



# Day 2: Session III

## Considerations in Comparative and Public Reporting

---

Presenters: Patrick Romano, UC Davis  
Shoshanna Sofaer, Baruch College  
AHRQ QI User Meeting  
September 26-27, 2005





# Selecting AHRQ Quality Indicators for public reporting and pay-for-performance

---

- Type or conceptual framework
- Face validity or salience to providers
- Impact or opportunity for improvement
- Reliability or precision
- Coding (criterion) validity
- Construct validity
- Susceptibility to bias



# Types of provider-level quality indicators

- **Structure: the conditions under which care is provided**
  - ✓ Volume (AAA repair, CEA, CABG, PCI, esophageal or pancreatic resection, pediatric heart surgery)
- **Process: the activities that constitute health care**
  - ✓ Use of desirable/undesirable procedures (C/S, VBAC, bilateral cardiac cath, incidental appendectomy, laparoscopic cholecystectomy)
- **Outcome: changes attributable to health care**
  - ✓ Risk-adjusted mortality (AMI, CHF, GI hemorrhage, hip fracture, pneumonia, stroke, AAA repair, CABG, craniotomy, esophageal resection, pancreatic resection, THA, pediatric heart surgery)
  - ✓ Risk-adjusted complications or “potential safety-related events” (Patient Safety Indicators)

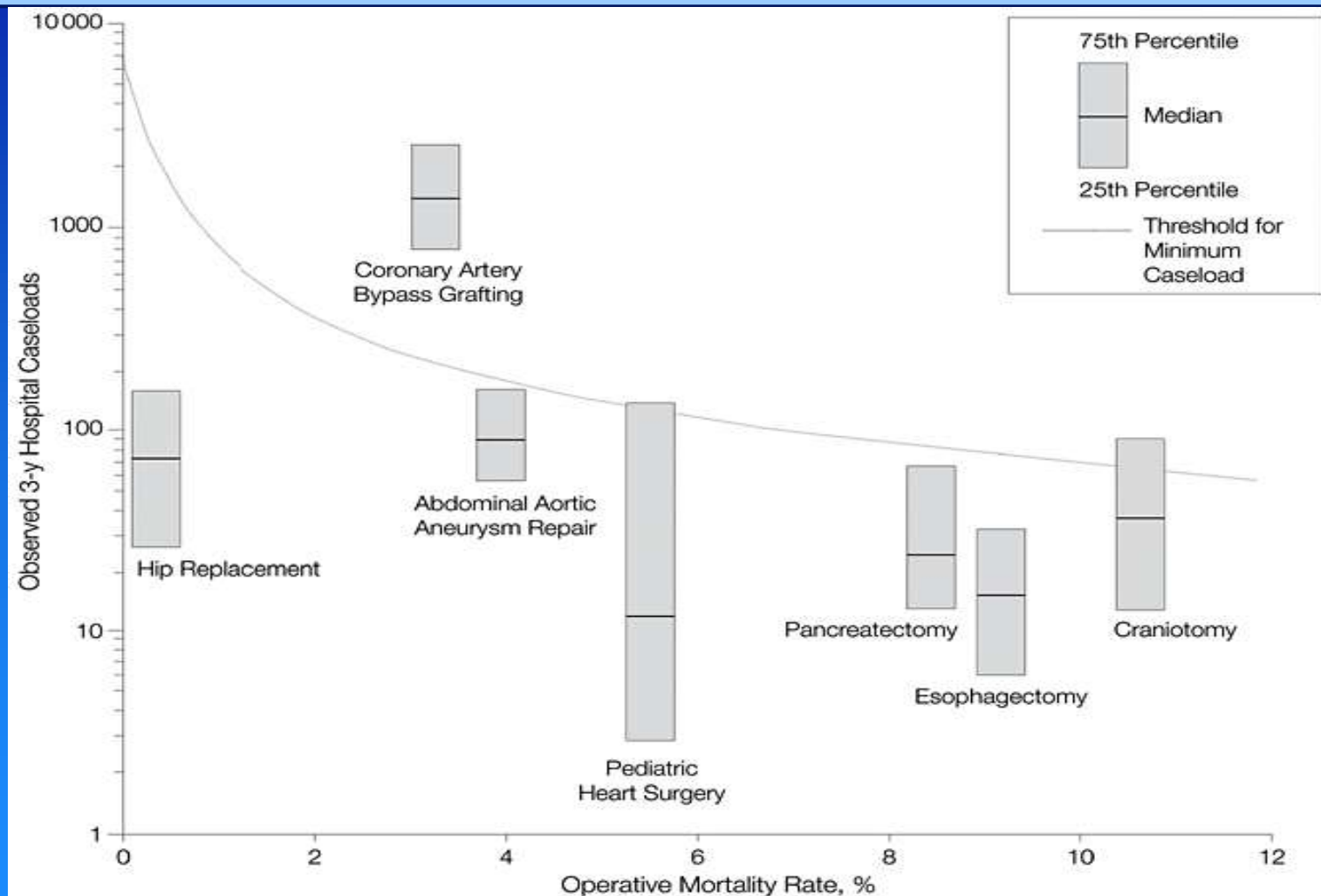


# Key features of structural measures

- Enabling factors that make it easier (harder) for professionals to provide high-quality care (i.e., facilitators or markers)
- Weakly associated with process/outcome measures
- Easy to measure, but hard to modify
- Few intervention studies, causal relationships unclear – do better structures lead to different processes, or do better processes lead to different structures?
- Use structural indicators when acceptable process or outcome measures are not available (“free ride” problem)
- Focus on modifiable structures OR settings in which hospitals that cannot modify structures are allowed to close (excess capacity)



# Minimum hospital volume needed to detect doubling of mortality rate ( $\alpha=0.05$ , $\beta=0.2$ )



Ref: Dimick, et al. *JAMA*. 2004;292:847-851.







# Impact: Estimated lives saved by implementing hospital volume standards (NIS)

Birkmeyer et al., Surgery 2001;130:415-22

<b>Volume indicator</b>	<b>RR mortality LVH vs HVH</b>	<b>Patients at LVHs in MSAs</b>	<b>Potential lives saved by volume standards</b>
CABG	1.38	164,261	1,486
Coronary angioplasty/PCI	1.33	121,292	345
AAA repair	1.60	18,534	464
Carotid endarterectomy	1.28	82,544	118
Esophagectomy	3.01	1,696	168



# Key features of process measures

---

- Directly actionable by health care providers (“opportunities for intervention”)
- Highly responsive to change
- Validated – or potentially “validatable” – in randomized trials (but NOT the AHRQ QIs)
- Illustrate the pathways by which interventions may lead to better patient outcomes
- Focus on modifiable processes that are salient to providers, and for which there is clear opportunity for improvement



# Key features of outcome measures

- What really matters to patients, families, communities
- Intrinsically meaningful and easy to understand
- Reflect not just what was done but how well it was done (difficult to measure directly)
- Morbidity measures tend to be reported inconsistently (due to poor MD documentation and/or coding)
- Outcome measures may be confounded by variation in observation units, discharge/transfer practices, LOS, severity of illness
- Many outcomes of interest are rare or delayed
- Are outcomes sufficiently under providers' control?
- Focus on outcomes that are conceptually and empirically attributable to providers (e.g., process linkages), and for which established benchmarks demonstrate opportunity for improvement.





