Reliability, Validity และแหล่งข้อมูลสำหรับการวิจัยระบบบริการสุขภาพ

รศ. นพ. จิรุตม์ ศรีรัตนบัลล์

ภาควิชาเวชศาสตร์ป้องกันและสังคม

คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

วัตถุประสงค์การเรียนรู้

- ✦ แหล่งข้อมูลสำหรับการวิจัยระบบบริการสุขภาพ
- ✦ **Reliability** และ **Validity** ของการวิจัย
- ✦ วิธีการและการแปลผลการประเมิน **Reliability** และ **Validity** ของเครื่องวิจัย

แหล่งข้อมูลเพื่อการทำวิจัย

- **ข้อมูลปฐมภูมิ (Primary Data) : ผู้วิจัยออกแบบวิธีการกำหนดเครื่องมือและเก็บข้อมูลเอง**
- **ข้อมูลทุติยภูมิ (Secondary Data) : ผู้วิจัยใช้ข้อมูลจากฐานข้อมูล หรือแหล่งที่มีผู้อื่นรวบรวม บันทึกหรือเก็บข้อมูลไว้ด้วยวิธีการต่างๆ**
  - ***ระเบียบวิธีการเก็บรวบรวมข้อมูลเป็นอย่างไร ?***
  - ***ความลำเอียงอันเนื่องจากจุดมุ่งหมายหลักของการเก็บข้อมูล***

ข้อมูลทุติยภูมิ (Secondary Data)

- แฟ้มข้อมูลการใช้บริการของหน่วยบริการ เช่น รพ., รพ.สต.
- ฐานข้อมูลเพื่อการเบิกจ่ายค่ารักษาพยาบาล เช่น ฐานข้อมูลในระบบกลุ่มวินิจฉัยโรคร่วม, ฐานข้อมูลในระบบ Disease Mgt. ของ สปสช.
- ข้อมูลในระบบรายงานของกระทรวงสาธารณสุข และ/หรือส่วนราชการอื่นๆ
- รายงานวิจัย การสำรวจและฐานข้อมูลจากการสำรวจชุมชนหรือสำรวจประชากรในวงกว้าง

การวัด / การประเมินค่า ?

✦ ค่าที่วัดได้ $X_i = t_i + (S + e_i)$

✦ คุณสมบัติของเครื่องมือ

- มี Random error ที่มีขนาดน้อยที่สุด หรือทำให้ Var (e) น้อยที่สุด
- ไม่มี Systematic error เลยจากการใช้เครื่องมือในการวัดหรือประเมิน หรือ ทำให้ $S = 0$

ความเชื่อถือได้ (Reliability)

หมายถึง ระดับที่วิธีการหรือเครื่องมือวิจัย จะวัดสิ่งใดสิ่งหนึ่ง ได้ผลเหมือนเดิมหรือได้ ผลเท่าเดิมเมื่อวัดซ้ำหลายๆครั้งในตัวอย่าง และสถานการณ์เหมือนเดิม

(Carmines and Zeller, 1979; Kelsey et al., 1986)

Concepts of Reliability

- **Equivalency reliability** : the extent to which two items measure identical concepts at an identical level of difficulty. (Co-occurrence → Correlation)
- **Stability reliability** : the agreement of measuring instruments over time. (sometimes called test, re-test reliability)
- **Internal consistency** : the extent to which tests or procedures assess the same characteristic, skill or quality. (Precision between the observers or of the measuring instruments used in a study)
- **Inter-rater reliability** : the extent to which two or more individuals agree. (Consistency of the implementation of a rating system)

การทดสอบซ้ำ (Retest method)

- ตั้งอยู่บนพื้นฐานที่ว่า ค่าที่เราวัดได้ในต่างเวลาจะ สัมพันธ์กัน เนื่องจากค่าที่ได้แต่ละครั้งสะท้อนถึง ค่าที่แท้จริง (t) เดียวกัน
- "Test-retest reliability"
- สหสัมพันธ์ (Correlations)
- ANOVA table, ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation),
- 2x2 table + Percent agreement, Kappa (K)

ความน่าเชื่อถือโดยการทดสอบซ้ำ

- ความน่าเชื่อถือของตัวอย่างที่วัด

  (Subject reliability)

- ความน่าเชื่อถือของผู้วัดคนใดคนหนึ่ง

  (Intra-observer reliability)

- ความน่าเชื่อถือของผู้ใช้เครื่องมือหรือผู้วัดหลายคน

  (Inter-observer reliability)

Kappa (K)

การวัดครั้งที่ 2 (Second measures) หรือ Rater 2		+	-	Total
การวัดครั้งที่ 1 (First measures) หรือ Rater 1	+	a	b	m_1
	-	c	d	m_2
	Total	n_1	n_2	N

