เอกสารคำสอน
เรื่อง
Quantitative Methods

กระบวนวิชาบทนำสุขภาพชุมชน 2

พ.วช. 315 (322315)

ปีการศึกษา 2556

ผศ.นพ. เกรียงไกร ศรีธนวิบุญชัย

ภาควิชาเวชศาสตร์ชุมชน

คำนำ

เอกสารคำสอนฉบับนี้เขียนขึ้นเพื่อใช้ประกอบการเรียนการสอนเรื่อง Quantitative methods กระบวนวิชาบทนำสุขภาพชุมชน 2 (พ.วช. 315) ปีการศึกษา 2556 สำหรับนักศึกษาแพทย์ชั้นปีที่ 3 คณะแพทยศาสตร์ มหาวิทยาลัยเชียงใหม่ เนื้อหาในเอกสารนี้ เป็นความรู้เบื้องต้นเกี่ยวกับวิธีการทางระบาดวิทยา และสถิติทางการแพทย์ ตามที่ระบุไว้ในเกณฑ์มาตรฐานแพทยสภา พ.ศ. 2545 และเพื่อให้ง่ายต่อการทำความเข้าใจของนักศึกษา ผู้เขียนจึงได้เขียนเนื้อหาแยกเป็นเรื่อง ๆ ตามที่เกณฑ์ฯ กำหนดไว้ และได้ระบุเลขข้อของเกณฑ์ฯ ที่สอดคล้องกับชื่อเรื่องแต่ละเรื่องไว้ด้วย

ความรู้เกี่ยวกับ Quantitative methods จะเป็นพื้นฐานสำคัญสำหรับนักศึกษา ในการเรียนระดับชั้นคลินิก และการประกอบวิชาชีพแพทย์ต่อไปในอนาคต โดยจะเป็นองค์ความรู้ที่จะช่วยให้นักศึกษาสามารถอ่านทำความเข้าใจรายงานวิจัยทางการแพทย์ และศึกษาต่อยอดเกี่ยวกับเวชศาสตร์อิงหลักฐาน (Evidenced-based medicine) ซึ่งเป็นหลักวิชาเกี่ยวกับการประเมินคุณค่าความน่าเชื่อถือของรายงานวิจัยทางการแพทย์ เพื่อช่วยในการตัดสินใจเลือกวิธีการป้องกันโรค การตรวจคัดกรองโรค การส่งตรวจวินิจฉัย การให้การวินิจฉัย การพยากรณ์โรค ตลอดจนวิธีการรักษาผู้ป่วย ที่เหมาะสมที่สุดจากองค์ความรู้ที่มี ณ เวลานั้น ๆ นอกจากนั้นความรู้ความเข้าใจเกี่ยวกับ Quantitative methods นี้ยังจำเป็นสำหรับใช้ในการออกแบบและทำวิจัยด้วยตนเองของแพทย์ในอนาคตอีกด้วย

เอกสารนี้เป็นเพียงเอกสารสำหรับอ่านประกอบเท่านั้น ให้ถือเอาเนื้อหาที่อาจารย์แต่ละท่านสอนในห้องเรียนเป็นขอบเขตของเนื้อหาที่นักศึกษาต้องเตรียมตัวสำหรับการสอบ นอกจากนั้นนักศึกษาควรแสวงหาความรู้จากตำราและแหล่งความรู้ต่าง ๆ เพิ่มเติมด้วย การเรียนวิชานี้ต่างจากเนื้อหาวิชาทางการแพทย์อื่น ๆ ตรงที่ต้องอาศัยการทำความเข้าใจมากกว่าการจดจำ ผู้เขียนจึงหวังว่า นักศึกษาจะหมั่นศึกษา ฝึกฝน และปฏิบัติเพิ่มเติมด้วยตนเอง ในประเด็นที่เกี่ยวข้อง ซึ่งจะช่วยให้เกิดความรู้ความเข้าใจที่ลึกซึ้งยิ่ง ๆ ขึ้น และส่งผลดีต่อการทำหน้าที่แพทย์ต่อไป

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย
พฤศจิกายน 2556

สารบัญ

หน้า

แผนการสอน	4
1.11.1.1 Scales of measurement	7
1.11.1.2 Distribution, central tendency, variability, probability	12
1.11.1.3 Disease prevalence and incidence	25
1.11.1.4 Disease outcomes	30
1.11.1.5 Associations	34
1.11.1.6 Health impact	44
1.11.1.7 Efficacy of drug therapy, 1.11.1.8 Risk of drug therapy	49
1.11.1.9 Sensitivity, specificity, PPV, NPV	55
1.11.2.1 Types of experimental studies	65
1.11.2.2 Types of observational studies	71
1.11.2.3 Sampling and sample size	80
1.11.2.4 Subject selection and exposure allocation	86
1.11.2.5 Outcome assessment	96
1.11.2.6 Internal and external validity	100
1.11.2.7 Level of evidence	104
1.11.3.1 Confidence Interval	105
1.11.3.2 Statistical significance and Type I error, 1.11.3.3 Statistical power and Type II error	111

1.11.1.1 Scales of measurement

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

ตัวแปร (Variable) และข้อมูล (Data)

ในการศึกษาเชิงปริมาณ เราพิจารณาชุดข้อมูล (Data set) ซึ่งประกอบกันขึ้นจากข้อมูลของตัวแปรต่าง ๆ *ตัวแปร* หมายถึง ลักษณะหรือคุณสมบัติร่วมกันของประชากรที่ผู้ศึกษาสนใจ ค่าของตัวแปรที่เปลี่ยนแปลงไปตามแต่ละบุคคล หรือสิ่งที่วัดเรียกว่า*ข้อมูล* ข้อมูลอาจเป็นสัญลักษณ์ หรือค่าตัวเลขที่บอกลักษณะของตัวแปรที่ศึกษา โดยทั่วไปการศึกษาเชิงปริมาณทางการแพทย์ สนใจบรรยายลักษณะ หาความแตกต่าง ประมาณค่า หรือความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ในชุดข้อมูลนั้น

<u>*ตัวอย่าง*</u>	*ตัวแปร*	*ข้อมูล*
	เพศ	*ชาย, หญิง*
	ฐานะทางเศรษฐกิจ	*หาไม่พอใช้, พอใช้, เหลือเก็บ*
	จำนวนบุตร	*1, 2, 0, 2, 3, 1, …*
	อุณหภูมิ(^{0}C)	*36.8, 37.9, 40.0, …*
	น้ำหนัก (Kg.)	*56.7, 71.4, 66.0, 49.5, …*

ประเภทของตัวแปร (Types of variable)

ลักษณะของตัวแปรมีความสำคัญอย่างยิ่งต่อการศึกษาเชิงปริมาณ เนื่องจากตัวแปรที่มีลักษณะต่างกัน จะมีวิธีวัด ตลอดจนการวิเคราะห์ข้อมูลที่แตกต่างกันออกไป ตัวแปรอาจแบ่งได้เป็น 2 ประเภท ใหญ่ ๆ คือ

1 ***ตัวแปรบอกลักษณะ (Qualitative variable)*** หรือตัวแปรจัดกลุ่ม (Categorical variable) คือตัวแปรที่อธิบายถึงลักษณะที่สนใจของประชากรศึกษานั้น ๆ ซึ่งแม้เราจะไม่สามารถใช้เครื่องมือวัดตัวแปรนี้ออกมาเป็นค่าตัวเลขได้ แต่เราสามารถนับและจัดแยกประชากรศึกษาออกเป็นกลุ่ม ๆ ตามลักษณะของตัวแปรได้ ข้อมูลลักษณะนี้นิยมบรรยายเป็นสัดส่วน (Proportion) หรือร้อยละ เช่นร้อยละของกลุ่มตัวอย่างที่เป็นเพศชายในประชากรศึกษาทั้งหมด ตัวแปรบอกลักษณะที่ใช้มากในการศึกษาทางการแพทย์จะมีลักษณะพิเศษ คือจัดแยกประชากรตามลักษณะตัวแปรออกได้เพียง 2 กลุ่ม เรียกว่า Binary variable หรือ Dichotomous variable

<u>*ตัวอย่าง*</u>	*ตัวแปร*	*ข้อมูล*	
	ฐานะทางเศรษฐกิจ	*หาไม่พอใช้, พอใช้, เหลือเก็บ*	
	เพศ	*ชาย, หญิง*	*Binary variable*
	การรอดชีวิต	*เสียชีวิต, รอดชีวิต*	*Binary variable*
	ผลเลือด HIV	*บวก, ลบ*	*Binary variable*

2. ***ตัวแปรบอกปริมาณ (Quantitative variable)*** คือตัวแปรที่ข้อมูลของประชากรศึกษาได้จากการวัด หรือจากการนับจำนวนเหตุการณ์ที่เกิดขึ้น ข้อมูลของตัวแปรประเภทนี้บอกปริมาณในลักษณะของ

ตัวเลข (Numerical) เพื่ออธิบายลักษณะของข้อมูลของตัวแปรเชิงปริมาณ นิยมบอกแนวโน้มเข้าสู่ส่วนกลาง (Central tendency) เช่น ค่าเฉลี่ย (Mean) และการกระจายของข้อมูล เช่น ค่าเบี่ยงเบนมาตรฐาน (Standard deviation, SD) ตัวแปรบอกปริมาณยังสามารถแบ่งย่อยออกได้เป็นอีก 2 ประเภท คือ ตัวแปรเชิงปริมาณไม่ต่อเนื่อง และตัวแปรเชิงปริมาณต่อเนื่อง อย่างไรก็ตามโดยทั่วไปการจัดการข้อมูลรวมถึงเทคนิคการวิเคราะห์ข้อมูลของตัวแปรย่อยทั้งสองประเภทไม่ต่างกัน

2.1. ตัวแปรไม่ต่อเนื่อง (Discrete variable) คือตัวแปรเชิงปริมาณที่ข้อมูลมีช่องว่างหรือช่วงเว้นระหว่างค่าหนึ่ง ๆ ไปยังค่าถัดไป และมักจะมีความหมายเมื่อกำหนดเป็นตัวเลขจำนวนเต็มเท่านั้น ไม่สามารถกำหนดเป็นทศนิยมได้ มักเป็นข้อมูลที่ได้จากการนับ

ตัวอย่าง	*ตัวแปร*	*ข้อมูล*
	จำนวนบุตร	*1, 2, 0, 2, 3, 1, …*
	จำนวนครั้งของการมา รพ./ปี	*5, 0, 3, 11, 4, 3, …*
	จำนวนบุหรี่ที่สูบ (มวน/วัน)	*20, 4, 9, 18, 15,…*

2.2. ตัวแปรต่อเนื่อง (Continuous variable) คือตัวแปรเชิงปริมาณที่ไม่มีช่องว่างระหว่างข้อมูล เป็นข้อมูลที่ได้จากการวัดจึงมักต้องระบุหน่วยที่วัดด้วย ข้อมูลลักษณะนี้สามารถมีความหมายแม้ไม่เป็นจำนวนเต็มทำให้มีจุดทศนิยมได้ ในทางทฤษฎีจะวัดลงไปถึงระดับย่อยแค่ไหนก็ได้ ขึ้นกับความละเอียดที่ต้องการ และความสามารถในการวัดของเครื่องมือ

ตัวอย่าง	*ตัวแปร*	*ข้อมูล*
	น้ำหนักตัว (Kg.)	*56.7, 71.4, 66.0, 49.5,…*
	น้ำตาลในเลือด (mg.%)	*96.5, 102.1, 75.8, 137.0,…*
	ความเข้มข้นของเลือด (%)	*41, 45, 34, 39, 43,…*

เพื่อประโยชน์ในการวิเคราะห์ข้อมูลและการนำไปอธิบายทางการแพทย์ บ่อยครั้งที่ผู้ศึกษาลดระดับของตัวแปร จากตัวแปรบอกปริมาณมาเป็นตัวแปรบอกลักษณะ โดยการจัดกลุ่มของข้อมูลบอกปริมาณเข้าด้วยกัน

ตัวอย่าง	*ตัวแปร*	*ข้อมูล*
	จำนวนบุหรี่ที่สูบ (มวน/วัน)	*20, 4, 9, 18, 15, …*
	↓	↓
	การสูบบุหรี่	*0 ไม่สูบ / 1-10 สูบน้อย / 11 – 20 สูบปานกลาง / 21 ขึ้นไป สูบมาก*

ระดับการวัดข้อมูล (Scales of measurement)

เราสามารถแบ่งระดับการวัดข้อมูลเชิงปริมาณตามการได้มาซึ่งข้อมูล และลักษณะของตัวแปรได้เป็น 4 ระดับ โดยตัวแปรบอกลักษณะมีระดับการวัด 2 ระดับ ได้แก่ ระดับกลุ่ม (Nominal scale) และระดับอันดับ (Ordinal scale) ส่วนตัวแปรบอกปริมาณ มีระดับการวัดอีก 2 ระดับ คือ ระดับช่วง (Interval scale) และระดับอัตราส่วน (Ratio scale) ระดับการวัดที่แตกต่างกันจะมีวิธีการจัดการข้อมูล และสถิติที่ใช้ในการวิเคราะห์ข้อมูลที่แตกต่างกัน

1. ***ระดับกลุ่ม* (Nominal scale)** เป็นระดับการวัดที่ต่ำที่สุด หมายถึงการนับแยกกลุ่มประชากรที่มีลักษณะเหมือนกันไว้ให้อยู่ในกลุ่มเดียวกันตามลักษณะตัวแปรที่กำหนด สัญลักษณ์ที่กำหนดให้แต่ละกลุ่มเป็นเพียงชื่อเรียกเท่านั้น แม้กำหนดเป็นตัวเลขก็ไม่มีความหมายเชิงปริมาณ สิ่งที่นำไปใช้คำนวณในการวิเคราะห์ข้อมูลคือสัดส่วนของจำนวนสมาชิกของแต่ละกลุ่ม เมื่อจัดเป็นกลุ่มแล้วข้อมูลแต่ละกลุ่มไม่สามารถเปรียบเทียบกัน และไม่สามารถจัดอันดับได้ ว่ากลุ่มใดอยู่สูงกว่า หรือมีค่ามากกว่ากลุ่มอื่น ๆ

<u>ตัวอย่าง</u>	*ตัวแปร*	*ข้อมูล*
	เพศ	*ชาย, หญิง*
	การรอดชีวิต	*เสียชีวิต, รอดชีวิต*
	ผลเลือด HIV	*บวก, ลบ*
	กรุ๊ปเลือด	*A, B, AB, O*
	หอผู้ป่วย	*อายุรกรรม, ศัลย์ฯ, สูติฯ, กุมารฯ, …..*

2. ***ระดับอันดับ* (Ordinal scale)** หมายถึงระดับการวัดโดยการคัดแยกกลุ่มประชากรที่มีลักษณะเหมือนกันให้อยู่ในกลุ่มเดียวกันเช่นเดียวกับการวัดระดับกลุ่ม แต่ข้อมูลที่ได้จะมีความหมายมากกว่า คือข้อมูลเมื่อแบ่งกลุ่มแล้ว สามารถเรียงลำดับได้ว่ากลุ่มใดมีค่ามากกว่าหรือน้อยกว่า หรืออยู่เหนือกลุ่มอื่น ๆ อย่างไรก็ตามเราไม่สามารถบอกขนาดความแตกต่างระหว่างกลุ่มว่าเป็นปริมาณเท่าใด ตัวอย่างเช่นเราบอกได้ว่า มะเร็งระยะที่สามรุนแรงกว่าระยะที่สอง และระยะที่สองรุนแรงกว่าระยะที่หนึ่ง แต่เราไม่สามารถบอกได้ว่า มะเร็งระยะที่สามรุนแรงกว่าระยะที่สองในขนาดที่เท่ากับมะเร็งระยะที่สองรุนแรงกว่าระยะที่หนึ่ง

<u>ตัวอย่าง</u>	*ตัวแปร*	*ข้อมูล*
	ฐานะทางเศรษฐกิจ	*หาไม่พอใช้, พอใช้, เหลือเก็บ*
	การสูบบุหรี่	*ไม่เคยสูบ, เคยสูบเลิกแล้ว, สูบน้อย, สูบมาก*
	Cancer staging	*I, II, III, IV*
	ความเจ็บปวด	*เล็กน้อย, ปานกลาง, รุนแรง, ทนไม่ได้*

3. ***ระดับช่วง* (interval scale)** คือการวัดตัวแปรเชิงปริมาณที่ค่าที่ได้เป็นตัวเลข บอกได้ว่าค่าที่ได้มีปริมาณมากน้อยกว่ากันเท่าใด อย่างไรก็ตามข้อมูลลักษณะนี้ จะไม่มี 0 สมบูรณ์ หมายถึง ถ้าวัดค่าได้ 0 ไม่ได้หมายความว่าไม่มีคุณสมบัตินั้น ๆ เลย ระยะห่างระหว่างข้อมูลที่ได้จากการวัดแบบนี้เปรียบเทียบกันได้ ทำให้สามารถนำมาบวกลบกันได้ แต่การนำข้อมูลมาคูณหารกันจะไม่มีความหมาย

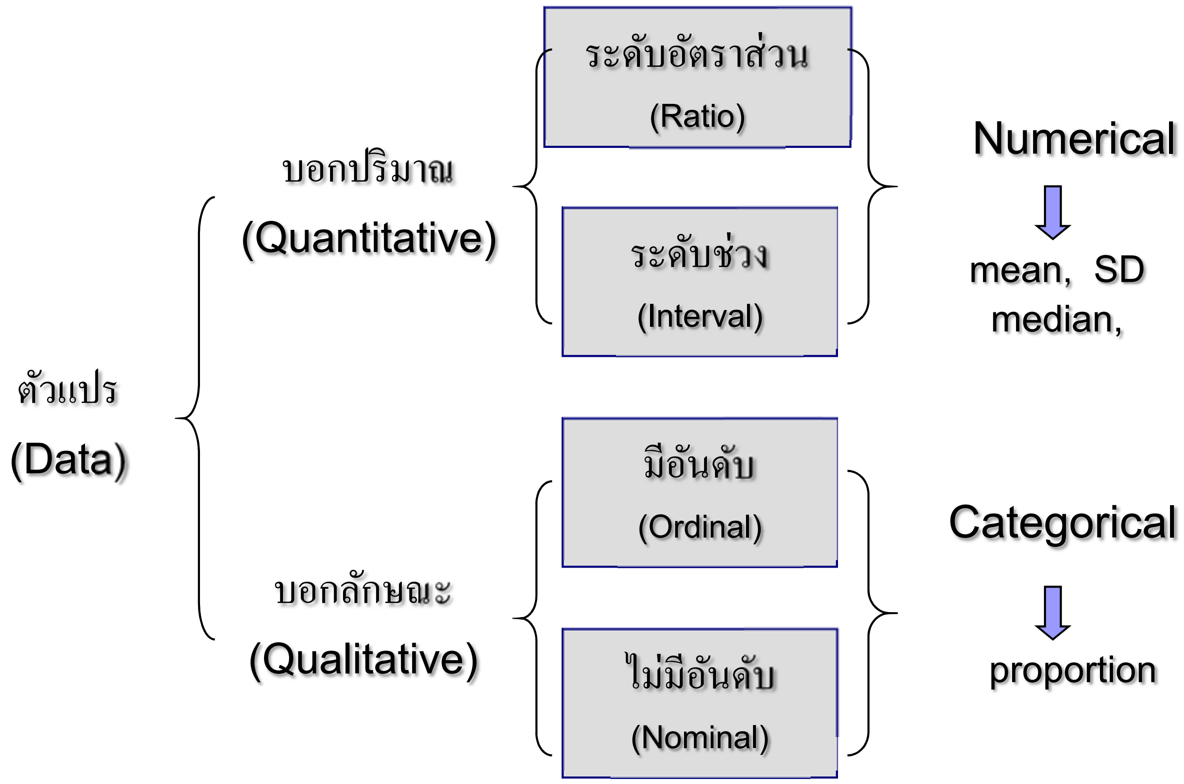
<u>ตัวอย่าง</u>	*ตัวแปร*	*ข้อมูล*
	อุณหภูมิ (^{0}C)	*36.8, 37.9, 40.0, …*
	คะแนนจากแบบทดสอบความเครียด	*12, 8, 9, 21, 18, …*