# AHRQ QI development: General process

---

- Literature review (all)
  - To identify quality concepts and potential indicators
  - To find previous work on indicator validity
- ICD-9-CM coding review (all)
  - To ensure correspondence between clinical concept and coding practice
- Clinical panel reviews (PSI's, pediatric QIs)
  - To refine indicator definition and risk groupings
  - To establish face validity when minimal literature
- Empirical analyses (all)
  - To explore alternative definitions
  - To assess nationwide rates, hospital variation, relationships among indicators
  - To develop methods to account for differences in risk





# AHRQ QI development: References

- **AHRQ Quality Indicator documentation web page at**  
**<http://www.qualityindicators.ahrq.gov/downloads.htm>**
  - *Refinement of the HCUP Quality Indicators (Technical Review)*, May 2001
  - *Measures of Patient Safety Based on Hospital Administrative Data - The Patient Safety Indicators*, August 2002
- **Peer-reviewed literature (examples):**
  - *AHRQ's Advances in Patient Safety: From Research to Implementation* (4-volume compendium)
  - Romano, et al. *Health Aff (Millwood)*. 2003; 22(2):154-66.
  - Zhan and Miller. *JAMA*. 2003; 290(14):1868-74.
  - Sedman, et al. *Pediatrics*. 2005; 115(1):135-45.
  - Rosen et al., *Med Care*. 2005; 43(9):873-84.





# Face validity: Clinical panel review

- Intended to establish consensual validity
- Modified RAND/UCLA Appropriateness Method
- Physicians of various specialties/subspecialties, nurses, other specialized professionals (e.g., midwife, pharmacist)
- Potential indicators were rated by 8 multispecialty panels; surgical indicators were also rated by 3 surgical panels
- All panelists rated all assigned indicators (1-9) on:
  - Overall usefulness
  - Likelihood of identifying the occurrence of an adverse event or complication (i.e., not present at admission)
  - Likelihood of being preventable (i.e., not an expected result of underlying conditions)
  - Likelihood of being due to medical error or negligence (i.e., not just lack of ideal or perfect care)
  - Likelihood of being clearly charted
  - Extent to which indicator is subject to case mix bias



# Evaluation framework for PSIs

Medical error and complications continuum



- Pre-conference ratings and comments/suggestions
- Individual ratings returned to panelists with distribution of ratings and other panelists' comments/suggestions
- Telephone conference call moderated by PI, with note-taker, focusing on high-variability items and panelists' suggestions (90-120 mins)
- Suggestions adopted only by consensus
- Post-conference ratings and comments/suggestions



# Example reviews of PSIs

## Multispecialty panels

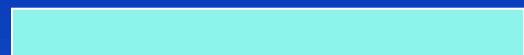
### Postop Pneumonia

### Decubitus Ulcer

Overall rating



(5)

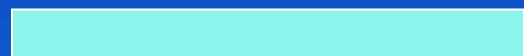


(8)

Not present on admission



(7)

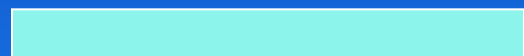


(8)

Preventability



(4)



(8)

Due to medical error



(2)



(8)

Charting by physicians



(6)



(7)

Not biased by case mix



(3)



(7)





# Final selection of PSIs

- Retained indicators for which “overall usefulness” rating was “Acceptable” or “Acceptable-”
  - Median score 7-9; AND
  - Definite agreement (“acceptable”) if no more than 1 or 2 panelists rated indicator below 7
  - Indeterminate agreement (“acceptable-”) if no more than 1 or 2 panelists rated indicator in 1-3 range
- 48 indicators reviewed (15 by 2 separate panels)
- 20 “accepted” based on face validity
  - 2 dropped due to operational concerns
- 17 “experimental” or promising indicators
- 11 rejected



# Panel ratings of PSI “preventability”

Acceptable	Acceptable (-)	Unclear	Unclear (-)
Decubitus ulcer	Complications of anesthesia	Death in low mortality DRG	Failure to rescue
Foreign body left in	Selected infections due to medical care	Postop hemorrhage/hematoma	Postop physiologic/metabolic derangement
Iatrogenic pneumothorax <sup>a</sup>	Postop PE or DVT <sup>b</sup>	Postop respiratory failure	
Postop hip fracture <sup>a</sup>	Transfusion reaction	Postop abdominopelvic wound dehiscence	
Technical difficulty with procedure	Birth trauma	Postop sepsis	
Obstetric trauma (all delivery types)			

<sup>a</sup> Panel ratings were based on definitions different than final definitions. For “Iatrogenic pneumothorax,” the rated denominator was restricted to patients receiving thoracentesis or central lines; the final definition expands the denominator to all patients (with same exclusions). For “In-hospital fracture” panelists rated the broader Experimental indicator, which was replaced in the Accepted set by “Postoperative hip fracture” due to operational concerns.

<sup>b</sup> Vascular complications were rated as Unclear (-) by surgical panel; multispecialty panel rating is shown here.





# International expert panel ratings of PSIs

## Organization for Economic Cooperation and Development

PSIs recommended	PSIs not recommended	Experimental or rejected PSIs recommended
Selected infections due to medical care	Death in low mortality DRG	Postop wound infection
Decubitus ulcer	Postop hemorrhage/ hematoma	In-hospital hip fracture or fall
Complications of anesthesia	Iatrogenic pneumothorax	
Postop PE or DVT	Postop abdominopelvic wound dehiscence	
Postop sepsis	Failure to rescue	
Technical difficulty with procedure	Postop physiologic/ metabolic derangement	
Transfusion reaction	Postop respiratory failure	
Foreign body left in		
Postop hip fracture		
Birth trauma		
Obstetric trauma (all delivery types)		



# Impact: Estimated cases in 2000 (NIS)

Romano et al., Health Aff 2003;22(2):154-66



Indicator	Frequency±95% CI	Rate/100
Postoperative septicemia	14,055 ± 1060	1.091
Postoperative thromboembolism	75,811 ± 4,156	0.919
Postoperative respiratory failure	12,842 ± 938	0.359
Postoperative physiologic or metabolic derangement	4,003 ± 419	0.089
Decubitus ulcer	201,459 ± 10,104	2.130
Selected infections due to medical care	54,490 ± 2,658	0.193
Postoperative hip fracture	5,207 ± 327	0.080
Accidental puncture or laceration	89,348 ± 5,669	0.324
Iatrogenic pneumothorax	19,397 ± 1,025	0.067
Postoperative hemorrhage/hematoma	17,014 ± 968	0.206



# Estimating the impact of preventing each PSI event on mortality, LOS, charges (ROI)

NIS 2000 analysis by Zhan & Miller, JAMA 2003;290:1868-74

Indicator	$\Delta$ Mort (%)	$\Delta$ LOS (d)	$\Delta$ Charge (\$)
Postoperative septicemia	21.9	10.9	\$57,700
Postoperative thromboembolism	6.6	5.4	21,700
Postoperative respiratory failure	21.8	9.1	53,500
Postoperative physiologic or metabolic derangement	19.8	8.9	54,800
Decubitus ulcer	7.2	4.0	10,800
Selected infections due to medical care	4.3	9.6	38,700
Postoperative hip fracture	4.5	5.2	13,400
Accidental puncture or laceration	2.2	1.3	8,300
Iatrogenic pneumothorax	7.0	4.4	17,300
Postoperative hemorrhage/hematoma	3.0	3.9	21,400