✦ **Chance-corrected percent agreement**

✦ **Prop. of agreement** : $P_0 = (a + d) / N$

Expected prop. : $P_e = (n_1m_1 + n_2m_2) / N^2$

Kappa : $K = (P_0 - P_e) / (1-P_e)$

Std.Err. : $se(K) = \sqrt{P_0(1-P_0) / N (1-P_e)^2}$

95% CI : K - 1.96se(K) to K + 1.96se(K)

Example of Kappa

Rater #2		*+*	*-*	*Total*
Rater #1	+	40	20	60
	-	10	30	40
	Total	50	50	100

P_0 = (40 +30) / 100 = 0.7 หรือ 70%

P_e = (50*60 + 50*40) / 100^2 = 0.5

K = (0.7 - 0.5) / (1 - 0.5) = 0.4 หรือ 40%

se(K) = [0.7*(1-0.7)] / [100*$(1-0.5)^2$] = 0.092

95% CI = [0.4-1.96*0.092, 0.4+1.96*0.092]= 0.22 to 0.58

ความสอดคล้องภายใน
(Internal consistency)

- อาศัยความสอดคล้องภายในคำถามหลายๆข้อ ที่ ถามเกี่ยวกับเรื่องเดียวกันในแบบสอบถามชุดหนึ่งๆ
- เป็นตัวประเมินเสมือนกับการวัดซ้ำแล้วซ้ำอีกใน การประเมินครั้งเดียว
- Cronbach's alpha (Cronbach, 1951)

$$\alpha = N\rho / [\, 1 + \rho\,(N-1)\,]$$

N= # of item ρ = mean inter-item correlations

ในกรณีตัวัดเป็นข้อมูล Dichotomous ใช้สถิติ KR-20

ความถูกต้อง (Validity)

- ระดับที่วิธีการหรือเครื่องมือวิจัยจะวัดสิ่งใดสิ่งหนึ่ง ได้ถูกต้อง ตรงประเด็น ตามที่ได้รับการสร้างขึ้นมา

[the extent to which a measuring device does what it is intended to do]

(Carmines and Zeller, 1979; Kelsey et al., 1986)

ทั้งนี้ ความถูกต้องของการวัดขึ้นกับ (1) ธรรมชาติของสิ่งที่จะวัด และ (2) ความสัมพันธ์ ระหว่างสิ่งที่สามารถวัดได้ กับที่มาหรือเหตุปัจจัยที่ต้องการวัดหรือประเมิน

การประเมินความถูกต้องของเครื่องมือวิจัย

- ความถูกต้องของเนื้อหา

  (Content validity)

- ความถูกต้องโดยอิงกับเกณฑ์

  (Criterion-related validity)

- ความถูกต้องตามการผูกโยง หรือเชิงโครงสร้าง

  (Construct validity)

Content validity

- หมายถึง ระดับความถูกต้องของเครื่องมือ โดยประเมินจากเนื้อหา (content) ของการวัดว่า ได้ ครอบคลุมเนื้อหาประเด็นที่ต้องการ
- ***Face validity*** : เมื่อประเมินดูเนื้อหาของเครื่องมือ วิจัยแล้วผู้เชี่ยวชาญลงความเห็นได้ว่าเครื่องมือนั้นสามารถนำมาใช้วัดสิ่งที่ต้องการวัดได้
- ***Sampling validity*** : เครื่องมือนั้นสามารถวัดสิ่งที่ ต้องการได้ครอบคลุมในทุกแง่มุมอย่างแท้จริง

ตัวอย่าง Content Validity Measurement

- Content Validity Index (CVI)
- Content Validity Ratio (CVR)

- Katz's Percentage (1958)
- Klein & Kosecoff's Correlation (1975)
- Rovinelli & Hambleton's Index of Item-Objective Congruence (1977)
- Aiken's V (1985) content-validity coefficient
- Hambleton's (1980) 12 Steps
- Crocker & Algina (1986)
- Penfield's (2003) Score Interval

ปัญหาของ Content validity

- สวนทางกับความสอดคล้องภายใน **(Internal consistency)**
- หาวิธีที่จะประเมิน **Content validity** อย่างแน่นอน ได้ยาก ทำให้มีนักวิจัยบางคนไม่ยอมรับแนวคิดนี้
- ต้องมีพื้นฐานความรู้ที่ใช้เป็นกรอบเนื้อหาในการ ประเมินเครื่องมือ

Criterion-related validity

- เปรียบเทียบความสอดคล้องของผลที่ได้ จากการวัด โดยอาศัยเครื่องมือวิจัย กับสิ่งที่สังเกตได้โดยตรง หรือ ผลจากวิธีวัดมาตรฐานที่เป็น **Gold Standard**
- ***Concurrent validity*** : เปรียบเทียบผลการวัดจาก เครื่องมือกับ **Criterion** จากการวัดที่เวลาเดียวกัน
- ***Predictive validity*** : เปรียบเทียบผลจากเครื่องมือ กับ **Criterion** ที่จะปรากฏขึ้นในอนาคต

