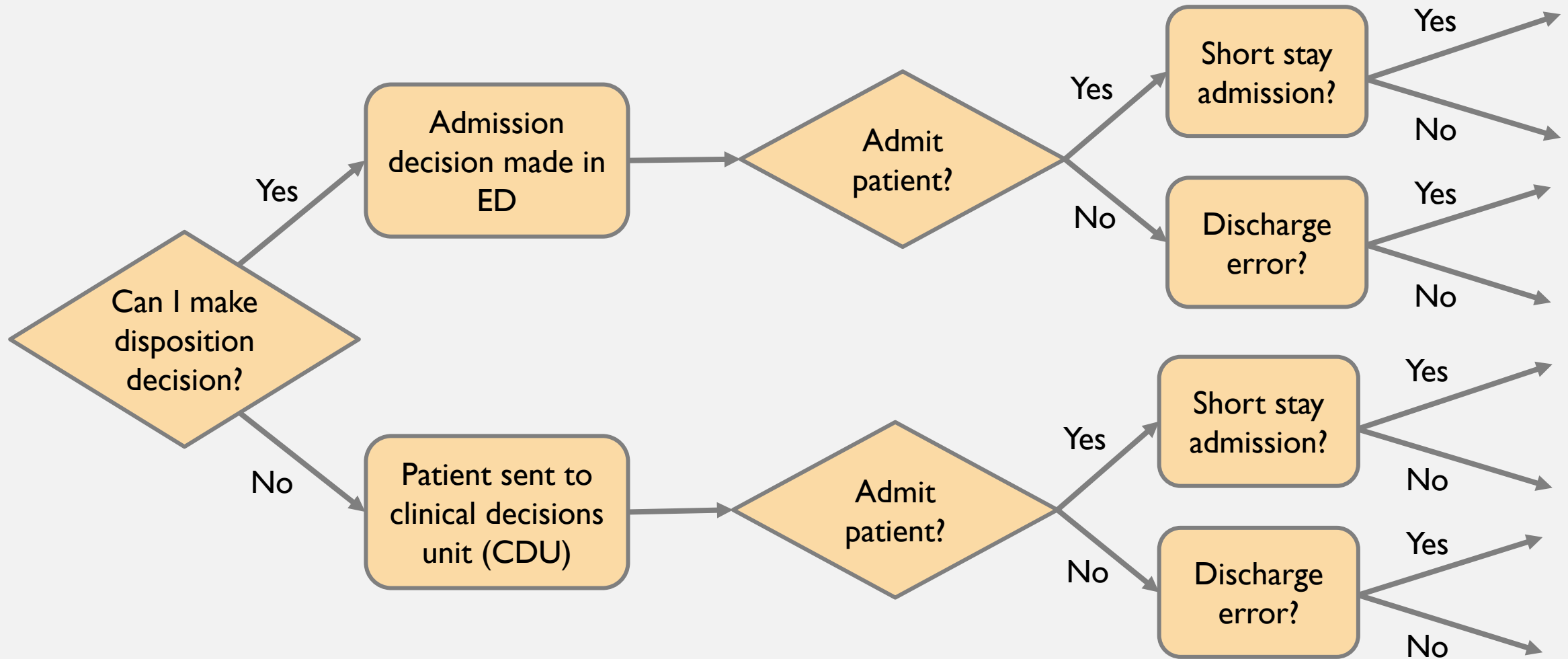
RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