# Estimating the impact of preventing each PSI event on mortality, LOS, charges (ROI)

VA PTF analysis by Rosen et al., Med Care 2005;43:873-84

Indicator	$\Delta$ Mort (%)	$\Delta$ LOS (d)	$\Delta$ Charge (\$)
Postoperative septicemia	35.7	18	\$39,531
Postoperative thromboembolism	10.2	7	12,856
Postoperative respiratory failure	29.3	19	39,848
Postoperative physiologic or metabolic derangement	44.5	15	37,460
Decubitus ulcer	10.9	5	5,887
Selected infections due to medical care	9.8	11	18,706
Postoperative hip fracture	17.9	10	18,906
Accidental puncture or laceration	3.9	3	11,626
Iatrogenic pneumothorax	10.1	5	8,039
Postoperative hemorrhage/hematoma	8.1	6	14,384



# Impact: Estimated cases in 2000 (NIS)

Romano et al., Health Aff 2003;22(2):154-66

Indicator	Frequency±95% CI	Rate per 100
Birth trauma	27,035 ± 5,674	0.667
Obstetric trauma –cesarean	5,523 ± 597	0.593
Obstetric trauma - vaginal w/out instrumentation	249,243 ± 12,570	8.659
Obstetric trauma - vaginal w instrumentation	60,622 ± 3,104	24.408
Postoperative abdominopelvic wound dehiscence	3,858 ± 289	0.193
Transfusion reaction	138 ± 49	0.0004
Complications of anesthesia	5,305 ± 455	0.056
Foreign body left during procedure	2,710 ± 204	0.008



# Impact of patient safety events in 2000

Zhan & Miller, JAMA 2003; replicated by Rosen et al., 2005

Indicator	$\Delta$ Mort (%)	$\Delta$ LOS (d)	$\Delta$ Charge (\$)
Birth trauma	-0.1 (NS)	-0.1 (NS)	300 (NS)
Obstetric trauma –cesarean	-0.0 (NS)	0.4	2,700
Obstetric trauma - vaginal w/out instrumentation	0.0 (NS)	0.05	-100 (NS)
Obstetric trauma - vaginal w instrumentation	0.0 (NS)	0.07	220
Postoperative abdominopelvic wound dehiscence	9.6	9.4	40,300
Transfusion reaction*	-1.0 (NS)	3.4 (NS)	18,900 (NS)
Complications of anesthesia*	0.2 (NS)	0.2 (NS)	1,600
Foreign body left during procedure†	2.1	2.1	13,300

\* All differences NS for transfusion reaction and complications of anesthesia in VA/PTF.

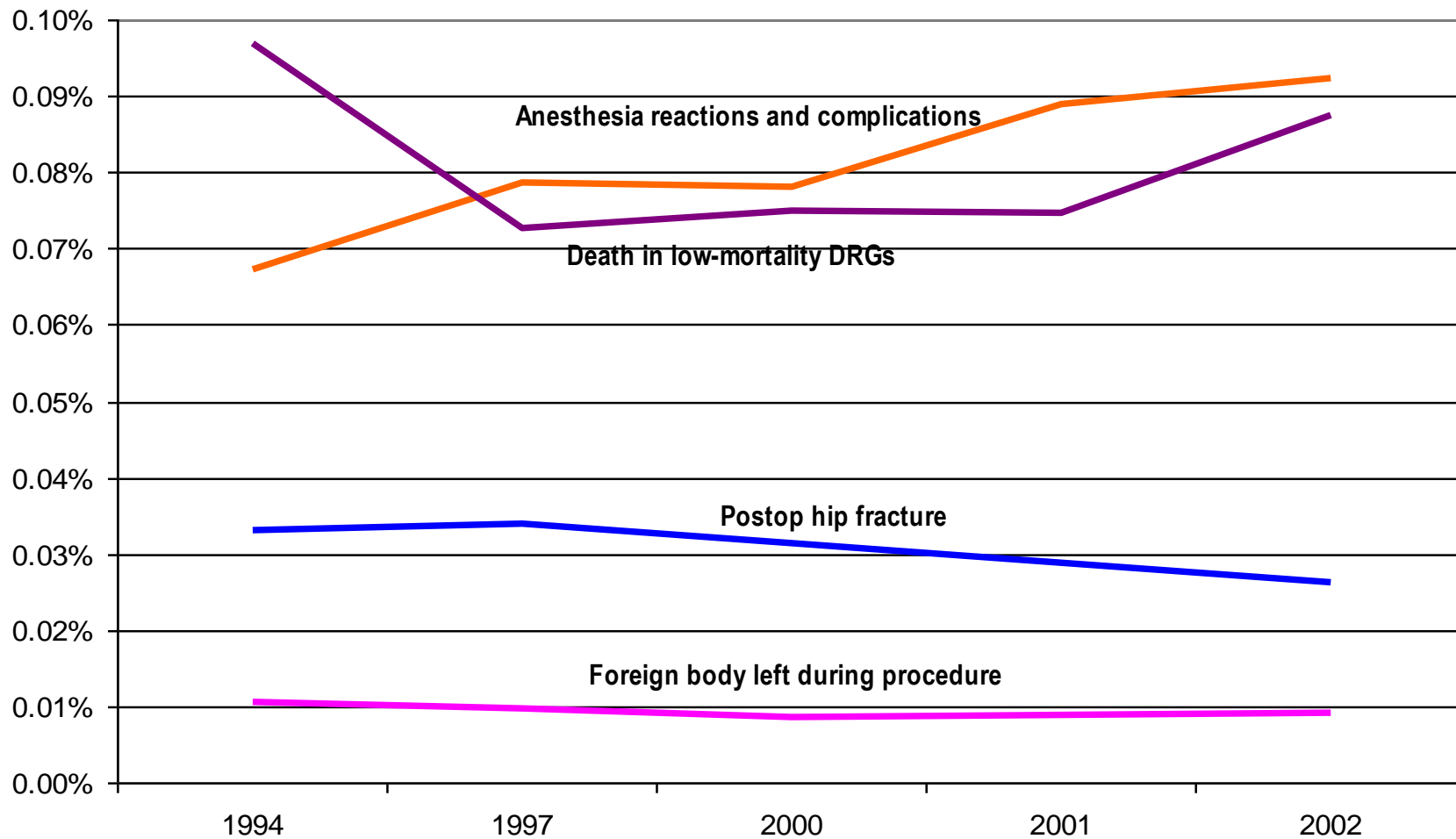
† Mortality difference NS for foreign body in VA/PTF.