Sensitivity & Specificity

Test	Disease present	Disease absent	
Positive	27	35	
Negative	10	77	
Total	37	112	

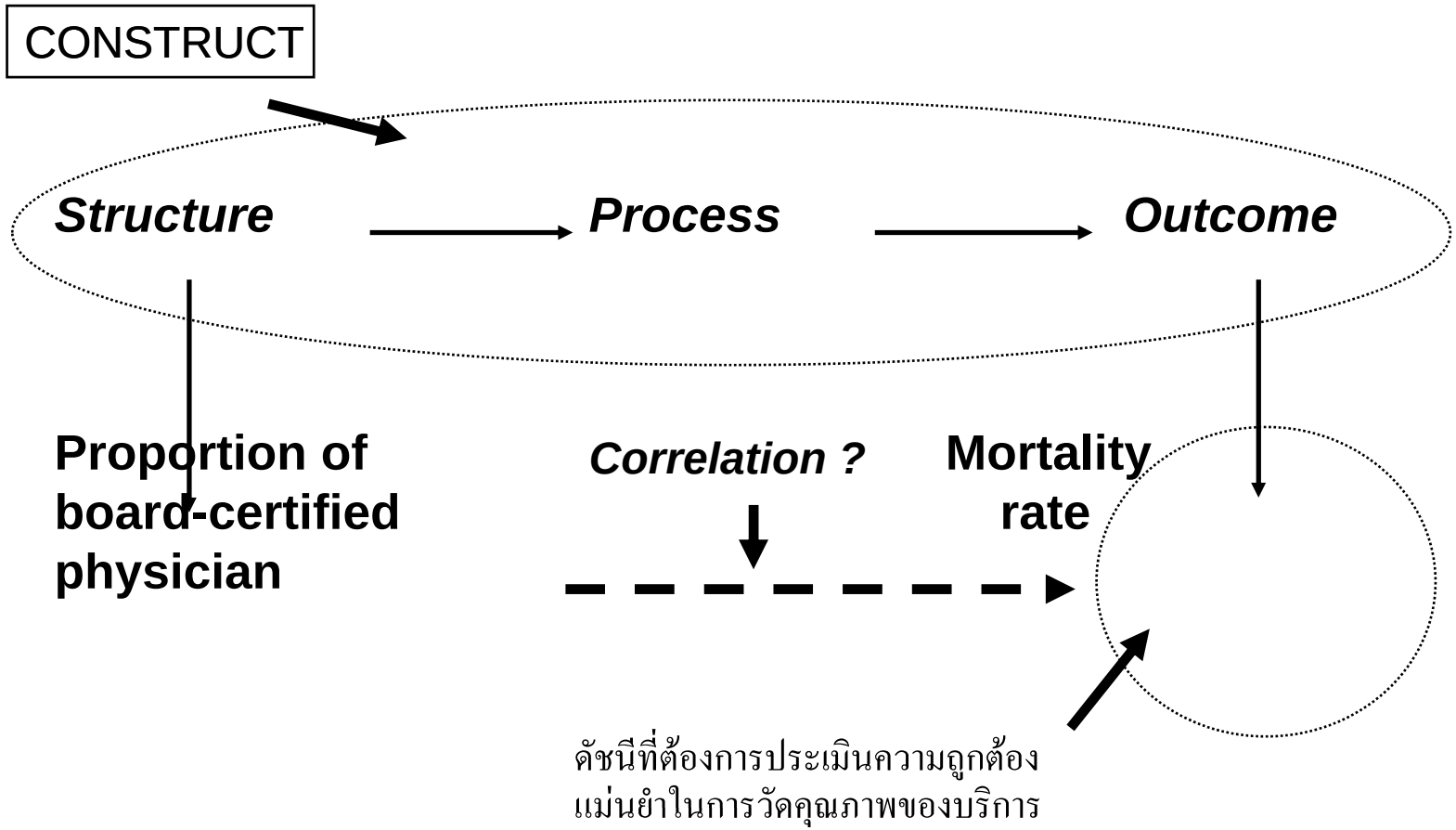
Sensitivity = 27 / (27+10) = 73%

Specificity = 77 / (35+77) = 69%

Construct validity

- เป็นการวัดระดับความสัมพันธ์ระหว่างเครื่องมือที่ใช้วัด กับตัวชี้วัดอื่นๆ ที่สอดคล้องกันภายใต้สมมติฐานเกี่ยวกับกรอบแนวคิดที่ต้องการจะวัด ที่ได้จากทฤษฎีเดียวกัน
- มักจะใช้เมื่อไม่มี **Criterion** หรือ เนื้อหาตามกรอบ แนวความคิดที่ได้รับการยอมรับอยู่ก่อน
- ประเมินทั้งทฤษฎีและเครื่องมือพร้อมกัน
- **Measured by the correlation between the intended independent variable (construct) and the proxy independent variable (indicator, sign) that is actually used. [Hunter and Schmidt (1990)]**