TWO-STAGE GATEKEEPING SYSTEM

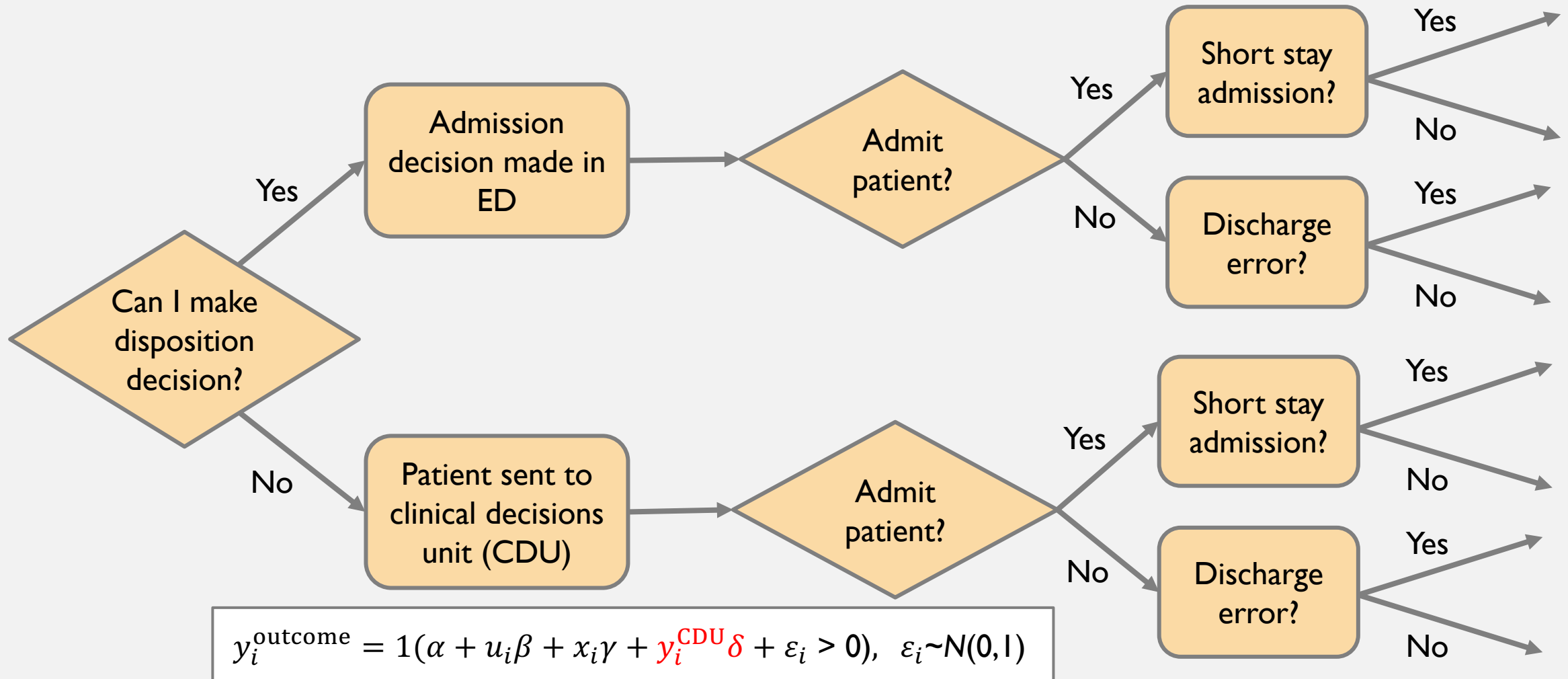**RQ-2:** *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*
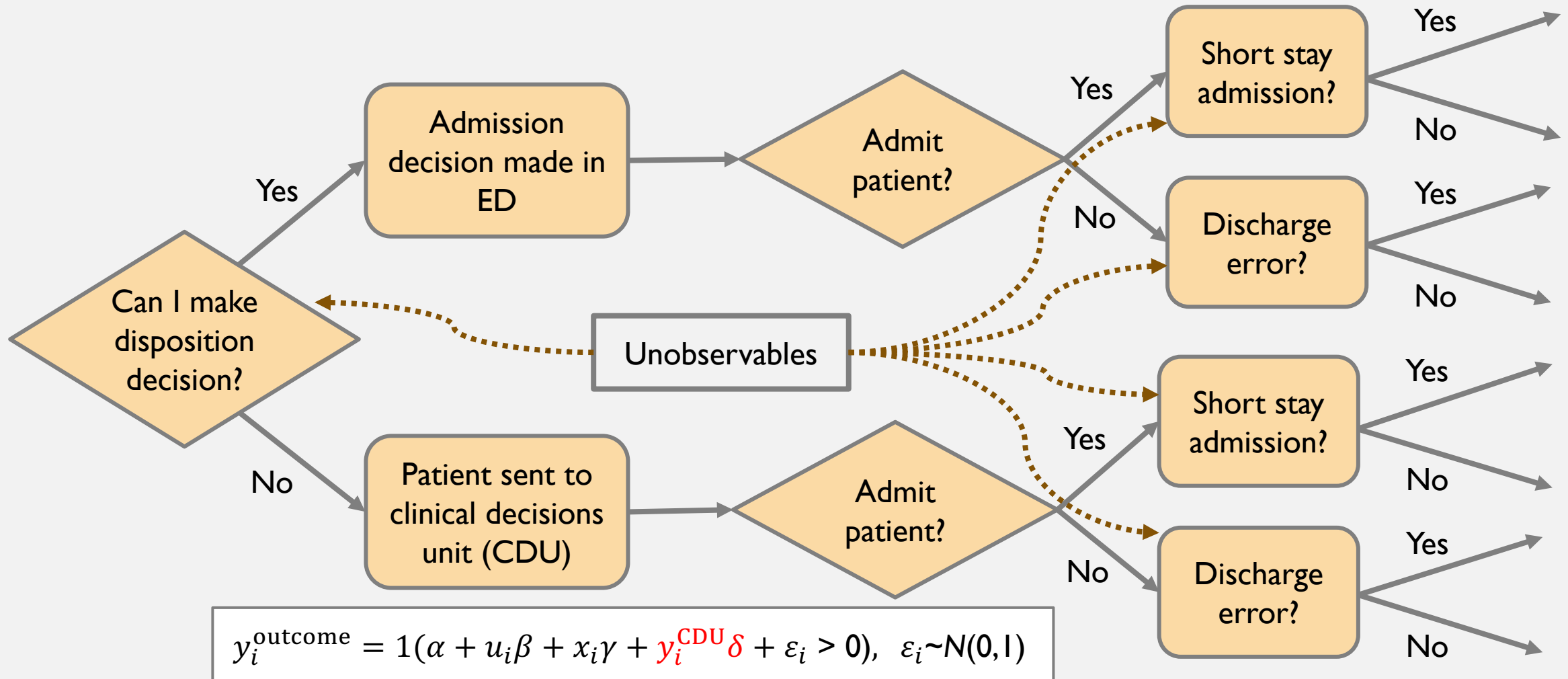
RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