# National trends in PSI rates, 1994-2002

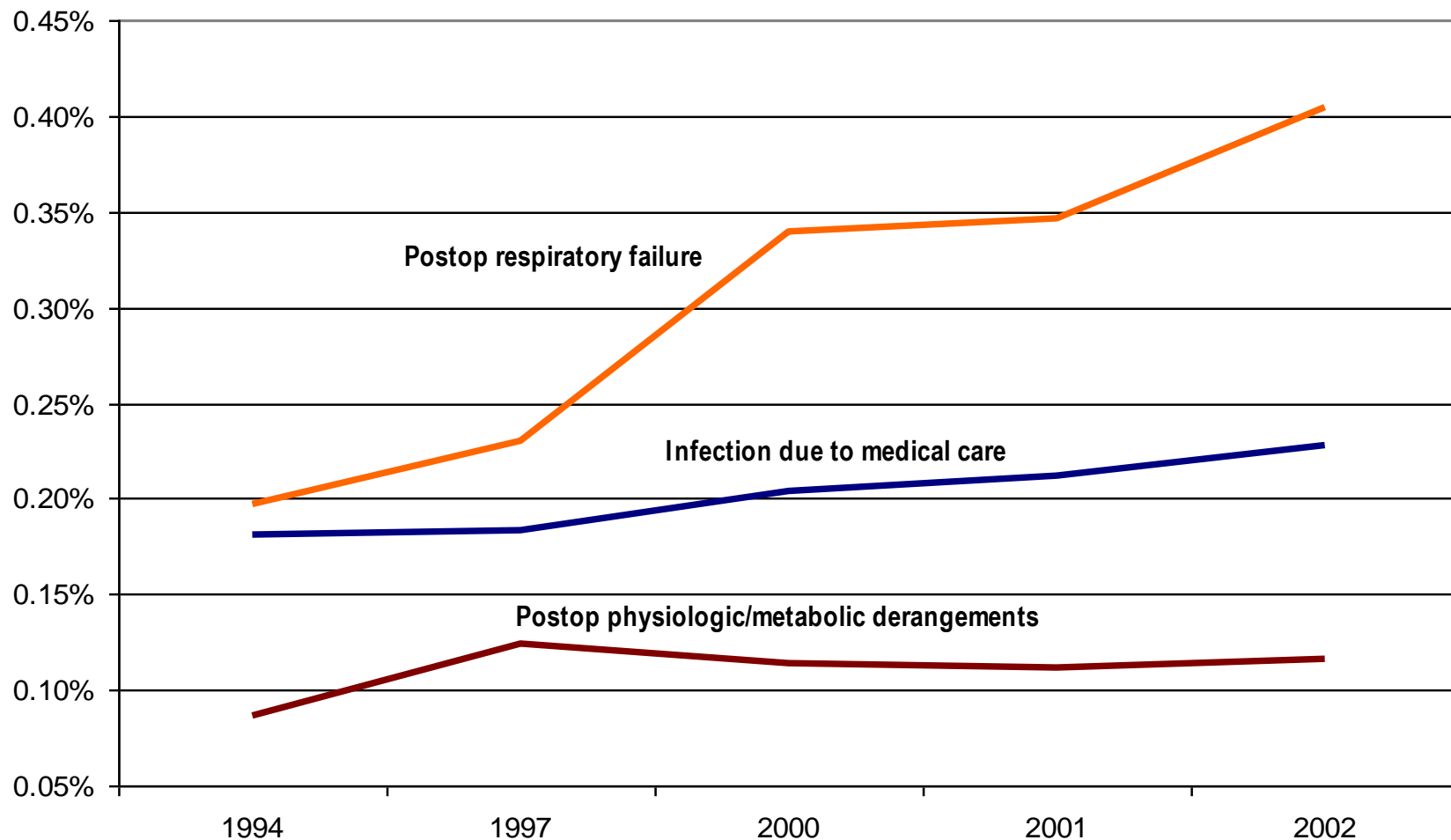
## Rare events (<0.1%)