จากตัวอย่างข้างต้น จะเห็นว่า อุณหภูมิ 0 ^{0}C ไม่ได้หมายถึงไม่มีความร้อนอยู่ และเช่นเดียวกัน หากใครก็ตามทำแบบทดสอบความเครียดแล้วได้ 0 คะแนน ก็ไม่ได้หมายถึงว่าผู้นั้นไม่มีความเครียดเลย

4. ***ระดับอัตราส่วน (Ratio scale)*** หมายถึงการวัดตัวแปรเชิงปริมาณที่มี 0 สมบูรณ์ ค่า 0 เท่ากับไม่มีคุณสมบัตินั้น ๆ เลย ทำให้เราสามารถนำข้อมูลลักษณะนี้มาบวกลบกันเหมือนข้อมูลระดับช่วง รวมถึงนำมาคูณหารกันเพื่อหาอัตราส่วนโดยยังมีความหมายได้ด้วย จัดเป็นการวัดระดับขั้นสูงสุด

ตัวอย่าง	*ตัวแปร*	*ข้อมูล*
	จำนวนบุตร	*1, 2, 0, 2, 3, 1,…*
	น้ำหนักตัว (Kg.)	*56.7, 71.4, 66.0, 49.5,…*
	น้ำตาลในเลือด (mg.%)	*96, 102, 75, 137,…*

แผนภูมิแสดงประเภทของตัวแปร และระดับการวัด



****

เอกสารอ้างอิง

1. Norman RG, Streiner LD. PDQ Statistics. 2nd ed. St. Louis; Mosby: 1997.
2. Knapp GR, Miller MC. Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins; 1992.
3. Danial WW. Biostatistics Basic concepts and Methodology for the Health Sciences. 9th Edition. Wiley@Sons, Inc, 2010

****

แบบฝึกหัด

1. การวิจัยในผู้ป่วยมะเร็งปอดที่รับไว้เป็นผู้ป่วยใน ตัวแปรใดใช้การวัดข้อมูลระดับ Ordinal scale
   A. Tumor staging
   B. Histopathology
   C. Body temperature
   D. ประวัติ Chemotherapy
   E. จำนวนวันที่นอนโรงพยาบาล

2 จำนวนบุตร จัดเป็นตัวแปรลักษณะใด
   A. Binary
   B. Discrete
   C. Continuous
   D. Qualitative
   E. Categorical

3. การศึกษาเกี่ยวกับการใช้ยาปฏิชีวนะในการรักษาโรคอุจจาระร่วงผู้ทำการวิจัยได้เก็บข้อมูลคนไข้ ตัวแปรใดเป็นแบบ Ratio
   A. อายุ
   B. เพศ
   C. ระดับการศึกษา
   D. การวินิจฉัยแยกโรค
   E. ชนิดของยาปฏิชีวนะ

4. ศูนย์มะเร็งแห่งหนึ่งศึกษาเกี่ยวกับความรุนแรงของมะเร็งเต้านมซึ่งมี 4 ระยะ คิดเป็นเปอร์เซ็นต์ ได้ดังนี้ 20, 35, 35, และ 10 การศึกษานี้มีลักษณะข้อมูลแบบใด
   A. Interval
   B. Ordinal
   C. Ratio
   D. Nominal

5. การศึกษาในผู้ติดเชื้อ HIV ตัวแปรใดต่อไปนี้ใช้ระดับการวัดข้อมูลแบบ Interval scale?
   A. CD4+ cell
   B. HIV staging
   C. HIV risk factor
   D. Body mass index
   E. Quality of life score

1.11.1.2 Variability, distribution, central tendency, probability

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

ความไม่แน่นอนของข้อมูล (Variability)

ข้อมูลทางการแพทย์ โดยเฉพาะอย่างยิ่งข้อมูลบอกปริมาณที่ได้จากการวัด มีความไม่แน่นอนเกิดขึ้นได้ การวัดเหตุการณ์ทางคลินิกเดียวกัน ทั้งที่อันที่จริงควรจะได้ค่าที่เท่ากันก็อาจได้ค่าที่เปลี่ยนไปไม่เท่ากันได้ ซึ่งเกิดจากปัจจัยหลายประการ แพทย์จึงควรทราบถึงหลักการและสาเหตุที่เป็นไปได้ต่าง ๆ ของปรากฏการณ์นี้

ความไม่แน่นอนของข้อมูลทางการแพทย์ แบ่งได้เป็น 2 ประเภทคือ

1. ***ความไม่แน่นอนทางชีวภาพ (Biologic variation)*** หมายถึงการวัดค่าทางคลินิกในบุคคลเดียวกัน แต่ได้ค่าต่างกันเนื่องจากปัจจัยทางชีวภาพภายในของบุคคลนั้น ๆ

   1.1 ความไม่แน่นอนทางชีวภาพแท้ (True biologic variation) เป็นความคลาดเคลื่อนเชิงสุ่ม (Random variation) ของผลการวัดตามหลักของความน่าจะเป็น เกิดจากปัจจัยภายในบุคคลตามธรรมชาติที่ไม่ทราบสาเหตุแน่นอน ทำให้การวัดแต่ละครั้งได้ค่าที่แตกต่างกัน

   <u>*ตัวอย่าง*</u>

   การวัดความดันโลหิตในผู้ป่วยรายหนึ่ง โดยทำการวัดซ้ำหลายครั้งต่อเนื่องกันไป ภายใต้สภาพการณ์ที่เหมือนกัน ปรากฏว่าได้ค่าความดันโลหิตที่แตกต่างกัน

   1.2 ความไม่แน่นอนจากเวลาการวัด (Temporal biologic variation) เกิดจากการเปลี่ยนแปลงทางคลินิกภายในบุคคล ตามระยะเวลาที่เปลี่ยนแปลงไป บางครั้งสัมพันธ์กับเวลาต่าง ๆของแต่ละวัน

   <u>*ตัวอย่าง*</u>

   การวัดความดันโลหิตในผู้ป่วยรายหนึ่ง พบว่าความดันโลหิตในเวลาบ่ายมักสูงกว่าตอนเช้า

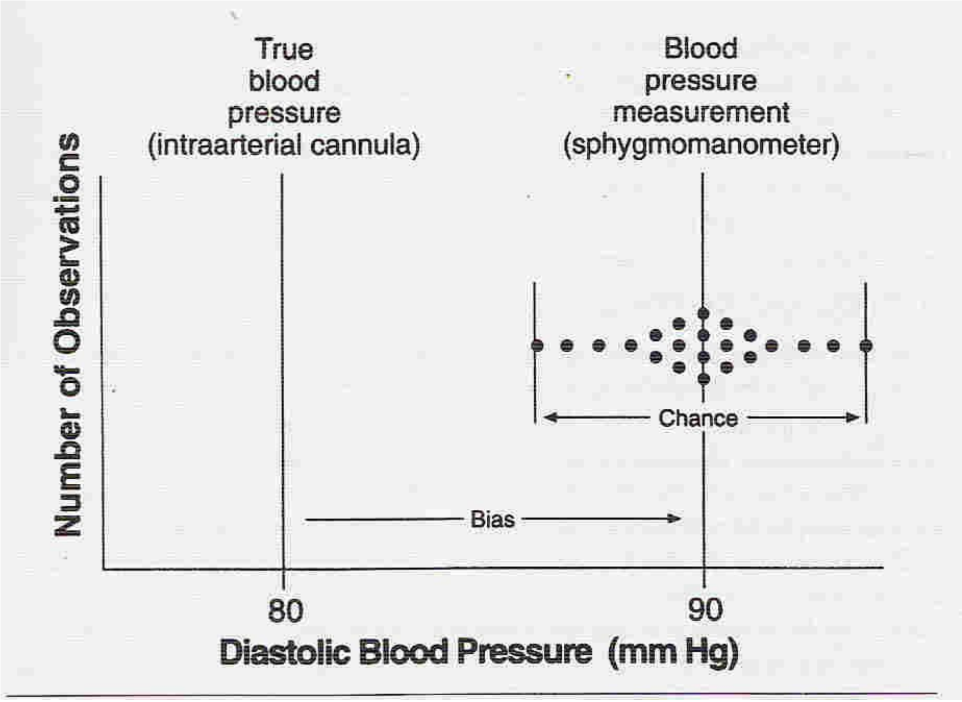
2. ***ความไม่แน่นอนที่เกิดจากการวัด (Measurement variation)*** บางครั้งเรียกว่าความผิดพลาดเนื่องจากการวัด (Measurement error) หมายถึง ความแตกต่างของข้อมูลที่เกิดเนื่องจากกระบวนการวัด อาจเนื่องจากตัวเครื่องมือเอง หรือจากตัวผู้วัด

   2.1. ความคลาดเคลื่อนเชิงสุ่มที่เกิดจากการวัด (Random measurement error) เกิดจากการแกว่งตัวของค่าที่วัดได้เนื่องจากความไม่แน่นอนตามธรรมชาติ (Chance) ซึ่งเป็นสิ่งที่เกิดขึ้นเอง ไม่สามารถป้องกันได้ และไม่ได้เกิดจากความผิดพลาดของกระบวนการวัด โดยโอกาสที่จะวัดได้ค่าสูงกว่าความจริง จะพอ ๆ กันกับโอกาสที่จะวัดได้ค่าต่ำกว่าความจริง ทำให้หากวัดซ้ำหลายครั้งในจำนวนตัวอย่างที่มากพอจะได้ค่าเฉลี่ยที่ใกล้เคียงความจริงเสมอ ไม่นำไปสู่อคติของผลการวัด

   2.2. ความคลาดเคลื่อนเชิงระบบที่เกิดจากการวัด (Systematic measurement error) บางครั้งเรียกว่าอคติ (Bias) เกิดเนื่องจากความผิดพลาดของกระบวนการวัด ซึ่งอาจเป็นผลมาจากวิธีการวัดที่ไม่ถูกต้อง หรือความไม่เที่ยงตรงของเครื่องมือที่ใช้วัด หรือความผิดพลาดของการอ่านค่า โดยทั่วไปความผิดพลาดลักษณะนี้ ค่าที่ได้จะมีทิศทางไปด้านใดด้านหนึ่ง เช่น สูงกว่าความจริง หรือต่ำกว่าความจริง ขึ้นอยู่กับทิศทางของความผิดพลาด การวัดที่ดีควรมีข้อผิดพลาดแบบนี้ให้น้อยที่สุด

ตัวอย่าง

แผนภาพด้านล่างแสดงความไม่แน่นอนที่เกิดจากการวัดความดันโลหิต Diastolic ในกลุ่มผู้ป่วย ซึ่งมีการแกว่งตัวตามธรรมชาติ (Chance) อยู่รอบ ๆค่า 90 mm.Hg. แต่สูงกว่าค่าจริง ที่ได้จากการวัดความดันโลหิตในหลอดเลือดแดงไป เฉลี่ยแล้วประมาณ 10 mm.Hg. (Bias)



โดยทั่วไปการศึกษาทางการแพทย์ให้ความสนใจ และต้องการวัดความแตกต่างตามความเป็นจริง ของค่าที่วัดได้ระหว่างบุคคล (Variation among patients) ที่เกิดเนื่องจากความแตกต่างทางชีวภาพ แพทย์จึงควรเข้าใจว่าค่าที่วัดได้ อาจเกิดความไม่แน่นอน หรือความคลาดเคลื่อนจากสาเหตุใดได้บ้าง ดังกล่าวข้างต้น เพื่อสามารถทำความเข้าใจข้อมูลได้อย่างถูกต้อง และสามารถออกแบบการวัดที่เหมาะสม เพื่อป้องกันหรือลดความคลาดเคลื่อนจากความเป็นจริงให้เหลือน้อยที่สุด

การกระจายของข้อมูล (Distribution)

ขั้นตอนแรกของการวิเคราะห์ข้อมูล คือการพรรณนา หรืออธิบายให้เห็นลักษณะของข้อมูลในภาพรวมว่ามีการมีลักษณะอย่างไร มีการกระจายตัวอย่างไร ซึ่งข้อมูลที่ได้จากการวัด และตัวแปรแต่ละประเภท จะมีวิธีการแสดงการกระจายของข้อมูลที่แตกต่างกันไป

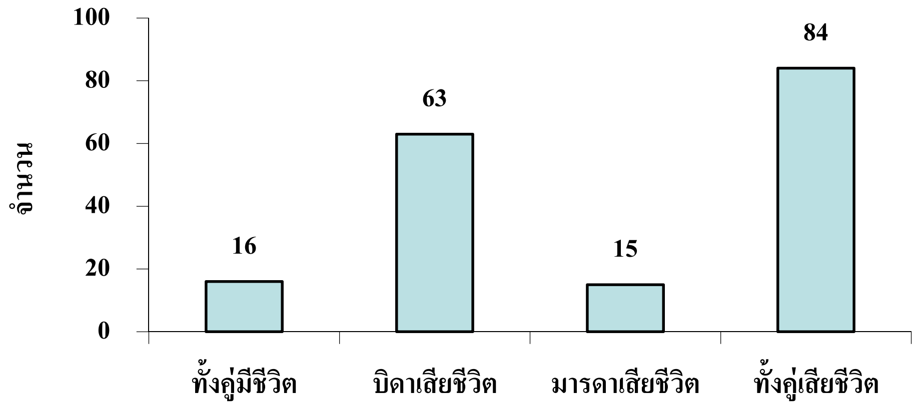
1. ***ข้อมูลบอกลักษณะ (Qualitative data)*** แสดงการกระจายของข้อมูลได้ โดยบอกจำนวน หรือสัดส่วน (ร้อยละ) ของประชากรแต่ละกลุ่ม อาจแสดงในลักษณะของการบรรยาย ตาราง หรือหากต้องการแสดงเป็นภาพ ก็อาจแสดงโดยใช้แผนภูมิแท่ง (Bar diagram) หรือแผนภูมิกง (Pie diagram)

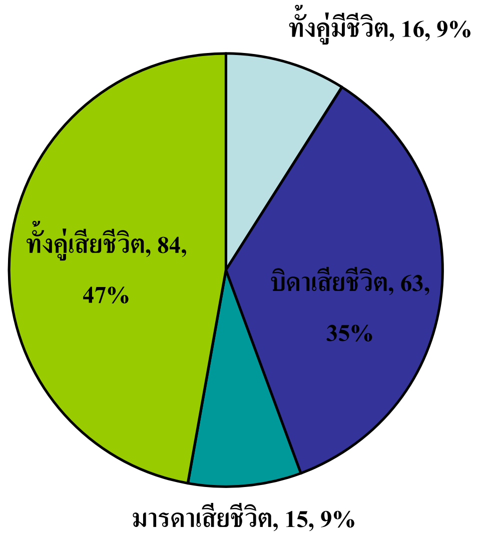
ตัวอย่าง

การมีชีวิตของบิดา มารดา ของเด็กที่ได้รับผลกระทบจากเอดส์

ลักษณะ	จำนวน	ร้อยละ
ทั้งคู่มีชีวิต	16	8.9
บิดาเสียชีวิต	63	35.4
มารดาเสียชีวิต	15	8.4
ทั้งคู่เสียชีวิต	84	47.2
รวม	178	100.0

การมีชีวิตของบิดา มารดา ของเด็กที่ได้รับผลกระทบจากเอดส์





2. ***ข้อมูลบอกปริมาณ (Quantitative data)*** การกระจายของข้อมูลบอกปริมาณ สามารถแสดงได้หลายแบบกว่าข้อมูลบอกคุณภาพ ได้แก่

2.1 ตารางแจกแจงความถี่ (Frequency table) เป็นตารางแสดงการกระจายของข้อมูล ด้วยการรายงานจำนวนในแต่ละระดับ หรือกลุ่มของตัวแปร เราอาจจัดกลุ่มของข้อมูลเป็นช่วง (Interval) โดยจัดข้อมูลที่มีค่าอยู่ในช่วงเดียวกันให้อยู่ในกลุ่มเดียวกัน อย่างไรก็ตาม เมื่อแบ่งเป็นกลุ่มแล้ว ข้อมูลทั้งหมดจะต้องจัดลงในกลุ่มใดกลุ่มหนึ่งได้ทั้งหมด และต้องจัดได้เพียงกลุ่มเดียวเท่านั้น

ตัวอย่าง

ตารางแจกแจงความถี่แสดงการกระจายของอายุของเด็กที่ได้รับผลกระทบจากเอดส์

อายุ	*ความถี่*	*ร้อยละ*	*ตำแหน่งเปอร์เซนไทล์*
6	3	1.5	1.5
7	6	2.9	4.4
8	9	4.4	8.7
9	10	4.9	13.6
10	17	8.3	21.8
11	27	13.1	35.0
12	31	15.0	50.0
13	28	13.6	63.6
14	22	10.7	74.3
15	34	16.5	90.8
16	12	5.8	96.6
17	6	2.9	99.5
18	1	.5	100.0
Total	206	100.0	

2.2 พิสัย (Range) หมายถึง ผลต่างของค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดหนึ่ง ๆ แม้จะคำนวณง่ายและความหมายเข้าใจง่าย แต่พิสัยมีที่ใช้น้อย เนื่องจากไม่ได้มาจากข้อมูลทุก ๆ ค่า แต่หามาจากค่าที่มากที่สุด และน้อยที่สุดเท่านั้น จึงไม่ได้บอกรายละเอียดเกี่ยวกับการกระจายตัวของข้อมูลอื่น ๆ

ตัวอย่าง

พิสัยของอายุเด็กที่ได้รับผลกระทบจากเอดส์

= ค่าสูงสุด – ค่าต่ำสุด

= 18 – 6 = 12 ปี

2.3 Interquartile range (IQR) คือค่าการกระจาย ซึ่งคำนวณได้จากผลต่างระหว่าง Q_3 กับ Q_1 ดังสูตรต่อไปนี้ IQR = $Q_3 - Q_1$ โดย Q_3 คือค่าที่เปอร์เซนต์ไทล์ที่ 75 และ Q_1 คือค่าที่เปอร์เซนต์ไทล์ที่ 25 IQR มีข้อดีกว่า range ตรงที่ไม่ sensitive ต่อข้อมูลที่มีการกระจายแบบไม่ปกติหรือข้อมูลที่มีค่าสุดโต่ง (outliers) IQR มักใช้แสดงคู่กับ median

2.3 ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation, SD) บอกถึงระยะห่างโดยเฉลี่ยของข้อมูลแต่ละค่าจากค่าเฉลี่ยเลขคณิต คำนวณจากรากที่สองของค่าเฉลี่ยกำลังสองของความแตกต่างของข้อมูลแต่ละตัวกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น ส่วนเบี่ยงเบนมาตรฐานเป็นค่าวัดการกระจายที่สำคัญ และใช้มากในการวิเคราะห์ทางสถิติ เพราะคำนวณจากตัวเลขทุกตัวในชุดข้อมูล จึงเป็นค่าที่ใช้บอกถึงการกระจายของข้อมูลได้ดีกว่าค่าพิสัย หากข้อมูลส่วนใหญ่อยู่ใกล้ค่าเฉลี่ยมาก ค่าเบี่ยงเบนมาตรฐานก็จะมีค่าน้อย ในทางกลับกัน ถ้าข้อมูลแต่ละจุดอยู่ห่างไกลจากค่าเฉลี่ยเป็นส่วนมาก ค่าเบี่ยงเบนมาตรฐานก็จะมีค่ามาก ค่า SD มักบอกคู่กับค่าเฉลี่ยเลขคณิต

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}}$$

<u>ตัวอย่าง</u>

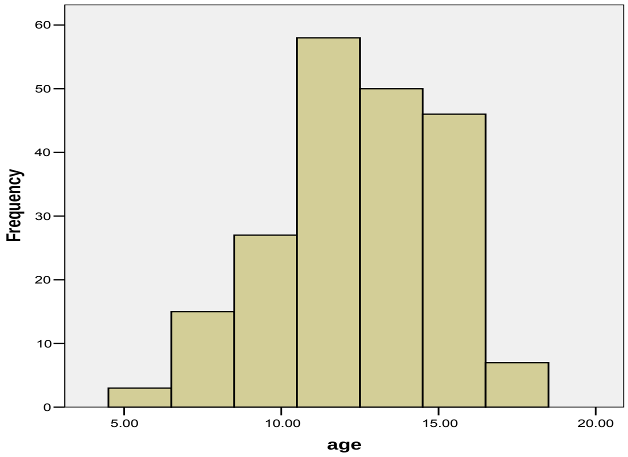
ส่วนเบี่ยงเบนมาตรฐานของอายุเด็กที่ได้รับผลกระทบจากเอดส์