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

RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

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

RECURSIVE BIVARIATE PROBIT MODEL

RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

Heckprobit

$$y_i^{\text{outcome}} = 1(\alpha + \textcolor{red}{u_i\beta} + x_i\gamma + \varepsilon_i > 0)$$
$$y_i^{\text{select}} = 1(\alpha' + u_i\beta' + x_i\gamma' + z_i\delta' + \varepsilon_i' > 0)$$

$$(\varepsilon_i, \varepsilon_i') \sim N\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

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

Coefficient of interest: β

Recursive bivariate probit

$$y_i^{\text{outcome}} = 1\big(\alpha + u_i\beta + x_i\gamma + \textcolor{red}{y_i^{\text{CDU}}\delta} + \varepsilon_i > 0\big)$$
$$y_i^{\text{CDU}} = 1(\alpha' + u_i\beta' + x_i\gamma' + z_i\delta' + \varepsilon_i' > 0)$$

$$(\varepsilon_i, \varepsilon_i') \sim N\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

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

Coefficient of interest: δ

BENEFITS OF TWO-STAGE GATEKEEPING

RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

	No patient routed through CDU	**All patients routed through CDU**
Short-stage admissions	5.23%	1.76%
Discharge errors	1.16%	0.67%

Average treatment effect (ATE):

- Short-stay admissions = -3.5%

- Discharge errors = -0.49%

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

BENEFITS OF TWO-STAGE GATEKEEPING

RQ-2: *Can streaming patients through two-stage gatekeeping system improve disposition accuracy?*

	No patient routed through CDU	**All patients routed through CDU**
Short-stage admissions	5.23%	1.76%
Discharge errors	1.16%	0.67%

Average treatment effect (ATE):

- Short-stay admissions = -3.5%

- Discharge errors = -0.49%

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

Average treatment effect on the treated (ATT):

- Short-stay admissions = -9.3%

- Discharge errors = -1.2%

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

MORE ED OR MORE CDU CAPACITY?

- **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 short-stay admissions (and slight increase in discharge errors)

- Keeping ED at its observed capacity and retain CDU:
  - leads to 1% reduction of short-stay admissions ⬅ → Retaining CDU capacity ~7x more effective than adding ED capacity
- 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

- Gatekeepers frequently must take referral decisions under 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 <u>and</u> reduce the discrepancy between the gatekeeper and the system optimal response

Thank you!

Questions?