# National trends in PSI rates, 1994-2002

## Low-frequency medical complications (0.05-0.5%)

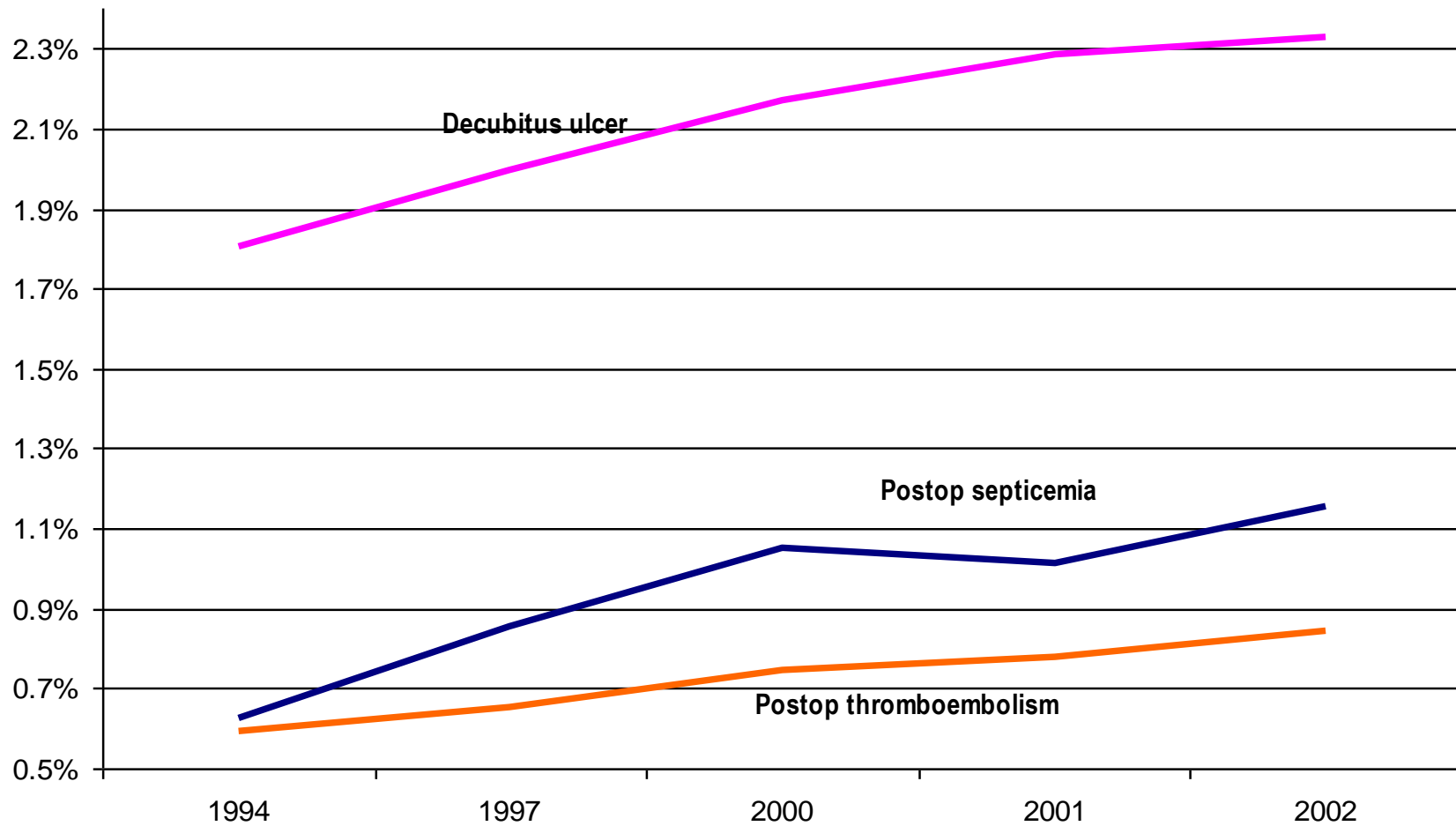






# National trends in PSI rates, 1994-2002

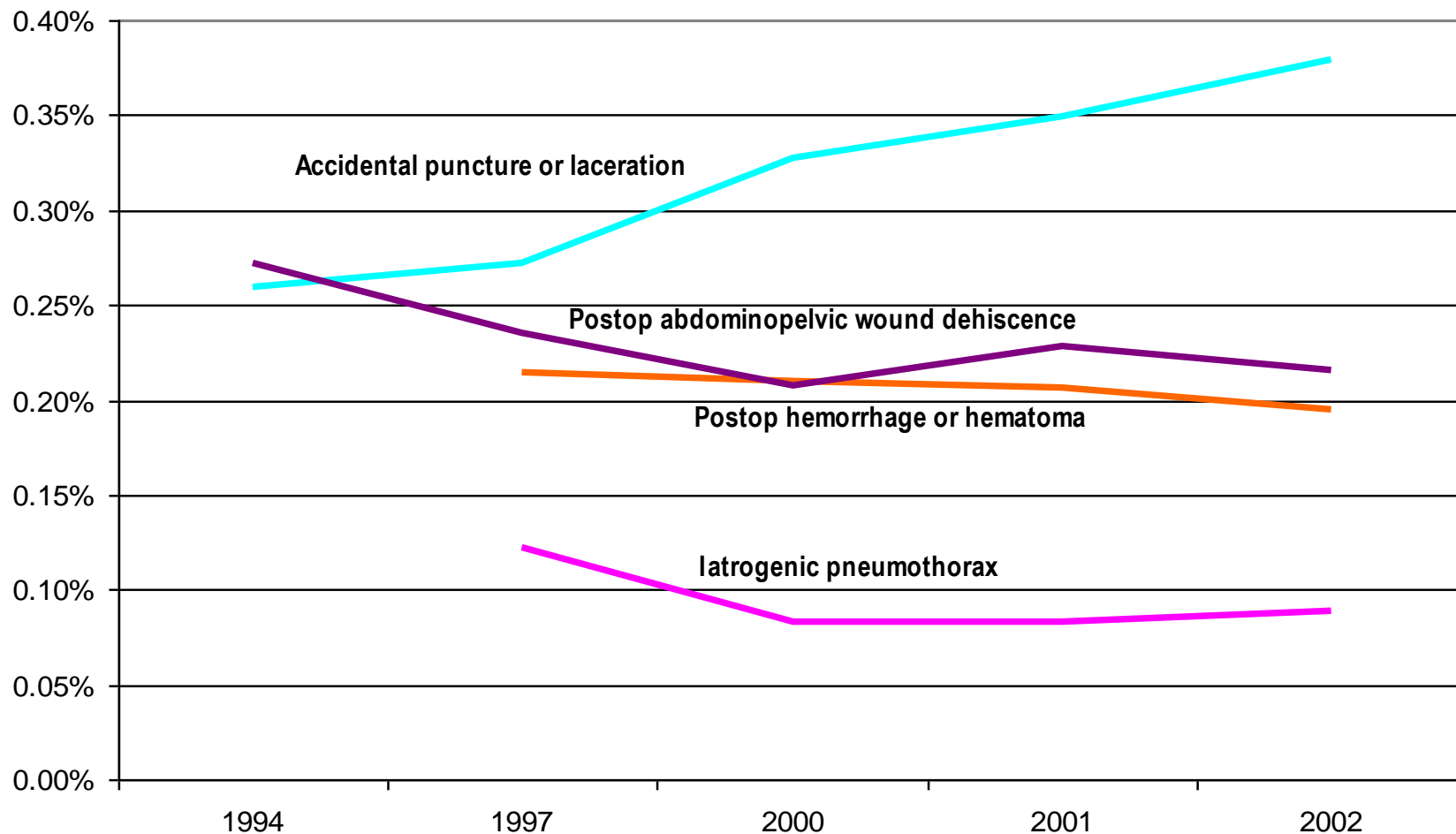
## High-frequency medical complications (0.5-2.5%)