= 2.57 ปี

2.4 ฮิสโตแกรม (Histogram) คือกราฟแสดงการแจกแจงความถี่ของข้อมูลบอกปริมาณ ที่มีลักษณะเป็นรูปสี่เหลี่ยมมุมฉากเรียงติดต่อกันจากค่าน้อยไปหามาก โดยความกว้างของรูปสี่เหลี่ยมมุมฉากแต่ละอันแทนความกว้างของแต่ละอันตรภาคชั้น ส่วนสูงของสี่เหลี่ยมมุมฉากแต่ละอันแทนสัดส่วนของกลุ่มตัวอย่างที่อยู่ในอันตรภาคชั้นนั้น ๆ เทียบกับกลุ่มตัวอย่างทั้งหมด และพื้นที่ของรูปสี่เหลี่ยมมุมฉากแต่ละรูปแทนจำนวนกลุ่มตัวอย่างที่อยู่ในแต่ละอันตรภาคชั้น

ตัวอย่าง

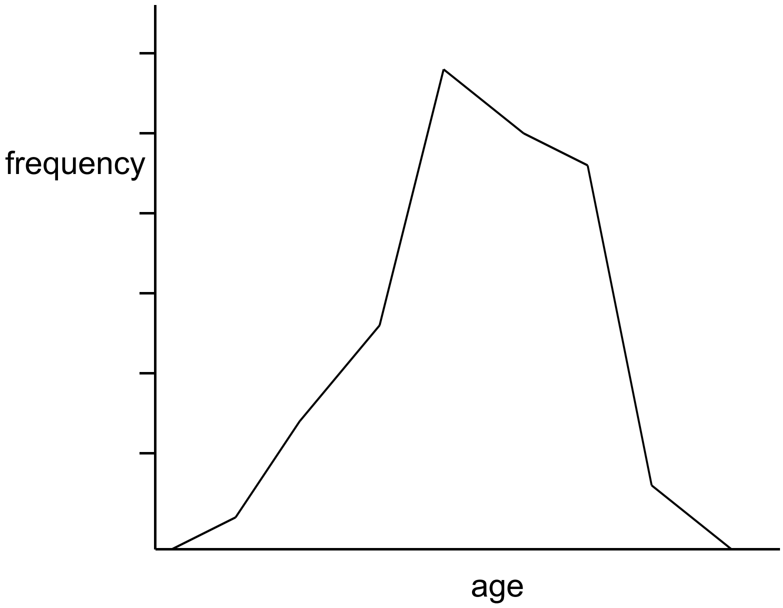
ฮิสโตแกรม แสดงการกระจายของอายุเด็กที่ได้รับผลกระทบจากเอดส์



2.5 รูปหลายเหลี่ยมของความถี่ (Frequency polygon) คือรูปหลายเหลี่ยมที่เกิดจาก การลากเส้นตรงโยงระหว่างจุดกึ่งกลางบนแท่งสี่เหลี่ยมมุมฉากที่อยู่ติดกันของฮิสโตแกรมของข้อมูลชุดนั้น ซึ่งแสดงการกระจายของข้อมูลด้วยภาพได้ดี และเป็นพื้นฐานของ Probability distribution ที่จะกล่าวถึงต่อไป

ตัวอย่าง

รูปหลายเหลี่ยมของความถี่ แสดงการกระจายของอายุเด็กที่ได้รับผลกระทบจากเอดส์



การวัดแนวโน้มเข้าสู่ส่วนกลาง (Central tendency)

ในการบรรยายลักษณะของชุดข้อมูลหนึ่ง ๆ เรามักหาค่ากลางมา 1 ค่า เพื่อเป็นตัวแทนของข้อมูลทั้งหมด เรียกว่าการวัดแนวโน้มเข้าสู่ส่วนกลาง ที่นิยมใช้กันมี 3 แบบ คือ

1. <u>ค่าเฉลี่ยเลขคณิต (Arithmetic mean)</u> คือผลเฉลี่ยของข้อมูลทุกตัวในข้อมูลชุดนั้น เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางที่รู้จักกันดีที่สุดและใช้มากที่สุดในทางสถิติ มักบอกร่วมกับค่า SD ค่าเฉลี่ยเลขคณิตใช้ได้เฉพาะกับข้อมูลบอกปริมาณ โดยสัญลักษณ์ที่ใช้แทนค่าเฉลี่ยของประชากร คือ μ

   <u>ตัวอย่าง</u>

   ค่าเฉลี่ยของอายุเด็กที่ได้รับผลกระทบจากเอดส์ (μ) = (6+6+6+7+7+….) / 206

   = 12.4 ปี

2. <u>มัธยฐาน (Median)</u> คือค่ากลางของข้อมูลเมื่อเรียงจากน้อยไปมากหรือจากมากไปน้อย ชุดข้อมูลที่มีจำนวนเป็นเลขคี่ มัธยฐาน คือ ข้อมูลที่อยู่ตรงกลาง ชุดข้อมูลที่มีจำนวนคู่ มัธยฐานจะเป็นค่าของข้อมูลสองตัวที่อยู่ตรงกลางรวมกันหารสอง Interquartile range เป็นการวัดการกระจายที่มักใช้ร่วมกับมัธยฐาน

   <u>ตัวอย่าง</u>

   มัธยฐานของอายุเด็กที่ได้รับผลกระทบจากเอดส์ = อายุกึ่งกลางเมื่อเรียงจากน้อยไปหามาก

   = (อายุลำดับที่ 103 + อายุลำดับที่ 104) / 2

   = (12 + 13) / 2 = 12.5 ปี

3. <u>ฐานนิยม (Mode)</u> คือค่าของข้อมูลที่มีความถี่มากที่สุด ข้อมูลชุดหนึ่ง ๆ อาจไม่มีฐานนิยมเลย เช่นกรณีที่ไม่มีข้อมูลที่ซ้ำกันเลย หรือมีฐานนิยมหลายค่าก็ได้ (ชุดข้อมูลที่มีฐานนิยม 2 ค่า เรียกว่า Bimodal) ฐานนิยมเป็นการวัดแนวโน้มเข้าสู่ส่วนกลางแบบเดียวที่สามารถใช้กับข้อมูลบอกคุณภาพได้

   <u>ตัวอย่าง</u>

   ฐานนิยมของอายุเด็กที่ได้รับผลกระทบจากเอดส์ = 15 ปี

   ฐานนิยมของการมีชีวิตของบิดา มารดา ของเด็กที่ได้รับผลกระทบจากเอดส์ได้แก่เด็กกลุ่มที่ทั้งบิดาและมารดาเสียชีวิต

ความสัมพันธ์ระหว่างการวัดแนวโน้มเข้าสู่ส่วนกลาง และลักษณะการกระจายของข้อมูล

หากเราสังเกต Histogram หรือ Frequency polygon ของข้อมูลบอกปริมาณส่วนใหญ่ จะเห็นว่ามีรูปร่างคล้ายเส้นโค้งยอดเดี่ยว มีลักษณะการกระจายได้ 3 แบบ คือ

1. โค้งปกติ (Normal, no skew) มีการกระจายของข้อมูลใกล้เคียงกันทั้งทางด้านน้อยและด้านมาก ไม่เอียงไปด้านใดด้านหนึ่ง การกระจายแบบนี้ ทั้งค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยม จะเท่ากัน หรือใกล้เคียงกัน
2. โค้งเบ้ซ้าย (Left skewed, negatively skewed) การกระจายของข้อมูลที่มีหางโค้งยาวไปทางด้านซ้าย แบบนี้ ฐานนิยมจะมีค่าสูงสุด รองลงมาเป็นมัธยฐาน ส่วนค่าเฉลี่ยมีค่าน้อยที่สุด ตัวอย่างชุดข้อมูลอายุเด็กที่ได้รับผลกระทบจากเอดส์ข้างต้น ก็มีลักษณะแบบนี้

3. โค้งเบ้ขวา (Right skewed, positively skewed) การกระจายของข้อมูลที่มีหางโค้งยาวไปทางด้านขวา แบบนี้ ค่าเฉลี่ยจะมีค่าสูงสุด รองลงมาเป็นมัธยฐาน ส่วนฐานนิยม มีค่าน้อยที่สุด

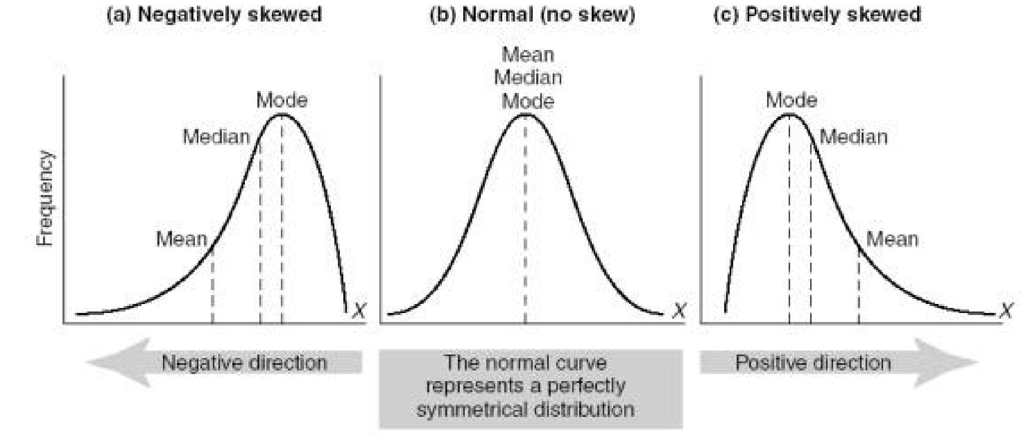


FIGURE 15.6 Examples of normal and skewed distributions

จะสังเกตได้ว่า กรณีที่โค้งมีการเบ้ (skewed) ไม่ว่าด้านใด ค่าฐานนิยมจะอยู่ที่ยอดโค้ง ส่วนค่า มัธยฐานจะอยู่ระหว่าง ค่าฐานนิยมและค่าเฉลี่ยเสมอ ค่ามัธยฐานจึงเป็นค่ากลางที่เหมาะสมกว่าค่าเฉลี่ยในการเป็นตัวแทนของข้อมูลในกรณีที่โค้งมีการเบ้

ตัวอย่าง

การกระจายของข้อมูลอายุเด็กที่ได้รับผลกระทบจากเอดส์มีลักษณะเป็นโค้งเบ้ซ้าย
ฐานนิยม (15 ปี) > มัธยฐาน (12.5 ปี) > ค่าเฉลี่ย (12 ปี)

ความน่าจะเป็น (Probability)

ทฤษฎีความน่าจะเป็น ๆ รากฐานของสถิติเชิงอนุมาน ตลอดจนการให้การวินิจฉัยและดูแลรักษาผู้ป่วย แพทย์อาจบอกญาติผู้ป่วยว่า ผู้ป่วยอาการหนักมาก โอกาสการรอดชีวิตเท่ากับ ห้าสิบ ห้าสิบ หมายถึงมีโอกาสรอดชีวิตร้อยละ 50 เป็นต้น ความน่าจะเป็นหมายถึงโอกาสการเกิดเหตุการณ์ที่สนใจจากเหตุการณ์ที่เป็นไปได้ทั้งหมด ความน่าจะเป็นจึงเป็นสัดส่วน และมีค่าได้ตั้งแต่ 0 (ไม่มีโอกาสเกิดเหตุการณ์นั้นเลย) จนถึง 1 (เกิดเหตุการณ์นั้นอย่างแน่นอน) และนิยมแทนด้วยอักษร P

$$P(E) = \frac{\text{number of times E \underline{occurs}}}{\text{number of times E \underline{can occur}}}$$

แนวคิดเกี่ยวกับความน่าจะเป็นอาจแบ่งได้เป็น 2 ลักษณะคือ Classical Probability ภายใต้แนวคิดนี้ โอกาสการเกิดเหตุการณ์ในการเล่นเกมส์ หรือการเล่นพนัน ที่เป็นลักษณะของการเสี่ยงดวง จะใช้อธิบายได้ดี ตัวอย่างเช่น ความน่าจะเป็นที่จะทอยลูกเต๋าหนึ่งครั้งแล้วออกหน้า 1 เท่ากับ 1 ใน 6 เสมอ (หากลูกเต๋าไม่มีการถ่วง) การคำนวณโอกาสความน่าจะเป็นแบบนี้เป็นการใช้เหตุผลในลักษณะนามธรรม เนื่องจากไม่เหตุผลให้เชื่อเป็นอย่างอื่นและไม่จำเป็นต้องทำการทดลอง แนวคิดเกี่ยวกับความน่าจะเป็นอีกลักษณะหนึ่ง คือ Relative

Frequency Probability ภายใต้แนวคิดนี้ การคาดความน่าจะเป็นของการเกิดเหตุการณ์ คำนวณจากจากการนับความถี่ของเหตุการณ์ที่เราสนใจที่เกิดขึ้นในกระบวนการซ้ำ ๆ ในจำนวนครั้งที่มากพอ

ตัวอย่าง

จากชุดข้อมูลเด็กที่ได้รับผลกระทบจากเอดส์ ความน่าจะเป็นของการสุ่มหาเด็กที่ได้รับผลกระทบจากเอดส์ 1 คน แล้วได้เด็กอายุ 15 ปี

= 34 / 206 = 0.165

ในการให้การวินิจฉัยโรคและดูแลรักษาผู้ป่วย แพทย์ใช้ผลการศึกษาทางระบาดวิทยาในกลุ่มผู้ป่วยจำนวนมากที่มีลักษณะคล้ายคลึงกัน และนำผลการศึกษาที่แสดงถึงความน่าจะเป็นในการเกิดเหตุการณ์ทางคลินิกต่าง ๆ ว่ามีมากน้อยเพียงใด มาใช้ประกอบการตัดสินใจให้การวินิจฉัยและดูแลรักษาผู้ป่วยแต่ละราย ในทางกลับกัน เราสามารถใช้หลักความน่าจะเป็นในการสรุปผลการวิจัยซึ่งทำในกลุ่มตัวอย่างจำนวนหนึ่ง ไปยังระดับประชากรได้ (สถิติเชิงอนุมาน)

Odds

Odds หมายถึงอัตราส่วนของโอกาสการเกิดเหตุการณ์ที่สนใจ หารด้วยโอกาสการเกิดเหตุการณ์ที่ไม่สนใจ

$$\text{Odds (E)} = \frac{\text{Number of times E occurs}}{\text{Number of times E cannot occur}}$$

โอกาสการเกิดเหตุการณ์ที่สนใจในทางการแพทย์มักน้อย (Rare event) เช่นโอกาสการเจ็บป่วยหรือเสียชีวิตด้วยโรคใดโรคหนึ่งในระยะเวลาที่กำหนด กรณีนี้ Probability และ Odds จะมีค่าใกล้เคียงกัน นอกจากนั้น Odds ยังมีคุณสมบัติพิเศษที่ดีในทางสถิติ บ่อยครั้ง Odds จึงถูกใช้แทน Probability ในการศึกษาทางระบาดวิทยา Odds มีความสัมพันธ์กับ Probability ดังสมการต่อไปนี้

$$\text{Odds (E)} = \frac{P}{1 - P}$$

ตัวอย่าง

Odds ของการสุ่มหาเด็กที่ได้รับผลกระทบจากเอดส์ 1 คน แล้วได้เด็กอายุ 15 ปี

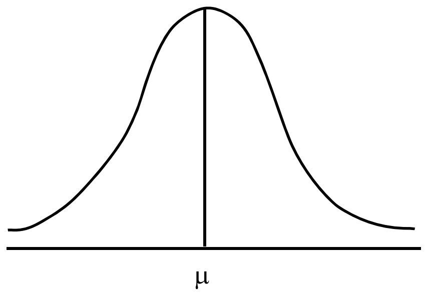
= 34 / (206 – 34) = 0.198

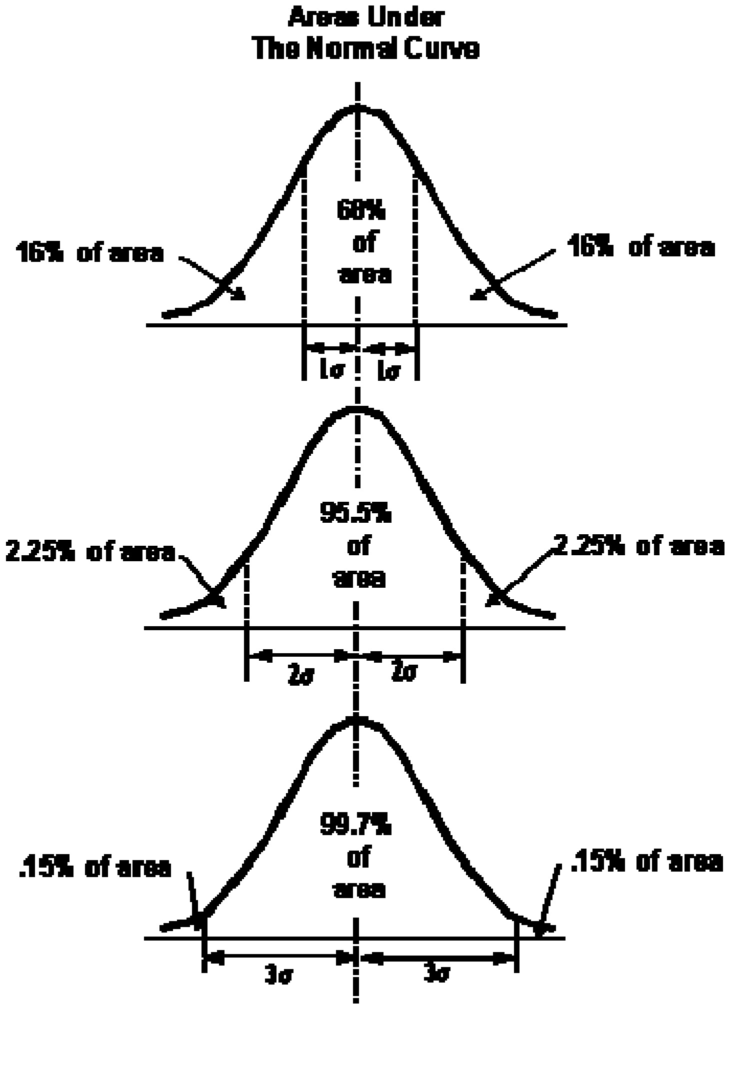
= 0.165 / (1 – 0.165) = 0.198

รูปแบบการกระจายของความน่าจะเป็น (Probability distribution)

หากเราแปลงแกน Y ของ Frequency polygon ให้เป็นร้อยละของความน่าจะเป็น จะพบว่าพื้นที่ใต้กราฟจะเท่ากับโอกาสความน่าจะเป็นของการเกิดเหตุการณ์ทั้งหมด คือเท่ากับ 1 เรียกว่ากราฟแสดง Probability distribution วิธีการทางสถิติส่วนใหญ่ตั้งอยู่บนสมมติฐานที่ว่า ข้อมูลที่ได้จากกลุ่มตัวอย่าง เป็นส่วนหนึ่งของข้อมูลในประชากรที่มีลักษณะการกระจายที่แน่นอนตายตัว เรียกว่า Theoretical probability distribution ซึ่งที่ใช้บ่อย และควรทำความรู้จัก ได้แก่

1. *Normal distribution* เป็น Theoretical probability distribution ที่สำคัญที่สุด และใช้บ่อยที่สุดในทางสถิติ ใช้กับข้อมูลบอกปริมาณ มีลักษณะเป็นรูประฆังคว่ำ สมมาตร มีจุดโค้งสูงสุดเพียงจุดเดียว ปลายโค้งทั้งสองข้างไม่จรดกับฐาน รูปร่างของกราฟทั้งหมดสามารถอธิบายได้ด้วยค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของประชากร





2. *Binomial distribution* เป็น Theoretical probability distribution ที่ใช้สำหรับวิเคราะห์ข้อมูลที่เกิดจากกระบวนการเชิงสุ่ม ที่ผลมีความเป็นไปได้เพียงสองอย่าง เช่นเมื่อเราต้องการหาความน่าจะเป็นของการรักษาหายหรือไม่หายจากการป่วยเป็นโรค ลักษณะของการกระจายของข้อมูลจะถูกกำหนดด้วยตัวแปร 2 ตัว คือขนาดของกลุ่มตัวอย่าง (n) และความน่าจะเป็นของเหตุการณ์ที่เราสนใจ (p) เมื่อขนาดของกลุ่มตัวอย่างน้อยการกระจายจะเป็นโค้งไม่สมมาตร แต่เมื่อขนาดของกลุ่มตัวอย่างมากขึ้นการกระจายจะมีลักษณะสมมาตร และมีลักษณะเหมือน normal distribution มากขึ้น

   ตัวอย่าง

   คนเรามีโอกาสมีเลือดกรุ๊ป B ร้อยละ 8 ถ้าสุ่มคนจากประชากรทั่วไปมา 10 คน เราสามารถใช้ Binomial distribution ประมาณความน่าจะเป็นที่จะพบคนมีเลือดกรุ๊ป B จำนวน 0, 1, 2, …, 10 คน ได้

3. *Poisson distribution* เป็น Theoretical probability distribution ที่ใช้สำหรับวิเคราะห์ข้อมูลที่เป็นจำนวนนับ (Count) หรืออัตรา (Rate) รูปร่างของการกระจายกำหนดได้ด้วยจำนวนเฉลี่ยของการเกิดเหตุการณ์ที่สนใจ (Lambda, λ) เพียงค่าเดียว และเช่นเดียวกับ Binomial distribution เมื่อค่าเฉลี่ยมีค่าน้อย กราฟจะมีรูปร่างไม่สมมาตร แต่เมื่อค่าเฉลี่ยของการเกิดเหตุการณ์สูงขึ้น กราฟจะมีลักษณะเหมือน normal distribution มากขึ้นเช่นเดียวกัน

   ตัวอย่าง

   ถ้าเรามีข้อมูลในอดีตว่า โดยเฉลี่ยแล้ว ห้องฉุกเฉินของโรงพยาบาลชุมชนแห่งหนึ่ง จะมีผู้ป่วยบาดเจ็บจากอุบัติเหตุจราจรเข้ามารับการรักษาวันละ 3.6 ราย เราสามารถจะคาดประมาณโอกาสที่ วันพรุ่งนี้จะมีผู้ป่วย 0, 1, 2, 3, 4, 5, ราย หรือมากกว่านั้น โดยใช้ Poisson distribution

****

เอกสารอ้างอิง

1. Altman GD. Practical statistics for medical research. London: Chapman & Hall; 1991.
2. Knapp GR, Miller MC. Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins; 1992.
3. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
4. Danial WW. Biostatistics Basic concepts and Methodology for the Health Sciences. 9th Edition. Wiley@Sons, Inc, 2010
5. Jaccard J, Becker A M. Statistics for the behavioral sciences. 5^{th} edition. Wadsworth; 2010.