Mortality rate in measuring quality of hospital services



Factor analysis

and

Construct validity

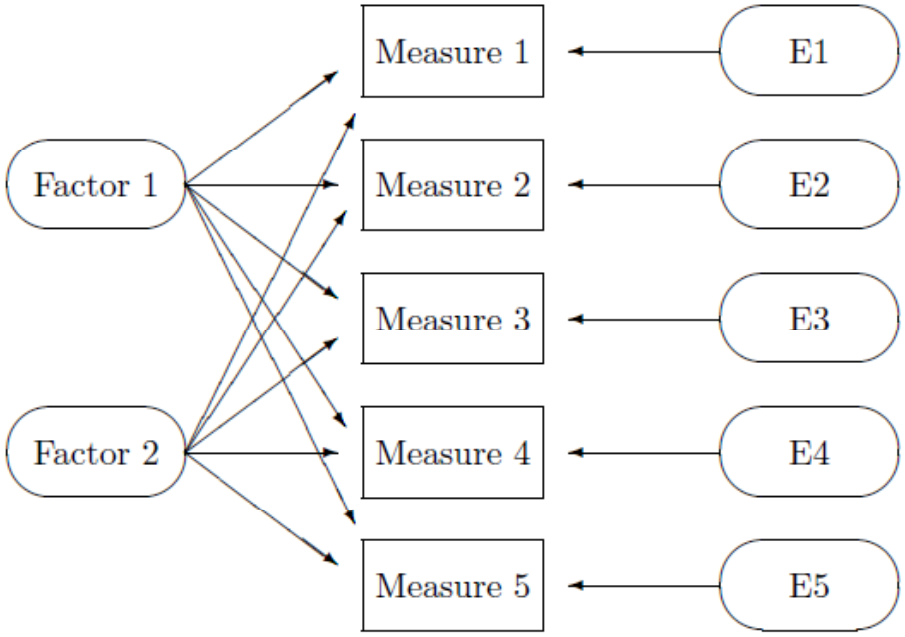
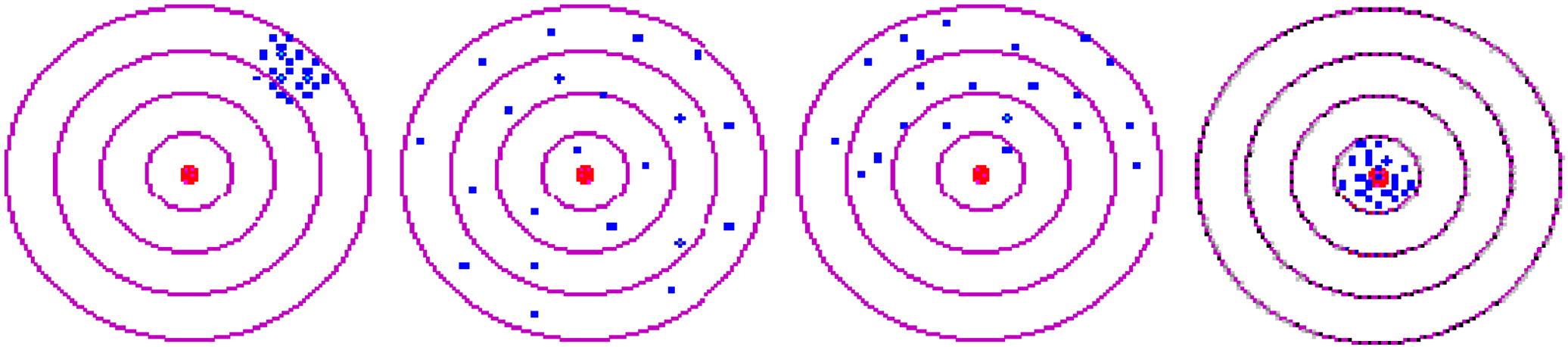


Table 2. *Differences Between Exploratory and Confirmatory Factor Analysis*

Exploratory Factor Analysis	Confirmatory Factor Analysis
Heuristic-weak literature base	Strong theory and/or strong empirical base
Determine the number of factor	Number of factors fixed a priori
Determine whether the factors are correlated or uncorrelated	Factors fixed a priori as correlated or uncorrelated
Variables free to load on all factors	Variable fixed to load on a specific factor or factors

Note. From *Applied multivariate statistics for the social sciences* (Stevens, 1996)

Reliability vs. Validity



Reliable
Not Valid

Valid
Not Reliable

Neither Reliable
Nor Valid

Both Reliable
And Valid

Questions
&
Answers