# National trends in PSI rates, 1994-2002

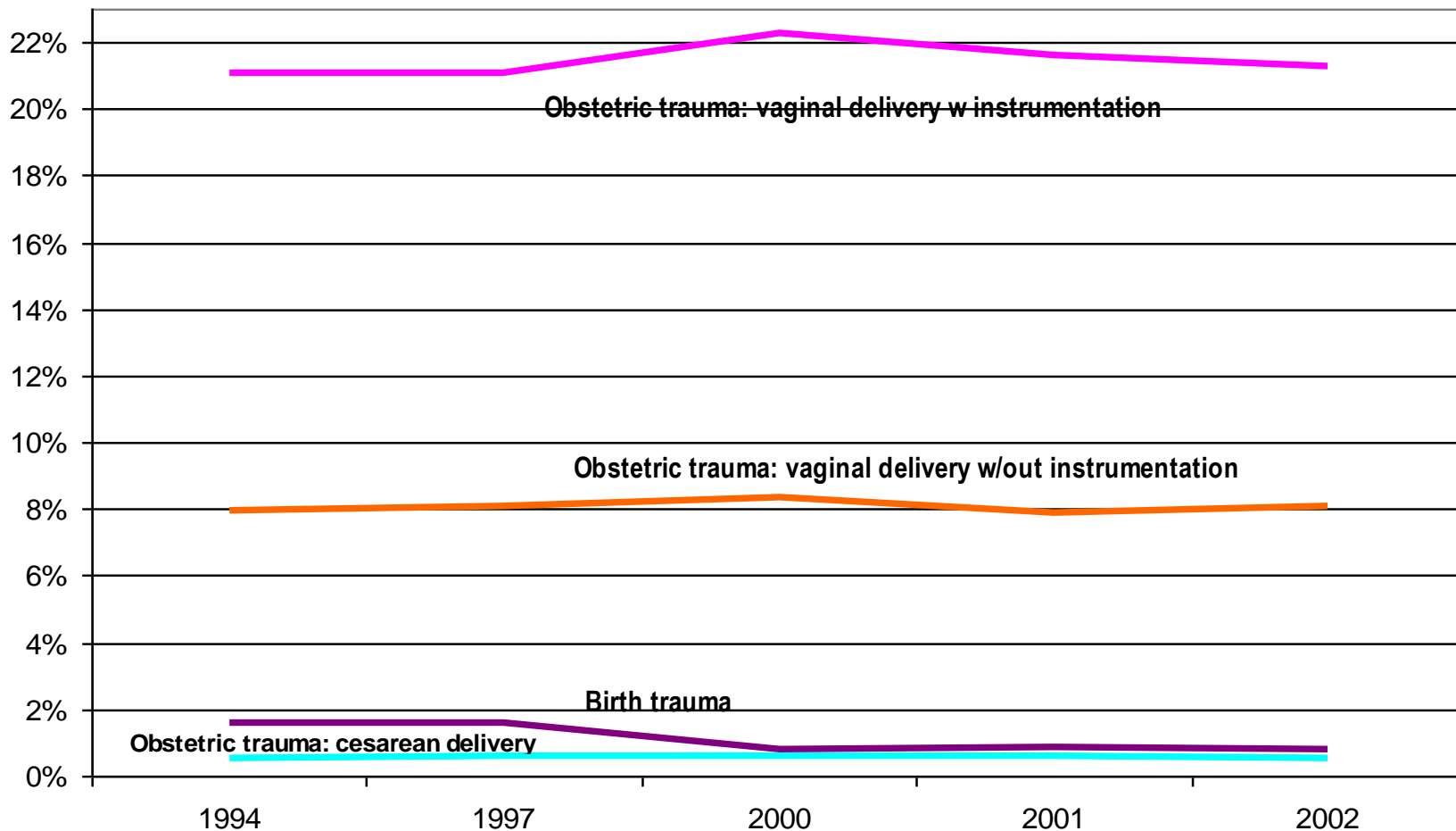
## Surgical/technical complications





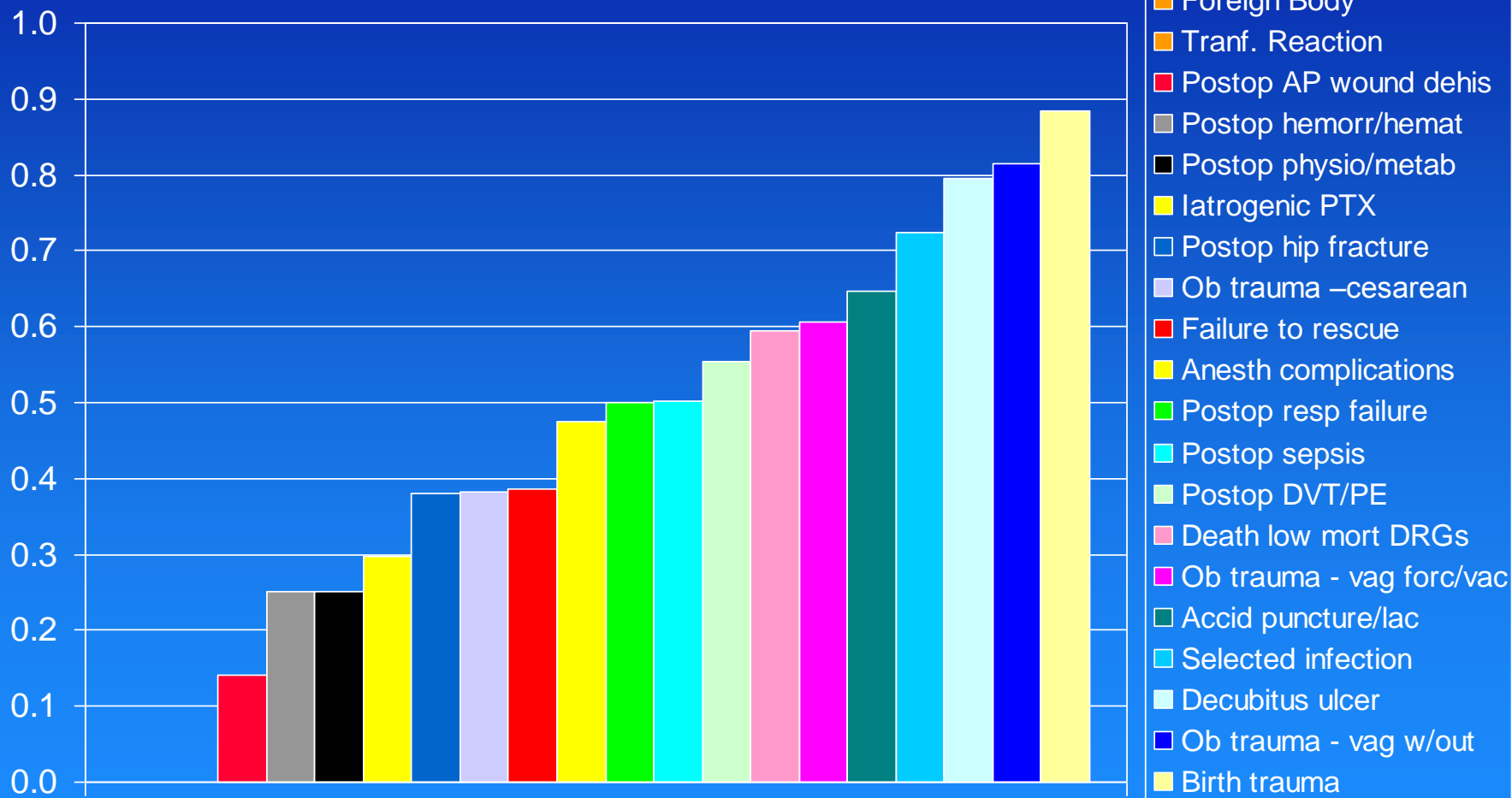
# National trends in PSI rates, 1994-2002

## Obstetric complications





# Reliability or precision: signal ratio

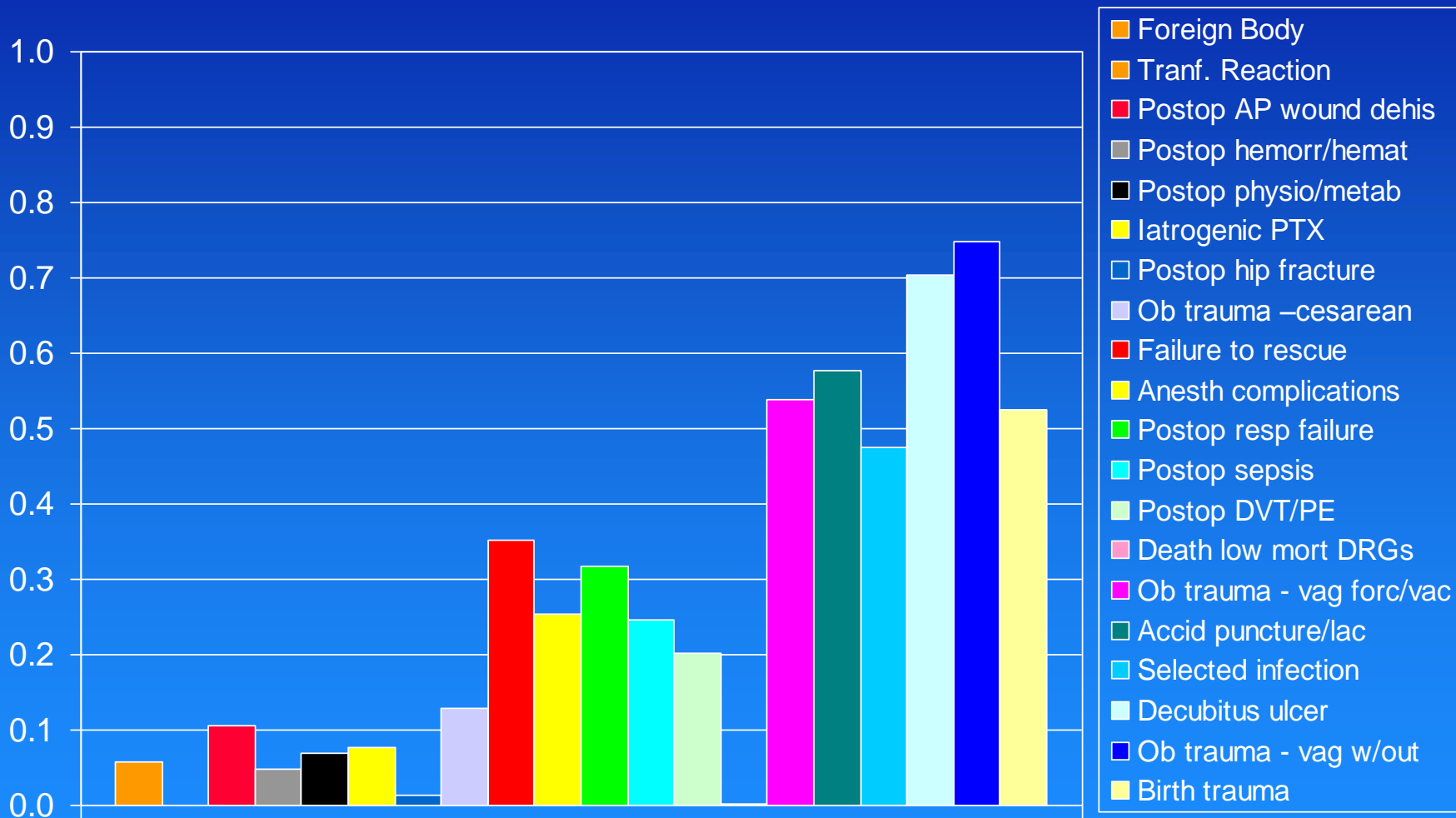


Source: 2002 State Inpatient Data. Average Signal Ratio across all hospitals (N=4,428)





# Year-to-year correlation of hospital effects



Source: 2001-2002 State Inpatient Data, hospitals with at least 1,000 discharges (N=4,428). Risk-adjusted unsmoothed rates.