****

แบบฝึกหัด

1. จำนวน CD4+ cell จากเลือดที่เจาะในเวลากลางคืนมักสูงกว่าเวลากลางวัน ปรากฏการณ์ดังกล่าวเข้าได้กับความไม่แน่นอนของการวัดลักษณะใด

A. Bias

B. Chance

C. Random error

D. Temporal variation

E. Measurement error

2. รายได้เฉลี่ยต่อเดือนของครอบครัวมักกระจุกตัวกันในช่วงรายได้ต่ำ และกระจายตัวห่างกันออกไปในช่วงรายได้สูง ๆ ข้อใดถูกเกี่ยวกับค่าแนวโน้มเข้าสู่ส่วนกลางของข้อมูลลักษณะนี้

A. mean > mode > median

B. mean > median > mode

C. mode > mean > median

D. mode > median > mean

E. median > mode > mean

3. แพทย์ผู้เชี่ยวชาญโรคหัวใจในเด็ก รวบรวมจำนวนผู้ป่วยที่มี Heart murmur intensity ตั้งแต่ grade1 ถึง grade 6 ได้ดังนี้ 32, 28, 13, 10, 3, 1

แผนภูมิใด เหมาะสมสำหรับใช้นำเสนอการกระจายของข้อมูลดังกล่าว

A. Histogram

B. Line graph

C. Bar diagram

D. Frequency table

E. Frequency polygon

4. ถ้า Odds ที่คู่สมรสจะมีบุตรเป็น Hb Bart's hydrops fetalis ในกรณีที่ทั้งสามี และภรรยา เป็นพาหะของ α-Thalassemia เท่ากับ 0.33 จงคำนวณหา Probability ของการเกิดเหตุการณ์ดังกล่าว

A. 0.16

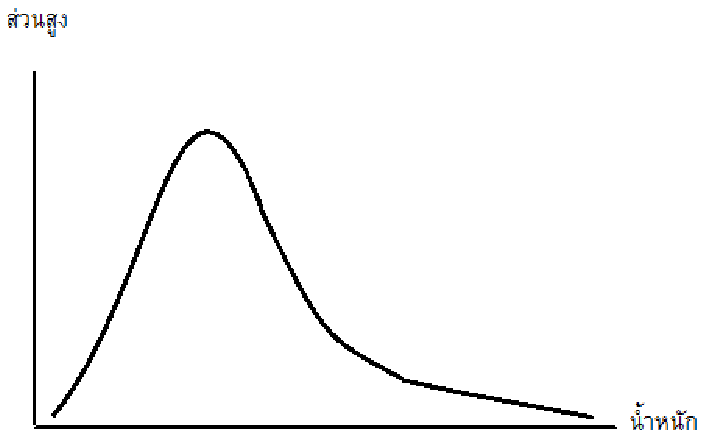
B. 0.25

C. 0.33

D. 0.50

E. 0.67

5. น้ำหนักเด็ก ป.4 ของเด็กนักเรียนห้องหนึ่งมีการกระจายดังภาพ ต้องใช้ค่าใดเป็นค่ามาตรฐาน



A. Mean กับ SD

B. Mean กับ range

C. Mode กับ range

D. Median กับ IQR

E. Median กับ SD

1.11.1.3 Disease prevalence and incidence

การวัดการเกิดโรคมีความสำคัญอย่างยิ่งต่อการวินิจฉัยและรักษาโรคของแพทย์ การศึกษาทางการแพทย์ส่วนใหญ่สนใจศึกษาเหตุการณ์ที่มีลักษณะเป็น Categorical dichotomous variables คือตัวแปรบอกลักษณะที่มีผลลัพธ์ได้เพียงสองอย่าง เช่นป่วยหรือไม่ป่วย เสียชีวิตหรือรอดชีวิต ผลตรวจทางห้องปฏิบัติการปกติหรือผิดปกติ การวัดตัวแปรลักษณะนี้ทำได้โดยนำจำนวนผู้ที่มีลักษณะที่สนใจเป็นตัวตั้ง และหารด้วยประชากรกลุ่มเสี่ยงต่อการเกิดภาวะนั้น ที่สำคัญ มี 2 ลักษณะ คือ ความชุก (Prevalence) และอุบัติการณ์ (Incidence)

ความชุก (Prevalence)

ความชุกหมายถึงสัดส่วนของประชากรที่กำลังเป็นโรคหรือมีลักษณะที่สนใจ เทียบกับประชากรทั้งหมด การคำนวณหาความชุกทำได้โดย นำจำนวนผู้ป่วยหรือผู้มีลักษณะที่สนใจ ณ เวลาที่ทำการศึกษา เป็นตัวตั้งหารด้วยจำนวนประชากรทั้งหมด ความชุกมักได้จากการศึกษาทางระบาดวิทยาเชิงสำรวจแบบภาคตัดขวาง (Cross-sectional study) ความชุกนิยมบอกเป็นร้อยละ โดยนำค่าสัดส่วนที่คำนวณได้คูณด้วย 100

P = C / N

เมื่อ P = ความชุก

C = จำนวนผู้ป่วย หรือผู้ที่มีลักษณะที่สนใจศึกษา ณ เวลาดังกล่าว (ทั้งรายเก่า และรายใหม่)

N = จำนวนประชากรขณะนั้น

<u>ตัวอย่าง</u> การสำรวจในผู้ที่มีอายุตั้งแต่ 40 ปี ขึ้นไป จำนวน 300 คน ในหมู่บ้านแห่งหนึ่ง ณ วันที่ 30 มิถุนายน 2552 พบว่ามีผู้ที่มีความดันโลหิตสูงกว่าเกณฑ์ปกติจำนวน 90 คน

P = C / N

= 90 / 300 = 0.30 = 30%

สรุปได้ว่า ความชุกของภาวะความดันโลหิตสูงในผู้ที่มีอายุตั้งแต่ 40 ปี ขึ้นไป ในหมู่บ้านแห่งนี้ ณ วันที่ทำการศึกษา เท่ากับร้อยละ 30

ความชุกบอกขนาดของความเจ็บป่วยหรือการลักษณะของกลุ่มตัวอย่างที่สนใจศึกษา ณ เวลานั้น ๆเหมาะสำหรับใช้อธิบายปัญหาการเจ็บป่วยด้วยโรคเรื้อรัง หรือปัจจัยเสี่ยงด้านสุขภาพที่ไม่เปลี่ยนแปลงบ่อย ความชุกยังบอกถึงขนาดของความต้องการบริการทางด้านสาธารณสุขของปัญหาสุขภาพดังกล่าว ช่วยในการวางแผนเกี่ยวกับการจัดการกำลังคน และขนาดของบริการอนามัยสำหรับชุมชน อย่างไรก็ตามความชุกไม่เหมาะสำหรับใช้ศึกษาขนาดของปัญหาโรคเฉียบพลันที่มีระยะเวลาเจ็บป่วยสั้น หรือโรคที่เกิดการระบาดเป็นครั้งคราว เนื่องจากหากไม่ได้ทำการศึกษาในช่วงที่มีการระบาดของโรคอาจตรวจไม่พบผู้ป่วยเลย (ความชุกเท่ากับศูนย์) นอกจากนั้น ความชุกยังมีข้อจำกัด คือไม่สามารถบอกได้ว่าผู้ป่วยที่พบเริ่มป่วยตั้งแต่เมื่อใด และป่วยมานานเท่าไหร่แล้ว

ความชุกยังอาจแบ่งได้เป็น 2 ลักษณะย่อย ๆคือ

1. <u>*Point prevalence*</u> หมายถึงความชุกของโรคในประชากร ณ จุดใดจุดหนึ่งของเวลา เหมาะสำหรับอธิบายขนาดของโรค หรือปัญหาสุขภาพเรื้อรังที่คงที่ เป็นแล้วเป็นเลยจะไม่หายจากภาวะนั้นอีก เช่น เบาหวาน ความดันโลหิตสูง การติดเชื้อเอชไอวี หรือการเป็นพาหะของไวรัสตับอักเสบบี เป็นต้น point prevalence มักได้จากการศึกษาเชิงสำรวจ (Survey study) ตัวอย่างเกี่ยวกับการศึกษาความชุกของภาวะความดันโลหิตสูงในชุมชนข้างต้นจัดเป็น point prevalence ปกติแล้วเมื่อกล่าวถึงความชุกเป็นการทั่วไป ไม่ได้ระบุประเภทอย่างเฉพาะเจาะจง จะหมายถึง point prevalence เสมอ

2. <u>*Period prevalence*</u> หมายถึงความชุกของโรคในประชากรภายในช่วงเวลาที่กำหนด เหมาะสำหรับอธิบายภาวการณ์เจ็บป่วยเรื้อรังแต่มีอาการเป็น ๆ หาย ๆ ไม่ได้เป็นตลอดเวลา เช่นอาการโรคระบบกระดูกและกล้ามเนื้อปวดศีรษะไมเกรน อาการทางจิต ประสาท เนื่องจากการวัดความชุกวิธีนี้จะช่วยอธิบายสภาพการณ์ของปัญหาได้ดีกว่า point prevalence

   <u>ตัวอย่าง</u> การสำรวจในประชากรหาผู้ที่มีอาการปวดหลัง ณ วันใดวันหนึ่ง อาจพบ point prevalence เพียงร้อยละ 2 แต่หากรวมผู้ที่มีอาการปวดหลังอย่างน้อย 1 ครั้ง ในรอบ 1 ปี อาจมี period prevalence ถึงร้อยละ 25 จะเห็นได้ว่า ในกรณีนี้ period prevalence อธิบายลักษณะปัญหาให้ดีกว่า point prevalence

อุบัติการณ์ (Incidence)

บางครั้งเรียกอัตราอุบัติการณ์ หมายถึงการเกิดผู้ป่วยรายใหม่ในประชากรกลุ่มเสี่ยงในช่วงเวลาที่ทำการศึกษา (หรือสังเกต) ในทางระบาดวิทยาถือว่าอุบัติการณ์มีประโยชน์กว่าความชุกในแง่การอธิบายความสัมพันธ์เชิงสาเหตุระหว่างปัจจัยเสี่ยงกับการเกิดโรค และสามารถบอกถึงปัญหาการแพร่ระบาดของโรคในชุมชนว่ารุนแรงเพียงใด รวมถึงใช้ประเมินมาตรการป้องกันควบคุมโรคว่าได้ผลหรือไม่ดีกว่าความชุก ค่าอุบัติการณ์มักได้จากการศึกษาทางระบาดวิทยาแบบไปข้างหน้า (Cohort study)

การหาอุบัติการณ์ของโรคทำได้ยากกว่าการหาความชุก เนื่องจากต้องใช้เวลานานในการติดตามเก็บข้อมูลและสังเกตการเกิดโรค อุบัติการณ์ แบ่งออกได้เป็น 2 แบบ คือ

1. <u>Cumulative incidence</u> (CI) คือโอกาสความน่าจะเป็นในการเกิดความเจ็บป่วยเป็นโรคที่สนใจของประชากรแต่ละคนโดยเฉลี่ยในช่วงระยะเวลาที่กำหนด เรียกอีกอย่างหนึ่งว่าความเสี่ยง (Risk) ต่อการเกิดเป็นโรค ภายใต้สมมติฐานที่ว่าผู้วิจัยสามารถติดตามประชากรทุกคนได้ตลอดช่วงเวลาที่ทำการศึกษา หรือจนกว่าจะป่วยเป็นโรค CI หาได้โดยนำจำนวนผู้ป่วยรายใหม่ที่เกิดขึ้นในช่วงเวลาที่ทำการศึกษาเป็นตัวตั้งหารด้วยจำนวนประชากรกลุ่มเสี่ยง ณ เวลาเริ่มต้นการศึกษา ซึ่งนับเฉพาะผู้ที่ยังมีโอกาสเจ็บป่วยด้วยโรคนั้น ๆ สำหรับโรคที่เมื่อเคยป่วยแล้วไม่สามารถเป็นซ้ำได้อีก จะหักจำนวนผู้ที่เคยป่วยด้วยโรคนั้นออกออกจากตัวหาร CI เป็นสัดส่วน (Proportion) มีค่าได้ตั้งแต่ 0 – 1 และต้องระบุระยะเวลาที่ทำการศึกษาประกอบด้วยเสมอ

$$CI = I / P$$

เมื่อ CI = Cumulative incidence

I = จำนวนผู้ป่วยรายใหม่

P = ประชากรกลุ่มเสี่ยง ณ เวลาเริ่มต้นการศึกษา

<u>ตัวอย่าง</u> การสำรวจในผู้ที่มีอายุตั้งแต่ 40 ปี ขึ้นไป จำนวน 300 คน ในหมู่บ้านแห่งหนึ่ง ณ วันที่ 30 มิถุนายน 2552 พบว่ามีผู้ที่มีความดันโลหิตสูงกว่าเกณฑ์ปกติจำนวน 90 คน หนึ่งปีต่อมา ณ วันที่ 30

มิถุนายน 2553 ผู้วิจัยกลับไปสำรวจในประชากรกลุ่มเดิม พบผู้ที่มีความดันโลหิตสูงกว่าเกณฑ์ปกติ จำนวน 99 คน

$$CI = I / P$$

$$= (99 – 90) / (300 – 90) \qquad = 9 / 210 \qquad = \text{ร้อยละ } 4.3$$

สรุปได้ว่า โดยเฉลี่ยแล้ว ผู้ที่มีอายุตั้งแต่ 40 ปี ขึ้นไป ที่อาศัยในหมู่บ้านแห่งนี้ มีโอกาสป่วยเป็นความดันโลหิตสูงร้อยละ 4.3 ต่อปี

2. Incidence density (ID) ชื่ออื่น ๆได้แก่ Incidence rate และ Hazard rate คืออัตรา หรือความเร็วชั่วขณะเวลาหนึ่งในการเกิดความเจ็บป่วยเป็นโรคที่สนใจของประชากรศึกษา ID จัดเป็นอัตราที่แท้จริงทางระบาดวิทยา การประมาณค่าอุบัติการณ์ลักษณะนี้ในประชากรมทำได้ 2 ลักษณะ คือ

2.1. *Incidence rate based on aggregate data* คือการประมาณอัตราป่วยหรืออัตราตาย ด้วยโรคที่สนใจศึกษา โดยใช้ข้อมูลระยะเวลาที่ติดตามในระดับประชากร มักเป็นการศึกษาที่มีการกำหนดขอบเขตงานวิจัยเป็นพื้นที่ทางภูมิศาสตร์ ตัวตั้งจะเป็นจำนวนเหตุการณ์ที่สนใจที่เกิดขึ้นใหม่ในช่วงเวลาที่ศึกษา ส่วนตัวหารเป็นจำนวนประชากรเฉลี่ยของช่วงเวลาหนึ่ง เช่นจำนวนประชากรกลางปี (ณ วันที่ 1 กรกฎาคม)

$$IR = I / P$$

เมื่อ IR = Incidence rate

I = จำนวนผู้ป่วยรายใหม่

P = จำนวนประชากรเฉลี่ย

ตัวอย่าง ระหว่างปี 2553 อำเภอ ก. มีผู้ป่วยวัณโรครายใหม่ 85 ราย ในขณะที่มีจำนวนประชากรกลางปี 53,000 คน อุบัติการณ์ (incidence rate) ของการเกิดวัณโรคเท่ากับ

$$IR = I / P$$

$$= 85 / 53,000$$

$$= 0.0016$$

$$= 1.6 / 1,000 \text{ ต่อปี}$$

สรุปได้ว่า โดยเฉลี่ยแล้ว อำเภอ ก. มีอัตราการป่วยเป็นวัณโรครายใหม่ประมาณ 1.6 รายต่อประชากร 1,000 คน ต่อปี

2.2. *Incidence density based on individual data* คือการประมาณอัตราป่วยหรืออัตราตาย ด้วยโรคที่สนใจ โดยพิจารณาลงไปในละเอียดมากขึ้น เนื่องจากการศึกษาวิจัยในชีวิตจริงประชากรบางคนอาจไม่อยู่ให้เราติดตามตลอดช่วงระยะเวลาที่กำหนด หรือเสียชีวิตจากภาวะอื่น ๆ ที่ไม่เกี่ยวข้องกับปัญหาสุขภาพที่สนใจ การคำนวณหาอุบัติการณ์แบบนี้ใช้ข้อมูลระยะเวลาที่ติดตามในระดับบุคคล โดยตัวตั้งจะเป็นจำนวนเหตุการณ์ที่สนใจที่เกิดขึ้นใหม่ในช่วงเวลาที่ศึกษา เช่นเดียวกับการหาอุบัติการณ์แบบอื่น ๆ ส่วนตัวหารจะเป็นระยะเวลาทั้งหมดรวมกันในการสังเกตประชากรกลุ่มเสี่ยงแต่ละคน เรียกว่า person-time at risk จะเห็นว่าระยะเวลาในการติดตามอาสาสมัครแต่ละคนไม่จำเป็นต้องเท่ากัน และสามารถนำมาบวกรวมกันได้ อาสาสมัครแต่ละคนไม่จำเป็นต้องถูกเริ่มสังเกตพร้อมกัน และไม่จำเป็นต้องออกจากโครงการพร้อมกัน

ID = I / PT

เมื่อ ID = Incidence density

I = จำนวนผู้ป่วยรายใหม่

PT = ระยะเวลาทั้งหมดในการสังเกตประชากรกลุ่มเสี่ยง

Incidence density แบบนี้ถือเป็นอัตราที่แท้จริง มีหน่วยเป็น $time^{-1}$ มีค่าได้ตั้งแต่ 0 - ∞ สามารถอธิบายเหตุการณ์ที่เกิดไม่บ่อย (rare events) ได้ดี เนื่องจากมีความสะดวกในการเลือกปรับทอนหน่วยให้เหมาะสม โดยความหมายไม่เปลี่ยนไป

<u>ตัวอย่าง</u> การศึกษาพบผู้ป่วยรายใหม่ 12 ราย จากการติดตามอาสาสมัครทั้งสิ้น 500 person-days

ID = 12 / 500 person-days

= 0.024 / person-day

= 2.4 / 100 person-days

= 8.76 / person-year

= 876 / 100 person-years

จะเห็นได้ว่า ทุกคำตอบมีความหมายเหมือนกันแล้วแต่ผู้วิจัยจะเลือกนำเสนอด้วยหน่วยไหนที่สอดคล้องกับสถานการณ์ และเข้าใจได้ง่ายที่สุด

ความสัมพันธ์ระหว่างความชุก และอุบัติการณ์

ความชุกมีความสัมพันธ์กับอุบัติการณ์ของโรคและระยะเวลาการเจ็บป่วยด้วยโรคนั้น ๆ ว่าการเจ็บป่วยแต่ละครั้งกินระยะเวลานานแค่ไหน ในภาวะสมดุลเมื่ออุบัติการณ์ของโรค (Incidence) และระยะเวลาการเจ็บป่วย (Duration) ของโรคคงที่ และความชุกของโรคไม่สูง (น้อยกว่าหรือเท่ากับ 0.05)

Point prevalence $\approx$ Incidence X Duration

<u>ตัวอย่าง</u>

อุบัติการณ์ของโรคเบาหวานคงที่ ๆ 1% ต่อปี และผู้ป่วยมีอายุหลังการวินิจฉัยเฉลี่ยประมาณ 15 ปี ความชุกจะเท่ากับ 0.01 X 15 = 15%

หมายความว่า ถ้าเราทำการสำรวจหาความชุกของการป่วยเป็นโรคเบาหวานในประชากรนี้ไม่ว่าทำเมื่อไหร่ก็จะพบว่าเท่ากับร้อยละ 15 เสมอ ถ้าสถานการณ์ไม่เปลี่ยนแปลง

****

เอกสารอ้างอิง

1. Knapp GR, Miller MC. Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins; 1992.
2. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
3. Szklo M, Nieto JF. Epidemiology beyond the basics. Sudbury; Jones and Bartlett Publishers: 2004.

****

แบบฝึกหัด

1. การศึกษาแบบ Survey study ให้ค่าการวัดการเกิดโรคใดเสมอ

A. Cumulative incidence

B. Incidence rate

C. Incidence density

D. Point prevalence

E. Period prevalence

2. ในสภาวะสมดุล ถ้าอุบัติการณ์ของโรคความดันโลหิตสูงในผู้ที่มีอายุ 40 ปี ขึ้นไป เท่ากับร้อยละ 2 ต่อปี อายุขัยเฉลี่ยเท่ากับ 80 ปี และความชุกของความดันโลหิตสูงในประชากรดังกล่าวเท่ากับร้อยละ 30 โดยเฉลี่ยแล้วผู้ป่วยแต่ละรายจะป่วยด้วยโรคนี้นานกี่ปี

A. 40

B. 30

C. 20

D. 15

E. 10

3. ผู้ติดเชื้อเอชไอวี 200 ราย ติดตามดูแลรักษา 1 ปี เสียชีวิตจากภาวะแทรกซ้อนของการติดเชื้อ 40 ราย ขาดการรักษา 100 ราย เสียชีวิตจากอุบัติเหตุ 10 ราย ที่เหลือยังสุขภาพดี ณ เวลา 1 ปี ค่า population-time เท่ากับ 48,000 คน-วัน จงคำนวณหา Incidence density ของการเสียชีวิตจากภาวะแทรกซ้อนของการติดเชื้อ

A. 20% ต่อ 1 ปี

B. 25% ต่อ 1 ปี

C. 30 คน ต่อ 100 คน-ปี

C. 35 คน ต่อ 100 คน-ปี

C. 40 คน ต่อ 100 คน-ปี

4. อุบัติการณ์ของโรคไข้หวัดนก บ่งบอกถึง

A. จำนวนผู้ป่วย

B. ความรุนแรงของโรค

C. ความสำเร็จในการรักษาผู้ป่วย

D. ผลของการควบคุมการแพร่ระบาดของโรค

E. งบประมาณที่ควรจัดเตรียมเพื่อรองรับการรักษา

1.11.1.4 Disease outcomes

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

การดำเนินโรค

เมื่อเกิดการเจ็บป่วยเป็นโรคใดโรคหนึ่ง ผู้ป่วยและญาติจะมีคำถามมากมายที่อยากถาม และได้คำตอบจากแพทย์ผู้ดูแล เช่น โรคที่เป็นมีความรุนแรงแค่ไหน จะเสียชีวิตหรือมีความพิการหรือไม่ จะมีความเจ็บปวดทรมานอย่างไร ซึ่งอันที่จริงก็คือ ต้องการทราบว่าลักษณะการดำเนินโรคตลอดจนผลของการเกิดโรคจะเป็นไปอย่างไรนั่นเอง เราอาจบรรยายลักษณะการดำเนินโรคได้ 2 แบบ คือ

1. *การดำเนินโรคทางคลินิก (Clinical course)* หมายถึงลักษณะการดำเนินโรคในผู้ป่วยที่ได้รับการดูแลรักษาทางการแพทย์ ซึ่งการรักษาที่ได้รับอาจไม่เหมือนกัน และนำมาซึ่งผลการดำเนินโรคที่แตกต่างกันได้ โดยทั่วไป ผู้ป่วยด้วยโรคที่มีอาการและอาการแสดงชัดเจน มักจะได้รับการดูแลรักษาเสมอ
2. *การดำเนินโรคตามธรรมชาติ (Natural history of disease)* หมายถึง ลักษณะการดำเนินโรคโดยเฉลี่ยหากปล่อยให้เป็นไปตามธรรมชาติ โดยไม่มีการดูแลรักษาทางการแพทย์เข้าไปแทรกแซง ข้อมูลได้จากการสังเกตผู้ป่วยที่ไม่ได้รับการรักษาพยาบาล ซึ่งอาจรวบรวมไว้ในอดีต สมัยที่ยังไม่มีการรักษาโรคนั้น ๆ หรือเป็นโรคที่ไม่ได้รับการวินิจฉัย การเจ็บป่วยที่ไม่แสดงอาการ หรือที่มีอาการเล็กน้อยไม่จำเป็นต้องได้รับการรักษา ตัวอย่างเช่นภาวะซึด โรคทางจิต ประสาท เป็นต้น

ผลลัพธ์ของการดำเนินโรค (Health outcomes)

หมายถึงเหตุการณ์ทางคลินิก (ด้านไม่ดี) ที่เป็นผลมาจากการดำเนินโรค และ/หรือ การรักษาพยาบาลที่ผู้ป่วยตระหนักถึงความสำคัญ ขณะที่แพทย์ก็ให้ความสนใจศึกษาเพื่อทำความเข้าใจ อธิบายปรากฏการณ์ตลอดจนคาดการว่าจะเกิดหรือไม่ อาจแบ่งได้ดังนี้

The 5 Ds

Death	คือการเสียชีวิต ถือว่าเป็นผลลัพธ์ที่ไม่ดี หากเกิดก่อนเวลาอันควร
Disease	กลุ่มอาการ อาการแสดง หรือความผิดปกติของผลตรวจทางห้องปฏิบัติการ
Discomfort	อาการไม่สุขสบาย เช่นความเจ็บปวด คลื่นไส้ เวียนศีรษะ
Disability	ความพิการ ซึ่งทำให้ไม่สามารถทำกิจกรรม หรือดำเนินชีวิตตามปกติได้
Dissatisfaction	ปฏิกิริยาด้านลบทางอารมณ์ต่อการเจ็บป่วย และกระบวนการรักษาพยาบาล เช่น ความโศกเศร้า ความไม่พอใจ หรือความโกรธ

ผลลัพธ์ของการดำเนินโรคบางอย่าง เป็นเพียงปรากฏการณ์ทางชีววิทยาที่ไม่ส่งผลต่อการรับรู้ของผู้ป่วย เช่นผลการวิเคราะห์ระดับสารเคมีในเลือด แพทย์จะให้ความสำคัญน้อยกว่า

การพยากรณ์โรค (Prognosis)

หมายถึง การคาดการณ์ถึงเหตุการณ์ทางคลินิกในอนาคต หลังจากมีการป่วยเป็นโรคและได้รับการวินิจฉัยแล้ว ผู้ป่วยด้วยภาวะเดียวกันอาจมีการพยากรณ์โรคต่างกัน ตามการมีหรือไม่มีปัจจัยพยากรณ์โรค (Prognostic factors) ตัวอย่างเช่น ผู้ป่วยไส้ติ่งอักเสบที่มีอาการไข้ร่วมด้วย ณ เวลาที่มาพบแพทย์ จะมีการพยากรณ์โรคที่เลวกว่าผู้ป่วยไส้ติ่งอักเสบที่ไม่มีอาการไข้ ในที่นี้อาการไข้ถือเป็นปัจจัยพยากรณ์โรคด้านไม่ดี (Poor prognostic factor) การจะทราบว่าปัจจัยใดเป็นปัจจัยพยากรณ์โรคด้านดีหรือไม่ดี ต้องทำการศึกษาเปรียบเทียบผลการดำเนินโรคในผู้ป่วยที่มี และไม่มีปัจจัยนั้น เรียกว่าการศึกษาปัจจัยพยากรณ์โรค (Prognostic study) การวิจัยแบบนี้ มีลักษณะคล้ายกับ Cohort study

ตัวชี้วัดการพยากรณ์โรค

แพทย์ศึกษาโอกาสการเกิดผลลัพธ์การดำเนินโรคทางคลินิกที่สำคัญในผู้ป่วยลักษณะคล้ายคลึงกันจำนวนมาก เพื่อนำมาพยากรณ์โรคในผู้ป่วยแต่ละราย เช่น

1. *การรอดชีพ (Survival rate, Survival proportion)* หมายถึงสัดส่วน (ร้อยละ) ของผู้ป่วยที่ยังมีชีวิตรอด ณ เวลาต่าง ๆ หลังการได้รับการวินิจฉัย นิยมใช้กับโรคเรื้อรังที่มีความรุนแรง และมีโอกาสเสียชีวิตสูง เช่น โรคมะเร็ง ซึ่งนิยมบอกการรอดชีพ ณ เวลา 5 ปี (5-year survival)
2. *อัตราป่วยตาย (Case fatality)* หมายถึง สัดส่วนของผู้เสียชีวิตจากโรคใดโรคหนึ่ง เทียบกับจำนวนผู้ที่ป่วยด้วยโรคนั้นทั้งหมด เป็นตัวชี้วัดที่ใช้บอกความร้ายแรงของโรค มักใช้ในโรคติดเชื้อเฉียบพลัน เช่น อัตราการเสียชีวิตจากการติดเชื้อไข้หวัดใหญ่สายพันธ์ใหม่ 2009 H1N1
3. *อัตราการตอบสนองต่อการรักษา (Response)* หมายถึงสัดส่วนของผู้ป่วยที่แสดงถึงผลการรักษาที่ดีขึ้นภายหลังได้รับการรักษา เช่น ระดับน้ำตาลในเลือดที่ลดลงในผู้ป่วยเบาหวานที่รับประทานยา
4. *การหายจากโรค (Remission)* หมายถึงสัดส่วนของผู้ป่วยที่เข้าสู่ระยะของการดำเนินโรคที่ไม่สามารถตรวจพบโรคได้ เช่นการตรวจไม่พบไวรัสในกระแสเลือดของผู้ป่วยเอดส์ภายหลังได้รับยาต้านไวรัสไประยะเวลาหนึ่ง
5. *การเป็นกลับซ้ำ (Recurrence)* หมายถึงสัดส่วนของผู้ป่วยที่กลับมามีอาการหรือเป็นโรคอีก ภายหลังจากที่หายจากโรคไประยะหนึ่ง เช่นการกลับเป็นซ้ำของโรคเริมที่อวัยวะเพศ

ดัชนีอนามัยที่สำคัญทางด้านสาธารณสุข (Major health indices)

ดัชนีอนามัยหมายถึงข้อมูลเชิงปริมาณที่ใช้บ่งชี้ภาวะสุขภาพอนามัยของชุมชน ดัชนีบางตัวเป็นมาตรฐานที่ใช้ร่วมกันในระดับนานาชาติ เพื่อเปรียบเทียบและติดตามความเปลี่ยนแปลงตลอดจนความก้าวหน้าของงานทางด้านสาธารณสุขของประเทศ ในที่นี้จะยกตัวอย่าง ดัชนีอนามัยที่สำคัญ และใช้บ่อยในทางสาธารณสุข ซึ่งได้แก่

ดัชนีอนามัยเกี่ยวกับการตาย (Mortality rates)

1. *อัตราตายอย่างหยาบ (Crude mortality rate, Crude death rate, All causes mortality rate)* หมายถึงสัดส่วนของจำนวนคนตายด้วยทุก ๆ สาเหตุ ต่อจำนวนประชากร 1,000 คน ต่อปี ส่วนใหญ่ตัวหารจะใช้ประชากรคาดประมาณกลางปี
2. *อัตราตายจำเพาะ (Specific mortality rate)* หมายถึง สัดส่วนของจำนวนคนตายในเฉพาะคนกลุ่มใดกลุ่มหนึ่ง ต่อจำนวนประชากร เช่น อัตราตายในเพศชาย หรืออัตราตายในกลุ่มคนอายุช่วงใดช่วงหนึ่ง
3. *อัตราทารกตาย (Infant mortality rate)* หมายถึง สัดส่วนของจำนวนทารกอายุต่ำกว่า 1 ปี ที่เสียชีวิต ต่อเด็กเกิดมีชีพพันคน เป็นดัชนีมาตรฐานสากลทางด้านการแพทย์และการสาธารณสุขที่มีความสำคัญมากที่สุด ใช้บอกระดับการพัฒนาทางด้านสาธารณสุข และการพัฒนาประเทศ อัตราตายของเด็กอายุต่ำกว่า 5 ปี (Under 5 mortality rate) ก็เป็นดัชนีอนามัยอีกตัวหนึ่งที่มีลักษณะใกล้เคียงกัน และใช้บ่อย
4. *อัตรามารดาตาย (Maternal mortality rate)* หมายถึงสัดส่วนของจำนวนมารดาตายเนื่องจากการตั้งครรภ์ การคลอด และการอยู่ไฟ ต่อเด็กเกิดมีชีพแสนคน

ดัชนีอนามัยเกี่ยวกับการป่วย (Morbidity rates)

ได้แก่ อัตราอุบัติการณ์ของโรค (Incidence) และอัตราความชุกของโรค (Prevalence) โดยระบุเป็นการเฉพาะไปแต่ละโรค เช่นอุบัติการณ์ของไข้เลือดออกต่อประชากรแสนคนในปี 2556 หรือความชุกของผู้ป่วยเบาหวานในประชากรที่มีอายุเกิน 40 ปี ในปี 2555 เป็นต้น

Standardization of rates

การเปรียบเทียบอัตราการเจ็บป่วยหรืออัตราการเสียชีวิตของประชากร 2 กลุ่ม บางครั้งไม่สามารถนำอัตราอย่างหยาบ (Crude rates) มาเปรียบเทียบกันได้ เพราะอาจสรุปผิดพลาดเนื่องจากปัจจัยที่แตกต่างกันของประชากรแต่ละกลุ่มซึ่งมากวน (Confound) ปัจจัยกวนที่สำคัญที่สุดคืออายุ เนื่องจากผู้ที่มีอายุมากมักมีโอกาสป่วยและการเสียชีวิตมากกว่าคนอายุน้อย การเปรียบเทียบอัตราตายอย่างหยาบของประชากรที่มีการกระจายของอายุหรือมีอายุเฉลี่ยน้อยกว่า กับอัตราตายอย่างหยาบของประชากรที่มีการกระจายของอายุหรืออายุเฉลี่ยมากกว่าจึงไม่ยุติธรรม

วิธีการคำนวณเพื่อให้สามารถเปรียบเทียบอัตราข้ามกลุ่มประชากรได้เรียกว่าการทำ Standardization บางครั้งเรียกว่า Adjustment มีสองวิธีคือ Direct standardization และ Indirect standardization อัตราที่ได้หลังทำ Standardization of rates แล้วจะสามารถนำมาเปรียบเทียบกันได้

****

เอกสารอ้างอิง

1. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.
2. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.

****

แบบฝึกหัด

1. อัตราใด บอกถึงความร้ายแรงของโรคได้ดีที่สุด

   A. Response

   B. Remission

   C. Recurrence

D. Case-fatality

E. 5 year survival

2. ดัชนีอนามัยเกี่ยวกับการตายใด นิยมคำนวณต่อแสนประชากร

A. All causes mortality rate

B. Maternal mortality rate

C. Under 5 mortality rate

D. Specific mortality rate

E. Crude death rate

3. ถ้าต้องการทราบว่าบุคลากรทางการแพทย์เสียชีวิตจากโรคมะเร็งมากกว่าประชาชนทั่วไปมากเท่าใด ต้องใช้ตัววัดตามข้อใด

A. Number of cancer death

B. Crude mortality rate

C. Age specific rate

D. Case fatality rate

E. Standardization of rate

4. ถ้าผู้ป่วยเบาหวานทุกๆ 500 ราย มีภาวะแทรกซ้อน Diabetic ketoacidosis 1 ราย และ 1 ใน 5 ของจำนวนนี้เสียชีวิต ให้คำนวณ Case fatality rate ของ Diabetic ketoacidosis

A. 0.02

B. 0.04

C. 0.2

D. 0.4

E. 1

1.11.1.5 Associations

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

สำหรับการศึกษาเชิงปริมาณ “ความสัมพันธ์” โดยทั่วไปหมายถึงความสัมพันธ์ทางสถิติระหว่างตัวแปรต้น (Independent variable) กับตัวแปรตาม (Dependent variable) ในการศึกษาเชิงวิเคราะห์ (Analytical studies) ตัวอย่างเช่นอายุที่เพิ่มขึ้นมีความสัมพันธ์กับความหนาแน่นของเนื้อกระดูกที่ลดลง ในที่นี้อายุเป็นตัวแปรต้น ขณะที่ความหนาแน่นของเนื้อกระดูกเป็นตัวแปรตาม โดยการที่ตัวแปรทั้งสองมีความสัมพันธ์กันนั้น ไม่จำเป็นเสมอไป ว่าตัวแปรต้นจะต้องเป็นสาเหตุของตัวแปรตาม

ความเสี่ยงต่อการเกิดโรค (Risk)

ระบาดวิทยา ให้ความสำคัญกับการหาความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงกับความเสี่ยงต่อการเกิดโรค (Risk) หากคนที่มีปัจจัยเสี่ยงป่วยเป็นโรคมากกว่าผู้ที่ไม่มีปัจจัยเสี่ยง ก็หมายถึงการมีปัจจัยเสี่ยงนั้นมีความสัมพันธ์กับโอกาสการเกิดโรค พิจารณาตัวอย่างข้างต้น การมีอายุที่เพิ่มขึ้นเป็นปัจจัยเสี่ยง (Risk factor) ของการมีความหนาแน่นของเนื้อกระดูกที่ลดลง ปัจจัยเสี่ยงที่ก่อให้เกิดโรคอาจแบ่งออกได้เป็นหลายลักษณะ เช่นปัจจัยเสี่ยงทางพันธุกรรม เชื้อโรค สารพิษ ลักษณะทางสังคม เศรษฐกิจ และปัจจัยเสี่ยงที่เป็นพฤติกรรมสุขภาพ

อย่างไรก็ตามหากเป็นการศึกษาเกี่ยวกับการป้องกันโรค ซึ่งผลที่พึงประสงค์คือการไม่ป่วยเป็นโรค หรือการศึกษาเกี่ยวกับการรักษาโรคในผู้ป่วย ในกรณีนี้ตัวแปรต้นจะไม่ถือเป็นปัจจัยเสี่ยง เนื่องจากไม่ได้ทำให้มีความเสี่ยงต่อการเกิดโรคมากขึ้น แต่กลับกันคือทำให้โอกาสในการเกิดโรคลดลงหรือการผลการดำเนินโรคด้านไม่ดีลดลด ซึ่งหากพบความสัมพันธ์ทางสถิติตัวแปรต้นจะกลายเป็นปัจจัยป้องกันโรค (Protective factors) เช่นวัคซีน เป็นต้น

ลักษณะของการมีปัจจัยเสี่ยง (Exposure to risk factors)