# Coding (criterion) validity based on literature review (MEDLINE/EMBASE)

---

- Validation studies of Iezzoni et al.'s CSP
  - At least one of three validation studies (coders, nurses, or physicians) confirmed PPV at least 75% among flagged cases
  - Nurse-identified process-of-care failures were more prevalent among flagged cases than among unflagged controls
- Other studies of coding validity
  - Very few in peer-reviewed journals, some in “gray literature”



# Validation (%) of Complications Screening Program

Med Care 2000;38:785-806,868-76; Int J Qual Health Care 1999;11:107-18

CSP Indicator	PSI	Coder: Complic Present	RN: process problem	MD: Complic present	MD: Quality problem
Postprocedural hemorrhage/hematoma	#9 narrower: requires proc code + dx	83 (surg) 49 (med)	66 vs 46 13 vs 5	57 (surg) 55 (med)	37 vs 2 31 vs 2
Postop pulmonary compromise	#11 narrower: includes only resp failure	72	52 vs 46	75	20 vs 2
DVT/PE	#12 surgical only Slight changes	59 (surg) 32 (med)	72 vs 46 69 vs 5	70 (surg) 28 (med)	50 vs 2 20 vs 2
In-hosp hip frx and falls	#8 surgical only, no E codes	57 (surg) 11 (med)	76 vs 46 54 vs 5	71 (surg) 11 (med)	24 vs 2 5 vs 2



# Criterion validity of PSIs linked to NSQIP, VA hospitals

Tsilimingras, Romano, et al., AcademyHealth 2005

Indicator	Sensitivity		PPV	
	Current Inpatient	Better Inpatient	Current Inpatient	Better Inpatient
Postop sepsis	32%	37%	44%	45%
Postop thromboembolism	56%	58%	22%	22%
Postop respiratory failure	19%	67%	74%	66%
Postop physiologic/ metabolic derangement	44%	48%	54%	63%
Postop abdominopelvic wound dehiscence	29%	61%	72%	57%



# Construct validity based on literature review (MEDLINE/EMBASE)

---

- Approaches to assessing construct validity
  - Is the outcome indicator associated with explicit processes of care (e.g., appropriate use of medications)?
  - Is the outcome indicator associated with implicit process of care (e.g., global ratings of quality)?
  - Is the outcome indicator associated with nurse staffing or skill mix, physician skill mix, or other aspects of hospital structure?

# Summary of construct validity evidence in literature

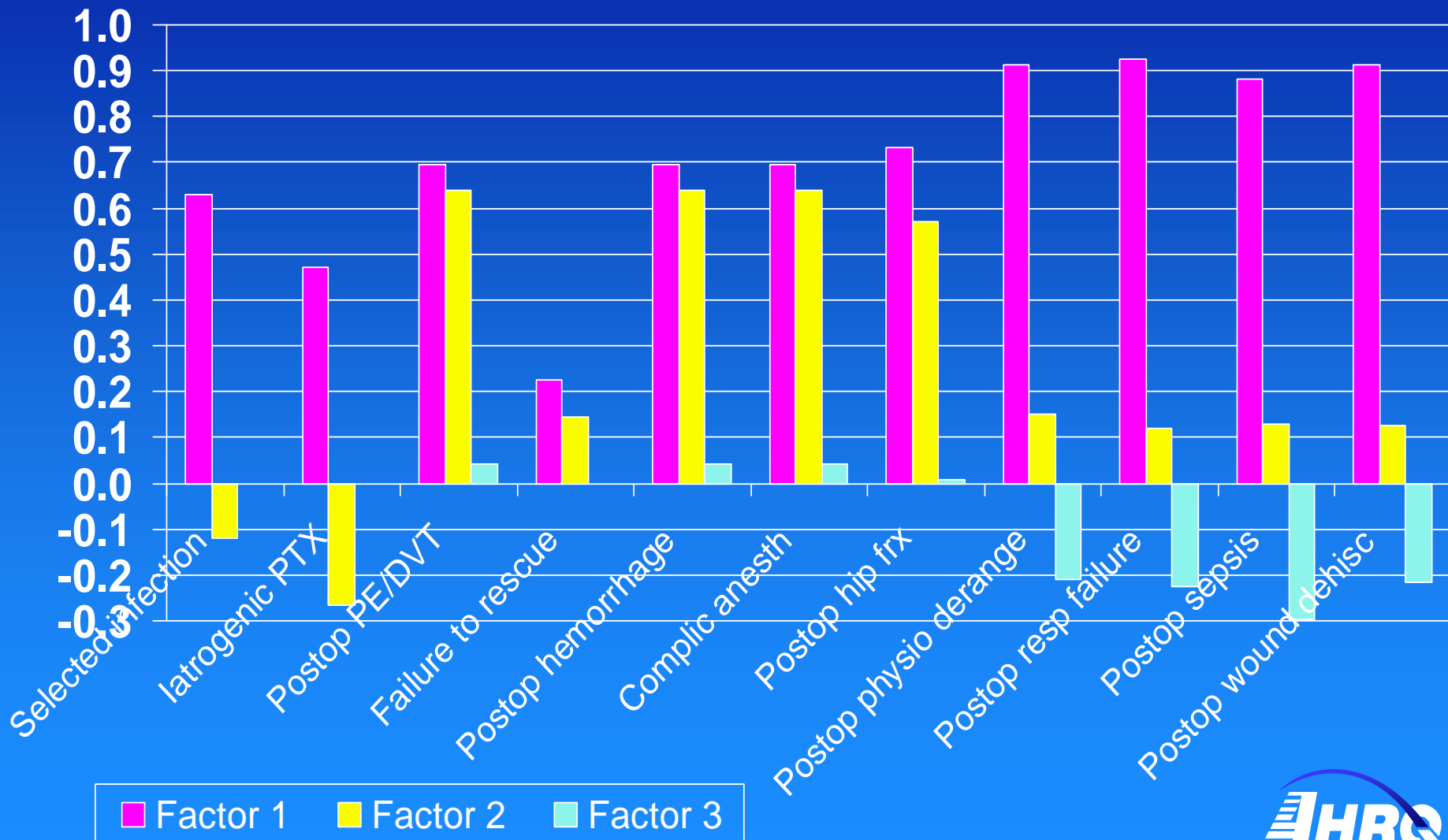
Indicator	Explicit process	Implicit process	Staffing
Complications of anesthesia			
Death in low mortality DRGs		+	
Decubitus ulcer			±
Failure to rescue			++
Foreign body left during procedure			
Iatrogenic pneumothorax			
Selected infections due to medical care			
Postop hip fracture	+	+	
Postop hemorrhage or hematoma	±	+	
Postop physiologic/metabolic derangements			--
Postop respiratory failure	±	+	±
Postop thromboembolism	+	+	±
Postop sepsis			--
Accidental puncture or laceration			
Transfusion reaction			
Postop abdominopelvic wound dehiscence			
Birth trauma			
Obstetric trauma – vaginal birth w instrumentation			
Obstetric trauma – vaginal w/out instrumentation			
Obstetric trauma – cesarean birth			