การ Expose ต่อปัจจัยเสี่ยงต่อการเกิดโรคอาจเป็นในลักษณะมีครั้งเดียว เช่นการรับประทานอาหารปนเปื้อนเชื้อโรคในงานเลี้ยง หรือมีปัจจัยเสี่ยงติดต่อกันเวลานาน เช่นการสูบบุหรี่และการมีภาวะความดันโลหิตสูง การวัดการมีปัจจัยเสี่ยงจึงมีได้หลายลักษณะ เช่นเคยมี ปริมาณสูงสุด ปริมาณรวม ระยะเวลาที่มี ฯลฯ ขึ้นอยู่กับองค์ความรู้ว่าปัจจัยเสี่ยงดังกล่าวอาจส่งผลต่อสุขภาพและต่อลักษณะของการเจ็บป่วยอย่างไร

เนื่องจากในความเป็นจริง การมีปัจจัยเสี่ยงไม่ได้นำไปสู่การเกิดโรคเสมอไป นอกจากนั้นการเกิดโรคส่วนใหญ่ต้องอาศัยเวลาไม่ได้เกิดทันทีตามหลังการมีปัจจัยเสี่ยง การจะสังเกตได้ว่าปัจจัยเสี่ยงใดมีความสัมพันธ์ หรือก่อให้เกิดโรคใดหรือไม่จึงมีความยุ่งยากซับซ้อน บางครั้งต้องใช้เวลานาน หรือต้องสังเกตประชากรจำนวนมากกว่าที่นักระบาดวิทยา หรือแพทย์จะสงสัยว่าปัจจัยเสี่ยงนั้น ๆ อาจเป็นสาเหตุของโรค และทำการศึกษาเพื่อหาความสัมพันธ์ดังกล่าวว่ามีจริงหรือไม่

ประโยชน์ของการทราบความเสี่ยงต่อการเกิดโรค

1. *ใช้คาดการณ์โอกาสการเกิดโรค* เราสามารถนำขนาดของความเสี่ยงต่อการเกิดโรคเมื่อมีปัจจัยเสี่ยงที่ได้จากการศึกษาในระดับประชากร มาใช้ในการพยากรณ์การเกิดโรคในระดับบุคคลที่มีปัจจัยเสี่ยงคล้ายคลึงกันได้ อย่างไรก็ตามการที่บุคคลมีปัจจัยเสี่ยงที่มีความสำคัญต่อการเกิดโรค ก็ไม่ได้หมายความว่าคน ๆ นั้น จะมีโอกาสสูงในการเกิดโรค เช่น สมมติว่าคนที่สูบบุหรี่หนักมีโอกาสเกิดมะเร็งปอดสูงกว่าผู้ไม่สูบบุหรี่ 20 เท่า ซึ่งดูเหมือนจะมีโอกาสป่วยเป็นมะเร็งปอดมาก แต่จริง ๆ แล้วคน ๆ นี้ หากคำนวณทางสถิติ เขาจะมีโอกาสป่วยเป็นมะเร็งปอดเพียง 1 ใน 100 ภายในระยะ 10 ปี

2. *ใช้ช่วยในการวินิจฉัยโรค* เมื่อมีผู้ป่วยที่ยังไม่ทราบว่าป่วยเป็นโรคอะไรมารับการรักษา ประวัติการมีปัจจัยเสี่ยงที่สำคัญ สามารถใช้ประกอบกับอาการและอาการแสดง ซึ่งช่วยให้แพทย์ให้การวินิจฉัยโรคได้ดีขึ้น

3. *ใช้ในการป้องกันควบคุมโรค* เมื่อเราทราบว่าอะไรเป็นปัจจัยเสี่ยงของการเกิดโรค การกำจัดปัจจัยเสี่ยงนั้นออกไปทั้งในระดับประชากรหรือระดับบุคคล จึงเป็นการป้องกันโรค ในทางระบาดวิทยาบางครั้งเราสามารถป้องกันควบคุมโรคอย่างได้ผลโดยการนำปัจจัยเสี่ยงออกไปก่อนที่จะทราบสาเหตุจริงของโรคเสียอีก เช่นกรณีของการระบาดของโรคเอดส์ในช่วงแรก ซึ่งยังไม่มีการค้นพบเชื้อไวรัสเอชไอวี ซึ่งเป็นสาเหตุที่แท้จริงของโรค ศูนย์ควบคุมโรคแห่งชาติของสหรัฐอเมริกาได้ออกประกาศให้ชายรักชายใช้ถุงยางอนามัยเมื่อมีเพศสัมพันธ์ เนื่องจากผลการศึกษาทางระบาดวิทยาชี้ว่าการร่วมเพศโดยไม่ป้องกันในคนกลุ่มนี้เป็นปัจจัยเสี่ยงของการเกิดโรค ซึ่งช่วยป้องกันการติดเชื้อได้ก่อนที่จะทราบสาเหตุที่แท้จริงของโรค

4. *ใช้หาสาเหตุของโรค* การศึกษาทางระบาดวิทยาเพื่อหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงที่สงสัยว่าอาจเป็นสาเหตุของโรคกับการเกิดโรค เป็นจุดเริ่มต้นที่สำคัญของการหาสาเหตุที่แท้จริงของโรค อย่างไรก็ตาม ปัจจัยเสี่ยงนั้น ๆ ในท้ายที่สุดอาจเป็นสาเหตุของโรคหรือไม่ก็ได้ ขึ้นอยู่กับธรรมชาติของการเกิดโรคและความสัมพันธ์ ตลอดจนความถูกต้องของวิธีการศึกษาและวิเคราะห์ข้อมูล

ตาราง 2 X 2 (2 by 2 Table)

ตาราง 2 X 2 ทางระบาดวิทยา เป็นตารางพื้นฐานที่ใช้บ่อยในการศึกษาทางระบาดวิทยา มักแสดงจำนวนประชากรที่มีหรือไม่มีโรคไว้ในแนวสดมภ์ (Column) และจำนวนประชากรที่มีหรือไม่มีปัจจัยเสี่ยงต่อการเกิดโรคไว้ในแนวแถว (Row) ตารางนี้ใช้ในการคำนวณหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงกับการเกิดโรคที่จะกล่าวถึงต่อไป ตาราง 2 X 2 ทางระบาดวิทยาถือเป็น Contingency table หมายถึงกลุ่มตัวอย่างแต่ละคนต้องถูกจัดให้อยู่ในช่องใดช่องหนึ่ง (a หรือ b หรือ c หรือ d) เพียงช่องเดียวเท่านั้น แต่ละช่องแสดงจำนวนกลุ่มตัวอย่างที่มีลักษณะต่าง ๆ ได้แก่

a = จำนวนผู้ที่มีปัจจัยเสี่ยงและป่วยเป็นโรค
b = จำนวนผู้ที่มีปัจจัยเสี่ยงแต่ไม่ป่วยเป็นโรค
c = จำนวนผู้ไม่มีปัจจัยเสี่ยงแต่ป่วยเป็นโรค และ
d = จำนวนผู้ไม่มีปัจจัยเสี่ยง และไม่ป่วยเป็นโรค

โดยจำนวนรวมของแต่ละแนวสดมภ์ แนวแถว และจำนวนรวมทั้งหมดต้องสอดคล้องกัน เช่น จำนวนรวมของผู้เป็นโรคทั้งหมด ต้องเท่ากับ a + c และ จำนวนรวมของผู้ไม่มีปัจจัยเสี่ยงต้องเท่ากับ c + d เป็นต้น

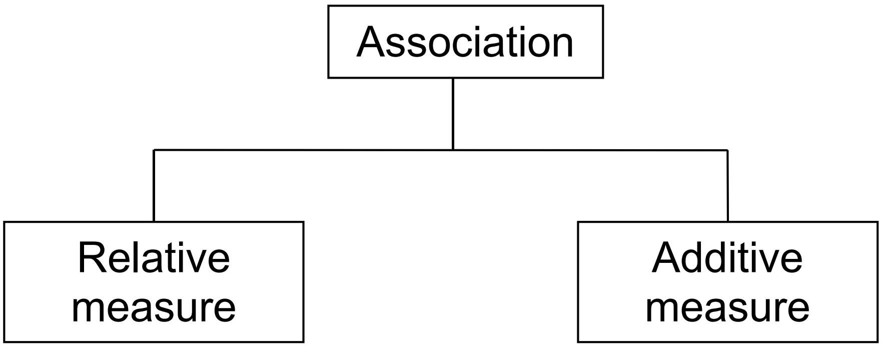
ตาราง 2 X 2 ทางระบาดวิทยา

ปัจจัยเสี่ยง	โรค (Disease)	รวม (Total)

(Exposure factor)	มี (Present)	ไม่มี (Absent)	
มี (Present)	a	b	a + b
ไม่มี (Absent)	c	d	c + d
รวม (Total)	a + c	b + d	a + b + c + d

ประเภทของความสัมพันธ์

วิธีการวัดความสัมพันธ์ (Measure of Association) ระหว่างการมีปัจจัยเสี่ยงกับการเกิดโรค อาจแบ่งได้เป็น 2 ลักษณะ คือ Relative measure (Multiplicative) และ Additive measure (Health impact)



Relative measure of association

หรืออีกชื่อหนึ่ง Multiplicative measure of association เป็นเทคนิคการหาความสัมพันธ์ในสเกลคูณหาร เป็นระดับการหาความสัมพันธ์ที่ใช้บ่อยในทางระบาดวิทยา เนื่องจากใช้อธิบายปรากฏการณ์ต่าง ๆ ได้ดี สามารถใช้ model ทางสถิติมาวิเคราะห์ข้อมูลเพิ่มเติม ได้ผลลัพธ์ที่สามารถเข้าใจได้ง่าย โดยเฉพาะอย่างยิ่งสำหรับการใช้อธิบายความเสี่ยงต่อการเกิดโรคที่เพิ่มขึ้นหรือลดลงในระดับบุคคล Multiplicative measure of association ในทางระบาดวิทยาที่ใช้บ่อยมี 2 แบบ คือ Relative risk และ Odds ratio

1. <u>Relative risk (RR)</u>

โดยทั่วไปเป็นการหาอัตราส่วน (Ratio) ระหว่างอุบัติการณ์การเกิดโรค (Incidence) ในกลุ่มประชากรที่มีปัจจัยเสี่ยงต่อการเกิดโรค หารด้วยอุบัติการณ์ในกลุ่มประชากรที่ไม่มีปัจจัยนั้น

Relative risk (RR) = Ie / Io

Ie = อุบัติการณ์ของโรคในกลุ่มประชากรที่มีปัจจัยเสี่ยงนั้น

Io = อุบัติการณ์ของโรคในกลุ่มประชากรที่ไม่มีปัจจัยเสี่ยงนั้น

หรือจากตาราง2 X 2

RR = (a / a + b) / (c / c + d)

RR เป็นชื่อที่ใช้เรียกการหาความสัมพันธ์ลักษณะนี้อย่างเป็นการทั่วไป เมื่อหา RR จากการศึกษาทางระบาดวิทยารูปแบบเฉพาะ RR จะมีชื่อเรียกที่แตกต่างกันออกไป แต่ความหมาย และวิธีการคำนวณยังคงเดิม RR มีค่าได้ตั้งแต่ 0 ถึง ∞ หาก RR มีค่าเท่ากับ 1 หมายถึง ปัจจัยนั้นไม่มีความสัมพันธ์กับการเกิดโรค RR ที่มากกว่า 1 หมายถึงการปัจจัยนั้นเป็นปัจจัยเสี่ยงต่อการเกิดโรค คือผู้มีปัจจัยนั้นจะมีโอกาสเสี่ยงต่อการเจ็บป่วยสูงกว่าผู้ไม่มีปัจจัยเสี่ยง เช่นผู้สูบบุหรี่มีโอกาสเสี่ยงต่อการป่วยเป็นมะเร็งปอดมากกว่าผู้ไม่สูบบุหรี่ ยิ่ง RR สูงกว่า 1 มากขึ้นเท่าใด ปัจจัยนั้นยิ่งมีความความสัมพันธ์กับโอกาสการเกิดโรคเป็นจำนวนเท่ามากขึ้นเท่านั้น ในทางตรงข้ามหาก RR ที่คำนวณได้มีค่าน้อยกว่า 1 หมายถึง การมีปัจจัยนั้นจะช่วยป้องกันโรค เช่นผู้ที่เคยฉีดวัคซีนป้องกันไข้หวัดใหญ่ เมื่อฤดูการระบาดของโรคมาถึงก็จะมีโอกาสป่วยเป็นโรคน้อยกว่าผู้ที่ไม่ได้ฉีด เป็นต้น

ตารางแสดงรูปแบบ และชื่อเรียกของ RR

รูปแบบการศึกษา	**ชื่อเรียก**	**การคำนวณ**
Cohort study of risk	Relative risk	Cumulative incidence among exposed ÷ Cumulative incidence among non-exposed
Cohort study of person-time	Incidence rate ratio	Incidence rate among exposed ÷ Incidence rate among non-expose
Cross-sectional study	Prevalence rate ratio	Prevalence among exposed ÷ Prevalence among non-exposed

2. <u>Odds ratio (OR)</u>

เนื่องจากในการศึกษาแบบ Case-control studies เราไม่สามารถหาอุบัติการณ์และความชุกของโรคจึงทำให้ไม่สามารถคำนวณหา RR ได้ วิชาระบาดวิทยาจึงได้พัฒนาวิธีการหาความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงกับการเกิดโรคที่มีค่าที่ใกล้เคียงกับ RR ขึ้น เรียกว่า Odds ratio (OR) ซึ่งคือ อัตราส่วนของ Odds ของการมีปัจจัยเสี่ยงในผู้ป่วย (Case) หารด้วย Odds ของการมีปัจจัยเสี่ยงในผู้ไม่ป่วย (Control)

ตาราง 2 X 2 ทางระบาดวิทยา กรณีการศึกษาแบบ Case-control study

ปัจจัยเสี่ยง (Exposure factor)	โรค (Disease)		รวม (Total)
	ผู้ป่วย (Case)	ผู้ไม่ป่วย (Control)	
มี (Present)	a	b	a + b
ไม่มี (Absent)	c	d	c + d
รวม (Total)	a + c	b + d	a + b + c + d

Odds ratio (OR) = E(case) / E(control)

E(case) = Odds ของการมีปัจจัยเสี่ยงใน case

E(control) = Odds ของการมีปัจจัยเสี่ยงใน control

หรือจากตาราง 2 X 2

$$OR = \frac{[a / (a + c) \div c / (a + c)]}{[b / (b + d) \div d / (b + d)]}$$

$$= (a / c) \div (b / d)$$

$$= (a)(d) / (b)(c)$$

เช่นเดียวกับ RR หาก OR ที่มีค่าเท่ากับ 1 หมายถึงปัจจัยนั้นไม่มีส่วนสัมพันธ์กับการเกิดโรค OR ยิ่งมากกว่า 1 เท่าใด ปัจจัยนั้นยิ่งมีส่วนสัมพันธ์กับโอกาสการเกิดโรคมากขึ้นเท่านั้น และ OR ที่มีค่าน้อยกว่า 1 หมายถึง ปัจจัยนั้นช่วยป้องกันโรค ในการศึกษาโรคที่พบไม่บ่อย (Rare diseases) OR จะมีค่าใกล้เคียงหรือเท่ากับ RR

Additive measure of association

หรืออีกชื่อหนึ่ง Health impact เป็นเทคนิคการหาความสัมพันธ์ระหว่างปัจจัยเสี่ยงกับการเกิดโรค ในสเกลบวก ลบ ที่ใช้มากในทางระบาดวิทยามี 4 แบบ คือ

1. *Attributable risk (AR)* ชื่ออื่น Risk different หรือ Excess risk คำนวณได้โดยการนำอุบัติการณ์ของโรคในผู้ที่มีปัจจัยเสี่ยงมาลบกับอุบัติการณ์ของโรคในผู้ที่ไม่มีปัจจัยเสี่ยง อีกความหมายหนึ่งคือ ความเสี่ยงต่อการป่วยเป็นโรคที่เพิ่มขึ้นมา (สเกลบวกลบ) เมื่อมีปัจจัยเสี่ยงนั้นนั่นเอง

   AR = Incidence (exposed) – Incidence (unexposed)

2. *Attributable risk percent (%AR)* สัดส่วนของการเกิดโรคเนื่องจากการมีปัจจัยเสี่ยงนั้นในผู้ที่มีปัจจัยเสี่ยง แล้วนำมาคูณด้วย 100 เพื่อทำให้เป็นร้อยละ หรืออีกนัยหนึ่ง ในจำนวนผู้ที่มีปัจจัยเสี่ยงและป่วยด้วยโรคนั้น 100 ราย มีกี่รายที่ป่วยเนื่องจากปัจจัยเสี่ยงนั้น ๆ (ผู้ป่วยส่วนที่เหลือป่วยเนื่องจากสาเหตุอื่น ๆ)

   %AR = (AR / Incidence in exposed) * 100

3. Population attributable risk (PAR) หมายถึงความเสี่ยงต่อการป่วยเป็นโรคที่เพิ่มขึ้นในประชากรเนื่องจากปัจจัยเสี่ยงนั้น ๆ คำนวณได้โดย

   PAR = Incidence in population – Incidence in unexposed

   หรือ

   = (AR) * (proportion of exposed in the population)

4. Population attributable risk percent (%PAR) หมายถึงสัดส่วนของการเกิดโรคในประชากรเนื่องจากการมีปัจจัยเสี่ยง แล้วนำมาคูณด้วย 100 เพื่อทำให้เป็นร้อยละ หรืออีกนัยหนึ่ง หมายถึงในประชากรที่ป่วยด้วยโรคนั้น 100 คน มีกี่คนที่ป่วยเนื่องจากปัจจัยเสี่ยงนั้น ๆ (ส่วนที่เหลือป่วยเนื่องจากสาเหตุอื่นๆ) ซึ่งจะเท่ากับสัดส่วนของการป่วยด้วยโรคนั้นที่จะลดลงในประชากร หากสามารถกำจัดปัจจัยเสี่ยงนั้นออกจากประชากรได้ทั้งหมด

   %PAR = (PAR / Incidence in population) * 100

Additive measure of association ใช้มากในการกำหนดนโยบายทางด้านสาธารณสุข เนื่องจากสามารถใช้คำนวณหาขนาดของผลกระทบด้านสุขภาพที่เกิดจากปัจจัยเสี่ยงนั้น ๆ ในระดับประชากร

สาเหตุ (Causation)

เป้าหมายสูงสุดของการวิจัยทางการแพทย์และการสาธารณสุขที่เป็นการศึกษาเชิงวิเคราะห์ คือการหาข้อสรุปให้ได้ว่า ปัจจัยเสี่ยงที่สนใจศึกษาเป็นสาเหตุของโรคหรือไม่อย่างถูกต้องตามความจริง ตัวอย่างเช่นในการทำการศึกษาเชิงทดลองเกี่ยวกับการรักษาโรค เราต้องการทราบว่ายาตัวใหม่ที่คิดค้นขึ้น มีประสิทธิผลดีกว่ายาตัวเดิมหรือไม่ หรือในการศึกษาแบบ Cohort study เราต้องการทราบว่า การมีปัจจัยเสี่ยงเป็นสาเหตุของอุบัติการณ์ที่เพิ่มขึ้นของโรคที่ตั้งใจศึกษาหรือไม่ อย่างไรก็ตามการพบความสัมพันธ์ไม่ได้หมายความว่าตัวแปรต้นจะเป็นสาเหตุของผลที่พบเสมอไป สิ่งที่ต้องพิจารณาประกอบกันเพื่อหาข้อสรุปว่าปรากฏการณ์ดังกล่าวเป็นสาเหตุที่แท้จริงหรือไม่ ได้แก่

1. <u>ความสัมพันธ์ในทางสถิติ (Statistical association)</u> ปัจจัยเสี่ยงหากจะเป็นสาเหตุของโรค โดยทั่วไปจะมีความสัมพันธ์กับการเกิดโรคอย่างมีนัยสำคัญทางสถิติเสมอหากขนาดของกลุ่มตัวอย่างมากพอและระเบียบวิธีวิจัยถูกต้องเหมาะสม อย่างไรก็ตามการที่ปัจจัยเสี่ยงมีความสัมพันธ์ทางสถิติกับการเกิดโรค ไม่ได้หมายความว่าจะมีความสำคัญทางคลินิก และไม่จำเป็นที่ปัจจัยเสี่ยงนั้นจะเป็นสาเหตุของโรคเสมอไป

2. <u>ความคลาดเคลื่อนอย่างเป็นระบบ (Systemic error)</u> หรือเรียกอีกอย่างหนึ่งว่า อคติ (Bias) หมายถึงความคลาดเคลื่อนของผลการวัดออกไปจากความเป็นจริงอย่างมีทิศทางชัดเจน หรืออีกนัยหนึ่ง คือการขาดความถูกต้อง (Validity) ของผลของการวัด Bias อาจแบ่งออกได้เป็น 3 กลุ่มใหญ่ ๆ คือ

   2.1 Selection bias หมายถึงความคลาดเคลื่อนของการวัด เนื่องจากสัดส่วนของประชากรกลุ่มต่าง ๆ ที่ถูกเลือกมาทำการศึกษา แตกต่างไปจากสัดส่วนที่เป็นจริงในประชากรจริง เช่นในการศึกษาแบบภาคตัดขวาง (Cross-sectional study) เพื่อความสัมพันธ์ระหว่างการสูบบุหรี่กับการป่วยด้วยโรคถุงลมปอดโป่งพอง ผู้วิจัยไม่พบความสัมพันธ์ระหว่างการสูบบุหรี่กับการป่วยเป็นโรคถุงลมปอดโป่งพองทั้ง ๆ ที่การสูบบุหรี่เป็นปัจจัยเสี่ยงของโรคถุงลมปอดโป่งพอง เนื่องจากผู้ที่เป็นโรคถุงลมปอดโป่งพองที่ยังสูบบุหรี่ ได้เสียชีวิตไปในสัดส่วนที่มากกว่า ผู้ที่เป็นโรคถุงลมปอดโป่งพองและหยุดสูบบุหรี่ (exposed case มีโอกาสเป็นกลุ่มตัวอย่างน้อยกว่ากลุ่มอื่น)

   2.2 Information bias หมายถึงความคลาดเคลื่อนของการวัดเนื่องจากอคติในระหว่างการวัดหรือรวบรวมข้อมูล เช่นการศึกษาเกี่ยวกับระดับความดันโลหิตที่ใช้เครื่องวัดที่ไม่ได้มาตรฐาน (Instrument bias) โดยเครื่องวัดความดันวัดค่าความดันโลหิตได้สูงหรือต่ำกว่าความเป็นจริงเสมอ หรือในการศึกษาที่เก็บข้อมูลโดยใช้วิธีสัมภาษณ์ แล้วผู้สัมภาษณ์พยายามเจาะจงถามคำถามกับกลุ่มผู้ป่วยจนได้ข้อมูลมากกว่ากลุ่มเปรียบเทียบ (Interviewer bias) เป็นต้น

   2.3 Confounding หมายถึงการสรุปความสัมพันธ์ระหว่างปัจจัยเสี่ยงที่สนใจศึกษา กับการเกิดโรคอย่างผิด ๆ เนื่องจากปัจจัยอื่น (Confounding factor) โดย Confounding factor ต้องมีความสัมพันธ์ทั้งกับปัจจัยเสี่ยงที่สนใจศึกษาและตัวโรค ตัวอย่างเช่น มีการศึกษาพบว่าการดื่มกาแฟเป็นปัจจัยเสี่ยงของการป่วยด้วยโรคหลอดเลือดหัวใจ แต่ต่อมาพบว่าเป็นข้อสรุปที่ผิดพลาด เนื่องจากผู้ที่ดื่มกาแฟมักสูบบุหรี่มากกว่าผู้ที่ไม่ดื่มกาแฟ (การสูบบุหรี่สัมพันธ์กับการดื่มกาแฟ) และบุหรี่ก็เป็นที่ทราบกันดีอยู่แล้วว่าสาเหตุของโรคหลอดเลือดหัวใจ (การสูบบุหรี่สัมพันธ์กับการป่วยเป็นโรคหลอด

เลือดหัวใจ) ความสัมพันธ์ระหว่างการดื่มกาแฟกับการป่วยด้วยโรคหัวใจและหลอดเลือดที่พบ จึงไม่ใช่ความสัมพันธ์จริง ในที่นี้การสูบบุหรี่ถือเป็น Confounding factor อย่างไรก็ตามตำราระบาดวิทยาบางเล่มไม่จัด Confounding เป็นอคติ แต่จัดเป็นความผิดพลาดในการสรุปผลการศึกษาอีกลักษณะหนึ่งแยกออกไปต่างหาก

3. ความคลาดเคลื่อนเชิงสุ่ม (Random error) ผลการศึกษาที่ได้อาจมีความคลาดเคลื่อนจากความเป็นจริงเนื่องจากความแปรปรวน (Variability) ตามธรรมชาติ ทั้งทางชีววิทยาและการวัดที่เกิดขึ้นเสมอ นอกจากนั้นแม้เราจะเลือกกลุ่มตัวอย่าง ๆ ดีที่สุดโดยการสุ่ม แต่กลุ่มตัวอย่างที่ใช้ศึกษาจริงอาจไม่เป็นตัวแทนที่แม่นยำของประชากรเสมอไป ซึ่งขนาดของกลุ่มตัวอย่างที่เพิ่มขึ้นจะช่วยลด Random error ลงได้

หากผู้ศึกษาทำการศึกษาอย่างมีคุณภาพสูงและมีขนาดตัวอย่างที่เพียงพอ หรืออีกนัยหนึ่งไม่มีข้อผิดพลาดในประเด็นดังกล่าวข้างต้นใด ๆ เลย และปัจจัยเสี่ยงกับการเกิดโรคมีความสัมพันธ์กันในทางสถิติ จึงบอกได้ว่า ความสัมพันธ์ที่พบเป็นความจริงและมีความถูกต้อง (Valid) อย่างไรก็ตาม เรายังไม่สามารถสรุปได้ว่าความสัมพันธ์นี้เป็นสาเหตุหรือไม่ ยังต้องอาศัยข้อพิจารณาอื่น ๆ ประกอบอีกซึ่ง Hill's criteria of causality เป็นเครื่องมืออีกอย่างหนึ่งที่ใช้ประกอบการพิจารณา

Hill's criteria of causality

ในปี ค.ศ.1965 Sir Austin Bradford Hill แพทย์ชาวอังกฤษ ได้ให้ข้อเสนอแนะในการพิจารณาว่าความสัมพันธ์ที่พบว่าเป็นสาเหตุหรือไม่ไว้ 9 ประการ คือ

1. ขนาดของความสัมพันธ์ (Strength of association) ขนาดของความสัมพันธ์ที่มาก เช่น ค่า RR สูง หรือ OR สูง เป็นหลักฐานสนับสนุนว่าปัจจัยนั้น ๆ เป็นสาเหตุของโรค อย่างไรก็ตามไม่ได้หมายความว่าความสัมพันธ์ที่มีขนาดน้อย จะไม่เป็นสาเหตุ
2. การศึกษาหลายครั้งได้ผลการศึกษาเหมือนกัน (Consistency) การทำการศึกษาเรื่องเดียวกันโดยนักวิจัยต่างกลุ่มกัน ทำคนละพื้นที่ คนละเวลา และในกลุ่มตัวอย่างที่มีลักษณะแตกต่างกัน หากได้ข้อสรุปเหมือนกัน จะเป็นข้อชี้บ่งที่ดีสำหรับการสรุปว่าปัจจัยเสี่ยงนั้นเป็นสาเหตุของโรค
3. ความจำเพาะ (Specificity) หมายถึงปัจจัยเสี่ยงหรือสาเหตุอย่างหนึ่งทำให้เกิดโรคเฉพาะเพียงอย่างหนึ่ง ใช้อธิบายสาเหตุของโรคติดเชื้อเป็นส่วนใหญ่ อย่างไรก็ตามไม่สามารถใช้ได้กับโรคเรื้อรังหรือปัจจัยเสี่ยงทางพฤติกรรมที่มีสาเหตุของโรคร่วมกันหลายอย่าง
4. ลำดับเวลาของสาเหตุ และผล (Temporality) ปัจจัยที่เป็นสาเหตุของโรคต้องมีหรือเกิดขึ้นก่อนการป่วยเป็นโรคเสมอ เป็นคุณสมบัติเพียงข้อเดียวจากทั้งหมด 9 ข้อ ที่ความสัมพันธ์ที่จะเป็นสาเหตุจำเป็นต้องมี อย่างไรก็ตามการมีคุณสมบัติข้อนี้มีน้ำหนักไม่มากในการบอกว่าความสัมพันธ์ดังกล่าวเป็นสาเหตุ
5. ขนาดของการเกิดโรคแปรตามขนาดของปัจจัยเสี่ยง (Dose response relationship) หากพบว่าการเพิ่มขนาดของปัจจัยเสี่ยงทำให้โอกาสการป่วยเป็นโรคสูงตามไปด้วย เป็นข้อบ่งชี้ที่ช่วยในการบอกว่า ปัจจัยดังกล่าวน่าจะเป็นสาเหตุของโรค

6. ความเป็นไปได้ในทางชีววิทยา (Biologic plausibility) ความสัมพันธ์ระหว่างปัจจัยเสี่ยงกับการเกิดโรคที่ทำการศึกษา หากมีหลักฐานทางการแพทย์หรือทางชีววิทยารองรับ ว่าปัจจัยเสี่ยงเข้าไปมีบทบาทในกลไกการเกิดโรคเป็นอย่างไร จะทำให้ความสัมพันธ์ดังกล่าวมีน้ำหนักมากขึ้น

7. ความสอดคล้องกันกับองค์ความรู้เดิม (Coherence) หมายถึง ความสัมพันธ์ที่กำลังศึกษาควรสอดคล้องและไม่ขัดแย้งกับองค์ความรู้ที่เกี่ยวข้องและมีอยู่เดิม

8. หลักฐานจากการผลการทดลอง (Experimental evidence) ความสัมพันธ์จะมีความน่าเชื่อว่าเป็นสาเหตุมากขึ้นหากสามารถมีผลการศึกษาเชิงทดลองมารองรับ คือเมื่อเพิ่มปัจจัยเสี่ยงเข้าไปจะทำให้เกิดโรค และเมื่อเอาปัจจัยเสี่ยงออกไปจะทำให้โรคหาย (Reversibility)

9. ความคล้ายคลึงกัน (Analogy) ความสัมพันธ์ที่พบมีความเป็นไปได้ที่จะเป็นสาเหตุ หากมีลักษณะที่คล้ายคลึงกับความสัมพันธ์เชิงสาเหตุที่เคยพิสูจน์แล้วว่าเป็นความจริง เช่นหากการศึกษาใหม่พบความสัมพันธ์ระหว่างการติดเชื้อไวรัสชนิดหนึ่งในมารดาขณะตั้งครรภ์กับความพิการในบุตร ก็น่าเชื่อว่าจะเป็นสาเหตุที่แท้จริงได้ เนื่องจากเรามีองค์ความรู้อยู่เดิมว่า มารดาที่ติดเชื้อหัดเยอรมัน (Rubella) จะทำให้บุตรในครรภ์มีโอกาสเกิดความพิการได้

****

เอกสารอ้างอิง

1. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
2. Szklo M, Nieto JF. Epidemiology beyond the basics. Sudbury; Jones and Bartlett Publishers: 2004.

****

แบบฝึกหัด

1. ในการศึกษาแบบ Case-control study หากผลการศึกษาพบ Odds Ratio น้อยกว่า 1 หมายถึง การ exposed ต่อ Independent variable

   A. ไม่สัมพันธ์กับความเสี่ยงในการป่วยเป็นโรค

B. ช่วยลดความเสี่ยงในการป่วยเป็นโรค
C. เพิ่มความเสี่ยงในการป่วยเป็นโรค
D. ช่วยลดโอกาสในการมีปัจจัยเสี่ยง
E. เพิ่มโอกาสในการมีปัจจัยเสี่ยง

2. “การพบความสัมพันธ์ระหว่างการรับประทานยาช่วยลดอาการแพ้ท้องชนิดหนึ่งกับความพิการในบุตร น่าเชื่อว่าความสัมพันธ์ดังกล่าวจะเป็นสาเหตุที่แท้จริงได้ เนื่องจากมีองค์ความรู้เดิมว่า มารดาที่รับประทานยา Thalidomide เพื่อลดอาการแพ้ท้องทำให้บุตรพิการได้”
คำกล่าวข้างต้น เข้าได้กับ Hill’s criteria of causation ข้อใด
A. Biologic plausibility
B. Temporal relationship
C. Analogy
D. Coherence
E. Consistency

3. การศึกษาแบบ Cohort study พบว่าผู้ติดเชื้อ HIV ที่ได้รับยาป้องกันวัณโรค และไม่ได้รับยาป้องกันวัณโรค มีอุบัติการณ์ของวัณโรคระยะแพร่กระจาย 1.6 และ 4.8 / person-year ตามลำดับ ให้คำนวณหา relative risk ของการป่วยเป็นวัณโรคระยะแพร่กระจายในผู้ที่ได้รับยาป้องกันวัณโรคเทียบกับผู้ที่ไม่ได้รับยาวัณโรค
A. 3
B. 3.2
C. 33%
D. 32%
E. 0.33

4. ในการศึกษาแบบ Case-control study เพื่อหาสาเหตุของภาวะ Autism ในเด็ก เจ้าหน้าที่โครงการวิจัยได้ใช้ความพยายามในการสอบถามประวัติการรับประทานยาระหว่างตั้งครรภ์ของมารดาที่มีบุตรป่วยเป็นโรคมากกว่ามารดาที่มีบุตรปกติ
การกระทำดังกล่าวอาจทำให้เกิดข้อผิดพลาดใด
A. Confounding
B. Selection bias
C. Information bias
D. Random error
E. Variability

5. จากข้อมูลผู้ป่วยโรคเบาหวาน 3,000 ราย พบผู้มีภาวะอ้วน 2,000 ราย ผู้วิจัยสรุปว่าโรคอ้วนสัมพันธ์กับโรคเบาหวาน
การสรุปผลดังกล่าวคลาดเคลื่อนจากการขาดข้อมูลใด
A. ประวัติการใช้ยา
B. การติดตามผู้ป่วย

C. กลุ่มประชากรควบคุม
D. พฤติกรรมการบริโภค
E. การประเมินตามกลุ่มอายุ

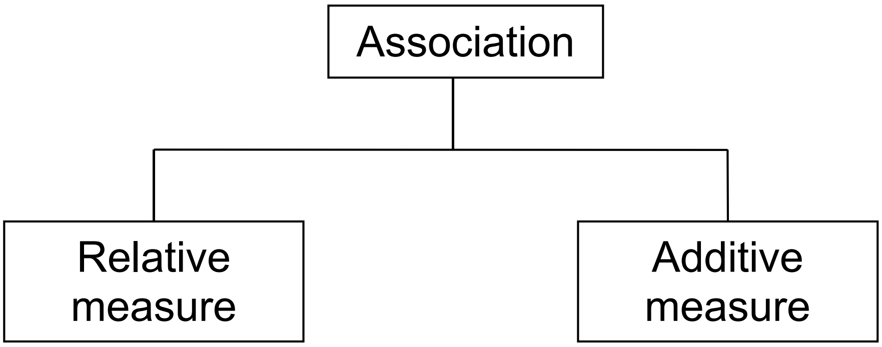
6. Cochrane Collaboration อาศัยหลักการข้อใดของ Hill's criteria of causation ในการทำงานมากที่สุด

A. Analogy
B. Coherence
C. Consistency
D. Experimental evidence
E. Strength of association

1.11.1.6 Health impact

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

วิธีการวัดความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงกับการเกิดโรคอาจแบ่งได้เป็น 2 ลักษณะ คือ Relative measure (Multiplicative) และ Additive measure (Health impact)



Relative measure of association บอกจำนวนเท่าของความเสี่ยงต่อการป่วยเป็นโรคที่เพิ่มขึ้นเนื่องจากการมีปัจจัยเสี่ยงโดยบอกใน scale คูณ-หาร (Multiplicative scale) จึงเหมาะสำหรับใช้อธิบายความสัมพันธ์ระหว่างปัจจัยเสี่ยงกับการเกิดโรคเพื่อการดูแลรักษาทางการแพทย์ (ระดับบุคคล) สำหรับ Additive measure of association เป็นการวัดความสัมพันธ์ในระดับ Additive scale คือ scale บวก-ลบ บอกถึงความเสี่ยงที่เพิ่มขึ้น (Excess risk) ในผู้ที่มีปัจจัยเสี่ยงเทียบกับผู้ที่ไม่มีปัจจัยเสี่ยง จึงมีประโยชน์ในทางการสาธารณสุข (ระดับประชากร) เนื่องจากช่วยบอกว่าการมีปัจจัยเสี่ยงส่งผลกระทบทำให้มีผู้ป่วยเพิ่มขึ้นเท่าใด บางครั้งจึงเรียกการหาความสัมพันธ์แบบนี้ว่า อัตราเสี่ยงกระทบ (Health impact)

รูปแบบการหาความสัมพันธ์ทางระบาดวิทยาแบบ Additive measure ที่สำคัญ มี 4 ลักษณะคือ

1. Attributable Risk in Exposed Individuals (AR_{exp})

คือผลต่างของความเสี่ยงต่อการป่วยเป็นโรค ระหว่างผู้ที่มีปัจจัยเสี่ยงกับผู้ที่ไม่มีปัจจัยเสี่ยง ในทางทฤษฎีหากเราเอาปัจจัยเสี่ยงนั้น ๆ ออกไป จะทำให้อุบัติการณ์ของโรคลดลงเท่ากับ AR_{exp} อย่างไรก็ตาม เนื่องจากในความเป็นจริง การกำจัดปัจจัยเสี่ยงดังกล่าวจะไม่สามารถลดความเสี่ยงต่อการป่วยเป็นโรคทั้งหมดที่เคยมีสะสมไว้ได้ ในทางปฏิบัติการจะลดความเสี่ยงให้ได้เท่าที่คำนวณได้ตาม AR_{exp} จึงต้องทำโดยการป้องกันไม่ให้ประชากรมีปัจจัยเสี่ยงนั้นเสียตั้งแต่ต้น AR_{exp} สามารถคำนวณได้โดย

$$AR_{exp} = Ie - Io$$

Ie = Incidence in exposed

Io = Incidence in unexposed

ตัวอย่าง

ตารางแสดงข้อมูลที่ได้จากการศึกษาความสัมพันธ์ระหว่างการใช้ยาเม็ดคุมกำเนิดกับการติดเชื้อในทางเดินปัสสาวะ

	การติดเชื้อในทางเดินปัสสาวะ	

การใช้ยาเม็ดคุมกำเนิด	*Yes*	*No*	*Total*
Yes	*27*	*455*	*482*
No	*31*	*1,831*	*1,862*
Total	*58*	*2,286*	*2,344*

AR $= Ie - Io$ $= a / (a + b) – c / (c + d)$

$= (27/482) - (31/1,862)$

$= 0.056 – 0.017 = 0.039$ *หรือ 3.9%*

หมายถึงการรับประทานยาเม็ดคุมกำเนิดจะเพิ่มความเสี่ยงต่อการติดเชื้อในทางปัสสาวะจากร้อยละ 1.7 (ในผู้ที่ไม่ได้รับประทาน) ไปเป็น ร้อยละ 5.6 (ในผู้ที่รับประทาน) หรือเสี่ยงเพิ่มขึ้น ร้อยละ 3.9

2. Percent AR_{exp} ($\%AR_{exp}$)

คือ สัดส่วนของความเสี่ยงต่อการเกิดโรคในผู้ที่ expose ต่อ risk factor ที่เนื่องมาจากการมีปัจจัยเสี่ยงนั้น คำนวณได้โดย

$$\%AR_{exp} = (Ie – Io) / Ie * 100$$

<u>ตัวอย่าง</u>

จากตัวอย่างเรื่องการศึกษาความสัมพันธ์ระหว่างการใช้ยาเม็ดคุมกำเนิดกับการติดเชื้อในทางเดินปัสสาวะ

$$\%AR_{exp} = (Ie – Io) / Ie \times 100$$

$$= (0.056 – 0.017) / 0.056 * 100$$

$$= 0.696 * 100 = 69.6\%$$

หมายถึง <u>ในผู้รับประทานยาเม็ดคุมกำเนิด</u> การป่วยเป็นโรคติดเชื้อในทางเดินปัสสาวะ ร้อยละ 69.6 เกิดเนื่องมาจากการรับประทานยาเม็ดคุมกำเนิด ส่วนที่เหลือเกิดเนื่องจากความเสี่ยงอื่น