# Construct validity: Do indicators track together?

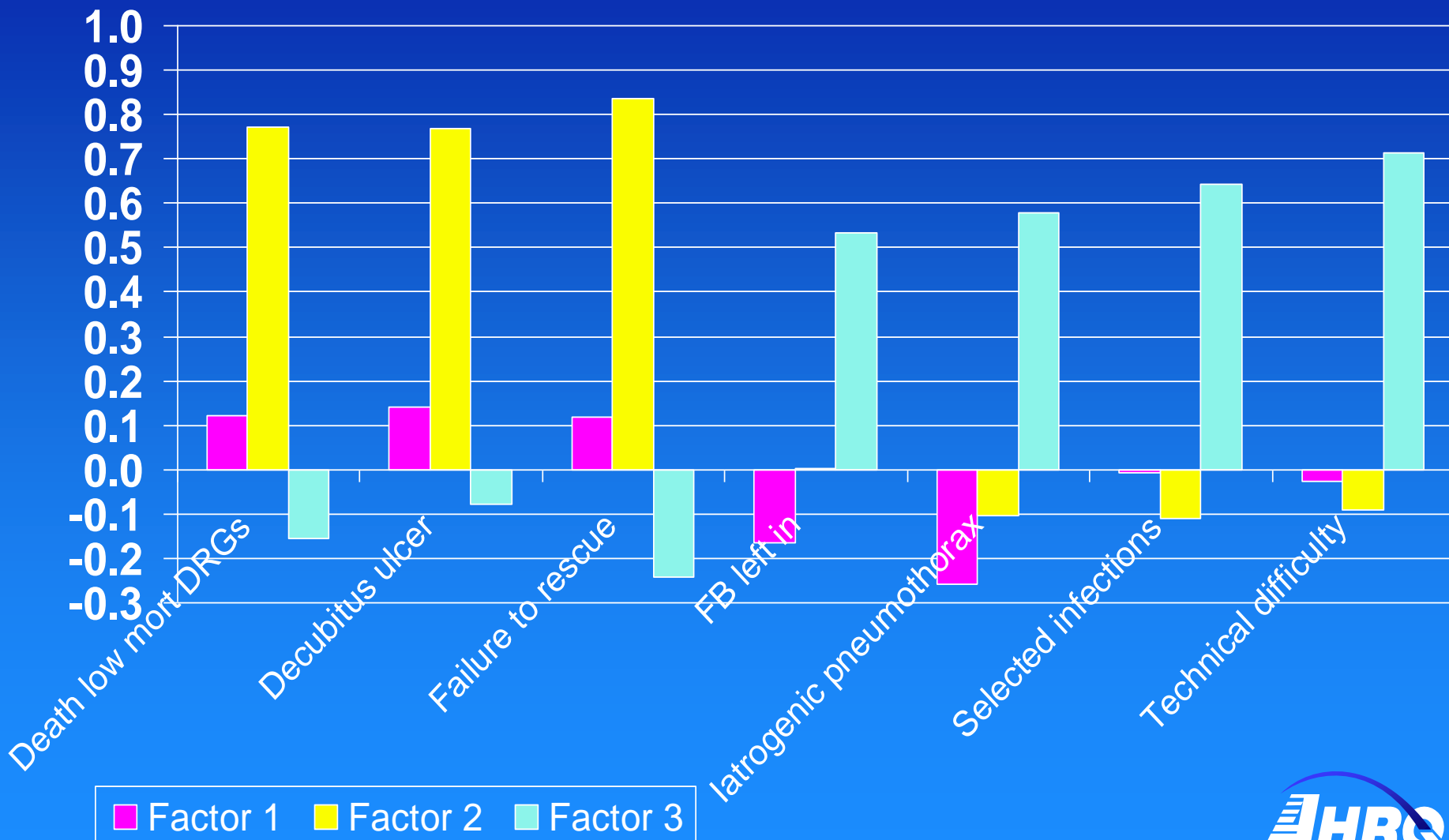
## Factor loadings from 2001 VA/PTF





# Construct validity: Do indicators track together?

## Factor loadings from 2001 VA/PTF





# PSI risk adjustment methods

---

- Must use only administrative data
- APR-DRGs and other canned packages may adjust for complications
- Final model
  - DRGs (complication DRGs aggregated)
  - Modified Comorbidity Index based on list developed by Elixhauser et al.
  - Age, Sex, Age-Sex interactions



# Susceptibility to bias at the hospital level: Impact of risk-adjustment, 1997 SID (summary)

High Bias	Medium Bias	Low Bias
Failure to rescue (44% change 2 deciles)	Postoperative respiratory failure (11%)	Postop abdominopelvic wound dehiscence (4%)
Accidental puncture or laceration (24%)	Postoperative hip fracture (8%)	Obstetric trauma – cesarean birth (2%)
Decubitus ulcer (26%)	Iatrogenic pneumothorax (14%)	Postop hemorrhage or hematoma (4%)
Postop thromboembolism (14%)	Postop physio/metabolic derangement (5%)	Complications of anesthesia (<1%)
Death in low mortality DRGs (13%)	Obstetric trauma – vaginal w instrument (5%)	Obstetric trauma – vaginal w/out instrument
Postoperative sepsis (11%)	Selected infections due to medical care (10%)	Birth trauma (0%)



# Measurement for quality-based purchasing and public reporting: Conclusions

- Quality-based purchasing and public reporting may stimulate improvement in quality of care, or at least more attention to quality indicators
- Measures/indicators must be selected based on local priorities and limitations of available data – AHRQ QIs appropriate for public reporting may differ across states and regions
- Results must be presented and disseminated in a manner that earns the confidence of providers, purchasers/consumers, and other stakeholders

Reference: Remus D, Fraser I. *Guidance for Using the AHRQ Quality Indicators for Hospital-level Public Reporting or Payment.*

AHRQ Publication No. 04-0086-EF.







# Acknowledgments

---

## **Funded by AHRQ**

Support for Quality Indicators II (Contract No. 290-04-0020)

- Mamatha Pancholi, AHRQ Project Officer
- Marybeth Farquhar, AHRQ QI Senior Advisor
- Mark Gritz and Jeffrey Geppert, Project Directors, Battelle Health and Life Sciences

## **Data used for analyses:**

Nationwide Inpatient Sample (NIS), 1995-2000. Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality

State Inpatient Databases (SID), 1997-2002 (36 states). Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality





# Acknowledgments

- We gratefully acknowledge the data organizations in participating states that contributed data to HCUP and that we used in this study: the Arizona Department of Health Services; California Office of Statewide Health and Development; Colorado Health and Hospital Association; CHIME, Inc. (Connecticut); Florida Agency for Health Care Administration; Georgia Hospital Association; Hawaii Health Information Corporation; Illinois Health Care Cost Containment Council; Iowa Hospital Association; Kansas Hospital Association; Maryland Health Services Cost Review Commission; Massachusetts Division of Health Care Finance and Policy; Missouri Hospital Industry Data Institute; New Jersey Department of Health and Senior Services; New York State Department of Health; Oregon Association of Hospitals and Health Systems; Pennsylvania Health Care Cost Containment Council; South Carolina State Budget and Control Board; Tennessee Hospital Association; Utah Department of Health; Washington State Department of Health; and Wisconsin Department of Health and Family Service.





# Questions?

---