$\%AR_{exp}$ ยังใช้ในการอธิบายประสิทธิผล (Efficacy) ของมาตรการรักษา หรือป้องกันโรค เช่น ยาหรือวัคซีนได้อีกด้วย วิธีการคำนวณจะคล้ายคลึงกับการคำนวณ $\%AR_{exp}$ ทั่วไป เพียงแต่ แทน Ie ด้วย risk ในกลุ่มเปรียบเทียบ และแทน Io ด้วย risk ในกลุ่มทดลอง ดังนี้

<u>ตัวอย่าง</u>

ในการศึกษาประสิทธิผลของวัคซีน พบว่ากลุ่มที่ได้วัคซีนจริง ป่วยเป็นโรคร้อยละ 5 ขณะที่กลุ่มที่ได้รับวัคซีนหลอก ป่วยเป็นโรค ร้อยละ 15

$$Efficacy = (15\% - 5\%) / 15\% * 100$$

$$= 66.7\%$$

หมายถึง วัคซีนมีประสิทธิผลในการป้องกันโรคร้อยละ 66.7

3. Population Attributable Risk (Pop AR, PAR)

หมายถึงสัดส่วนของความเสี่ยงต่อการเกิดโรคในประชากรที่เนื่องมาจากการมีปัจจัยเสี่ยง หรือเท่ากับผลต่างของความเสี่ยงต่อการเกิดโรคในประชากรทั้งหมดกับความเสี่ยงของการเกิดโรคในประชากรที่ไม่มีปัจจัยเสี่ยง ในทางทฤษฎีหากเราเอาปัจจัยเสี่ยงนั้น ๆ ออกจากประชากรจะช่วยลดความเสี่ยงลงได้เท่ากับ PAR

PAR = Incidence in population – Incidence in unexposed

PAR แปรผันตามความชุกของการมีปัจจัยเสี่ยงในประชากร โดย PAR จะมีค่าใกล้มาทาง incidence in unexposed เมื่อความชุกของการมีปัจจัยเสี่ยงต่ำ และจะมีค่าใกล้ไปทาง incidence in exposed เมื่อความชุกของการมีปัจจัยเสี่ยงสูง

ตัวอย่าง

จากตัวอย่างเรื่องการศึกษาความสัมพันธ์ระหว่างการใช้ยาเม็ดคุมกำเนิดกับการติดเชื้อในทางเดินปัสสาวะ

PAR = a + c / (a+b+c+d) – c / (c + d)

= (58/2,344) - (31/1,862)

= 0.025 – 0.017

= 0.008 หรือ 0.8%

หมายถึงการรับประทานยาเม็ดคุมกำเนิดจะเพิ่มความเสี่ยงต่อการติดเชื้อในทางปัสสาวะจาก ร้อยละ 1.7 (ในผู้ที่ไม่ได้รับประทาน) ไปเป็น ร้อยละ 2.5 (ในประชากรศึกษาทั้งหมด) หรือเสี่ยงเพิ่มขึ้น ร้อยละ 0.8

4. Percent Population Attributable Risk (%Pop AR, %PAR)

คือ สัดส่วนของความเสี่ยงต่อการเกิดโรคในประชากรที่เนื่องมาจากการมีปัจจัยเสี่ยงนั้น คำนวณได้โดย

%PAR = (PAR / Incidence in population) * 100

หรือ = { Pe (RR – 1) / Pe (RR – 1) + 1 } * 100

เมื่อ Pe = Proportion of exposure in population

RR = Relative risk

ตัวอย่าง

จากตัวอย่างเรื่องการศึกษาความสัมพันธ์ระหว่างการใช้ยาเม็ดคุมกำเนิดกับการติดเชื้อในทางเดินปัสสาวะ

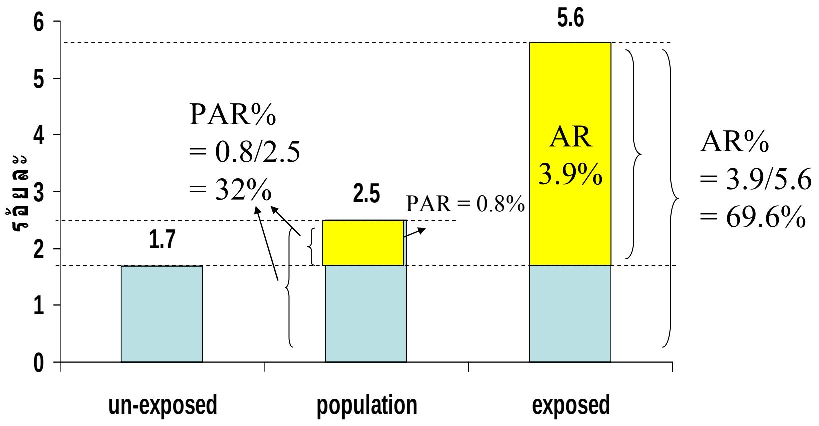
*%PAR = (PAR / Incidence in population) * 100*

*= (0.008 / 0.025) * 100*

*= 0.32 * 100   = 32%*

หมายถึง <u>ในประชากรศึกษาทั้งหมด</u> การป่วยเป็นโรคติดเชื้อในทางเดินปัสสาวะร้อยละ 32 เกิดเนื่องมาจากการรับประทานยาเม็ดคุมกำเนิดส่วนที่เหลือเกิดเนื่องจากสาเหตุอื่น

แผนภาพแสดงความสัมพันธ์ระหว่าง AR, %AR, PAR, %PAR



สาเหตุ

****

เอกสารอ้างอิง

1. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
2. Szklo M, Nieto JF. Epidemiology beyond the basics. Sudbury; Jones and Bartlett Publishers: 2004.

****

แบบฝึกหัด

1. ในการศึกษาประสิทธิผลของวัคซีนป้องกันไข้เลือดออก กลุ่มที่ได้วัคซีนจริง ป่วยเป็นโรคร้อยละ 2 ขณะที่กลุ่มที่ได้รับวัคซีนหลอก ป่วยเป็นโรค ร้อยละ 18 วัคซีนนี้มีประสิทธิผลร้อยละเท่าไหร่

A. 10

B. 11

C. 20

D. 89

E. 90

2. การศึกษาแบบ cohort study โดยติดตามคนงานโรงงานเฟอร์นิเจอร์ซึ่งมีสภาพการได้ยินปกติ เป็นเวลา 10 ปี พบว่า คนงานแผนกเลื่อยไม้มีภาวะหูเสื่อม 18 คน และมีสภาพการได้ยินปกติ 32 คน ส่วนคนงานแผนอื่น ๆ มีภาวะหูเสื่อม 3 คน และมีสภาพการได้ยินปกติ 22 คน

%Attributable risk ของการทำงานแผนกเลื่อยไม้ ต่อการเกิดภาวะหูเสื่อม มีค่าเท่ากับ

A. 33

B. 50

C. 67

D. 85

E. 90

3. "การป่วยเป็นมะเร็งปอดในผู้ที่สูบบุหรี่เป็นประจำ มีสาเหตุมาจากการสูบบุหรี่ร้อยละ 75"
คำกล่าวข้างต้นเข้าได้กับข้อใด

A. AR

B. %AR

C. PAR

D. %PAR

E. PRR

1.11.1.7 Efficacy of drug therapy, 1.11.1.8 Risk of drug therapy

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

ปัจจุบันการศึกษาแบบ Randomized control trials (RCT) ถือเป็นรูปแบบการวิจัยที่เป็นมาตรฐานทองที่ใช้ในการทดสอบประสิทธิผลของยารักษาโรคหรือมาตรการในการป้องกันโรคใหม่ ๆ ว่าได้ผลเป็นอย่างไร RCT เมื่อทำอย่างมีคุณภาพจะมีอคติ (Bias) น้อย และมี Internal validity สูงเมื่อเทียบกับการศึกษาแบบอื่น เพื่อประโยชน์ในการทำความเข้าใจกับผลการศึกษาและใช้ผลการศึกษาประกอบการดูแลรักษาผู้ป่วย แพทย์จำเป็นต้องเข้าใจถึงรูปแบบต่าง ๆ ในการวิเคราะห์ผล และประเมินผลการศึกษารูปแบบนี้

รูปแบบของการวิเคราะห์ผลการศึกษาของโครงการวิจัยแบบ RCT

การศึกษาวิจัยในชีวิตจริงเราอาจพบปัญหาหลายประการที่ส่งผลให้อาสาสมัครที่ถูกสุ่มให้อยู่ในกลุ่ม Intervention ในตอนแรกแต่ในท้ายที่สุดอาจไม่ได้รับ Intervention เนื่องจากต้องถูกสลับไปอยู่ในกลุ่มเปรียบเทียบ หรือได้รับ Intervention ไม่ครบตามขนาดและระยะเวลาที่กำหนด ทำให้อาสาสมัครที่ได้รับ Intervention อาจไม่เหมือนกับอาสาสมัครที่ถูกสุ่มได้ในตอนต้น ทำให้การศึกษาแบบ RCT มีแนวทางการวิเคราะห์ข้อมูล และนำเสนอผลการศึกษาได้ใน 2 ลักษณะคือ

1. <u>Intention-to-treat analysis</u>

หมายถึงการวิเคราะห์และนำเสนอผลการวิจัย โดยถือตามผลการสุ่มตัวอย่างตั้งแต่ตอนต้นของการศึกษาเป็นหลัก หากอาสาสมัครรายใดถูกสุ่มให้อยู่ในกลุ่มที่ได้รับ Intervention ก็จะถูกจัดให้อยู่ในกลุ่มนี้ตอนวิเคราะห์ผลการศึกษาเสมอ ไม่ว่าในความเป็นจริงอาสาสมัครรายนี้จะได้รับ Intervention หรือไม่ก็ตาม Intention-to-treat analysis จึงเป็นการตอบคำถามที่ว่า การรักษาแบบใดให้ผลดีกว่ากัน ณ ตอนทำที่ตัดสินใจเลือกวิธีการรักษา

ข้อดีของการวิเคราะห์แบบนี้คือ เป็นการตอบคำถามที่แพทย์ต้องเผชิญในชีวิตจริงเมื่อต้องพิจารณาให้การรักษาผู้ป่วย ซึ่งก็คือจะตัดสินใจให้การรักษานี้หรือไม่ นอกจากนั้นยังเป็นการคงไว้ซึ่งคุณสมบัติที่สำคัญที่ทำให้ RCT เป็นมาตรฐานทองของการวิจัยทางการแพทย์ คือการสุ่มเลือกกลุ่มตัวอย่าง อย่างไรก็ตามการวิเคราะห์แบบนี้ก็มีข้อด้อยได้แก่ ถ้ามีการเปลี่ยนแปลงไปมากของอาสาสมัครภายหลังการสุ่ม เช่นมีการออกจากการศึกษากลางคัน หรือสลับกลุ่มกันจำนวนมาก จะทำให้อำนาจการทดสอบทางสถิติ (Statistical power) ของการวิจัยลดลง จนอาจตรวจไม่พบความแตกต่างของผลการศึกษาระหว่างอาสาสมัครสองกลุ่ม ผู้วิจัยจึงต้องพยายามทำการวิจัยอย่างมีคุณภาพ เพื่อให้ได้ความร่วมมือในการรักษาจากอาสาสมัคร (Compliance) สูง และการขาดการรักษา (Loss to follow up) ต่ำ การวิเคราะห์และนำเสนอผลการศึกษาแบบ Intention-to-treat นี้ ถือเป็นวิธีมาตรฐาน และได้รับการยอมรับมากกว่าการวิเคราะห์แบบ Explanatory analysis ที่จะกล่าวถึงต่อไป

2. <u>Explanatory analysis (On-treatment analysis, As treated analysis)</u>

หมายถึงการวิเคราะห์และนำเสนอผล โดยถือเอาการได้รับหรือไม่ได้รับ Intervention จริง ระหว่างการวิจัยเป็นหลัก เช่นหากอาสาสมัครที่ถูกสุ่มให้อยู่ในกลุ่ม Intervention แล้วภายหลังถูกสลับไปอยู่กลุ่ม Control ก็ให้จัดอยู่ในกลุ่ม Control ตอนวิเคราะห์ผล การวิเคราะห์แบบนี้จึงเป็นการตอบคำถามที่ว่า Intervention ให้ Outcome ที่ดีกว่าในผู้ที่ได้รับ Intervention นั้นจริง ๆ เมื่อเทียบกับผู้ที่ไม่ได้รับหรือไม่

การวิเคราะห์แบบนี้ช่วยให้มีโอกาสพบความแตกต่างระหว่างกลุ่ม Intervention กับกลุ่ม Control อย่างมีนัยสำคัญทางสถิติได้ง่ายขึ้น แต่เป็นวิธีที่ได้รับการยอมรับน้อยกว่าแบบ Intention-to-treat analysis เนื่องจากไม่ได้เป็นการตอบคำถามวิจัยในตอนต้นที่ว่า หากเราเสนอการรักษานี้แก่ผู้ป่วยแล้วจะได้ผลการรักษาดีกว่าหรือไม่ นอกจากนั้นหากมีความเปลี่ยนแปลงไปของอาสาสมัครในการศึกษาจริงกับ

อาสาสมัครที่สุ่มได้ในตอนต้นมาก การวิจัยนี้จะขาดคุณสมบัติที่สำคัญของ RCT คือการ Randomization และทำให้มีโอกาสสูงที่จะสรุปผลการวิจัยผิดพลาดเนื่องจาก Confounding

การประเมินผลการรักษา และผลข้างเคียงของการรักษา ในการศึกษาแบบ RCT

เนื่องจากอาจกล่าวได้ว่า RCT เป็นรูปแบบเฉพาะของ Cohort study การสรุปผลการศึกษาจึงเป็นการเปรียบเทียบอัตราการเกิดเหตุการณ์ที่สนใจในกลุ่มตัวอย่างสองกลุ่มคล้าย ๆ กัน โดยสำหรับ RCT เป็นการเปรียบเทียบอัตราการเกิดเหตุการณ์ที่สนใจในอาสาสมัครที่ได้รับ Intervention (Experimental event rate, EER) กับเหตุการณ์ในอาสาสมัครกลุ่ม Control ที่ไม่ได้รับ Intervention (Control event rate, CER) ในการศึกษาที่เป็นการทดสอบประสิทธิผลของยา EER หมายถึงอัตราการเกิดผลการรักษาที่ดีในอาสาสมัครที่ได้รับยาจริง ส่วน CER คืออัตราการเกิดผลการรักษาที่ดีในอาสาสมัครที่ได้รับยามาตรฐานเดิม หรือยาหลอก (Placebo)

การประเมินผลการรักษาส่วนใหญ่เน้นไปที่ผลด้านดีของยาหรือการรักษานั้น (Efficacy of drug therapy) เป็นสำคัญ ได้แก่การหายจากโรค อาการที่ดีขึ้นของผู้ป่วย หรือการป้องกันความพิการ อย่างไรก็ตามยังมีการประเมินผลในด้านตรงข้าม คือในด้านที่ไม่ดีด้วย (Adverse effects) เรียกว่าการประเมินผลข้างเคียงของการรักษา (Risk of drug therapy) เช่นอาการแพ้ยาต่าง ๆ หรือผลข้างเคียงทางคลินิกอื่นที่เกิดจากการรักษานั้น ผลการประเมินทั้งสองด้านใช้ประกอบกันในการตัดสินใจว่า ควรนำยาหรือการรักษาดังกล่าวมาใช้กับผู้ป่วยหรือไม่

ตัวอย่างต่อไปนี้ ใช้ประกอบการทำความเข้าใจกับการประเมินผลการรักษาทั้งสองด้าน

“การรักษาผู้ป่วยโรคเบาหวานด้วยการฉีดอินซูลินเทียบกับยากินตามปกติ ช่วยลดอัตราการเกิดจอประสาทตาเสื่อมลง จาก 38% เหลือ 13% ขณะเดียวกันก็เพิ่มโอกาสการป่วยเนื่องจากน้ำตาลในเลือดต่ำจาก 23% เป็น 57%”

<u>การประเมินผลด้านดีของการรักษา (Efficacy of drug therapy)</u>

ผลด้านดีของการรักษาที่ได้จากการศึกษาแบบ RCT มีการคำนวณ และนำเสนอได้หลายแบบ แต่ละแบบมีความหมาย ตลอดจนที่ใช้ที่แตกต่างกัน

1. *Relative risk reduction (RRR)* หมายถึงสัดส่วนของโอกาสเสี่ยงของการเกิด Outcome ด้านไม่ดีที่ลดลง เนื่องจากการได้รับ Intervention เมื่อเทียบกับ Control เป็นค่าที่บอกประสิทธิผลของการรักษา (Efficacy)

    คำนวณโดยสูตร

    RRR = |EER – CER| / CER

    จากตัวอย่างข้างต้น = *|13% – 38%| / 38%*

    = *66%*

    หมายถึงการรักษาผู้ป่วยโรคเบาหวานด้วยอินซูลินเทียบกับการรักษาปกติ มีผลดีคือช่วยลดโอกาสการเกิดจอประสาทตาเสื่อมในผู้ป่วยลงร้อยละ 66

2. *Absolute risk reduction (ARR)* หมายถึงโอกาสการเกิด Outcome ด้านไม่ดีที่ลดลง เนื่องจากการได้รับ Intervention เมื่อเทียบกับ Control

คำนวณโดยสูตร

ARR = |EER – CER|

จากตัวอย่างข้างต้น = |13% – 38%|
= 25% = 0.25

หมายถึงผู้ป่วยโรคเบาหวานที่รักษาด้วยอินซูลินเมื่อเทียบกับการรักษาปกติ เกิดจอประสาทตาเสื่อมน้อยกว่าผู้ที่ได้รับการรักษาปกติไปร้อยละ 25

3. *Number needed to treat (NNT)* หมายถึงจำนวนผู้ที่ต้องได้รับการรักษา เพื่อป้องกันการเกิดผลที่ไม่ดี หรือเพื่อให้ได้ผลการรักษาที่ดีเพิ่มขึ้น 1 ราย

คำนวณโดยสูตร

NNT = 1 / ARR

จากตัวอย่างข้างต้น = 1 / 0.25
= 4

หมายถึงต้องเปลี่ยนการรักษาผู้ป่วยโรคเบาหวานจากการรักษาปกติเป็นการรักษาด้วยอินซูลิน 4 ราย จึงจะสามารถป้องกันการเกิดจอประสาทตาเสื่อมจริงในผู้ป่วย 1 ราย

NNT ยิ่งสูง หมายถึงต้องรักษาผู้ป่วยจำนวนมากเพื่อเปลี่ยนผลการรักษาให้เป็นด้านดีเพิ่มขึ้น 1 ราย เพราะฉะนั้นในการศึกษายารักษาโรคใหม่ ๆ NNT ยิ่งต่ำจึงยิ่งดี

<u>การประเมินผลข้างเคียงของการรักษา (Risk of drug therapy)</u>

1. *Relative risk increase (RRI)* หมายถึงสัดส่วนของโอกาสเสี่ยงของการเกิดผลข้างเคียงที่เพิ่มขึ้นเนื่องจากการได้รับ Intervention เมื่อเทียบกับ Control

คำนวณโดยสูตร

RRI = |EER – CER| / CER

จากตัวอย่างข้างต้น = |57% – 23%| / 23%
= 148%

หมายถึงการรักษาผู้ป่วยโรคเบาหวานด้วยอินซูลินเทียบกับการรักษาปกติ มีผลเสียคือ ทำให้เกิดโอกาสป่วยเนื่องจากน้ำตาลในเลือดต่ำเพิ่มขึ้นร้อยละ 148

2. *Absolute risk increase (ARI)* หมายถึงโอกาสการเกิดผลข้างเคียงที่เพิ่มขึ้นเนื่องจากการได้รับ Intervention เมื่อเทียบกับ Control

คำนวนโดยสูตร

ARI = |EER – CER|

จากตัวอย่างข้างต้น = |57% – 23%|

= 34%   = 0.34

หมายถึงผู้ป่วยโรคเบาหวานที่รักษาด้วยอินซูลินเกิดป่วยเนื่องจากน้ำตาลในเลือดต่ำมากกว่าผู้ที่ได้รับการรักษาปกติไปร้อยละ 34

3. *Number needed to harm (NNH)* หมายถึงจำนวนผู้ป่วยที่ให้การรักษาโดยเฉลี่ยที่จะทำให้เกิดผลข้างเคียงเพิ่มขึ้น 1 ราย

คำนวนโดยสูตร

NNH = 1 / ARI

จากตัวอย่างข้างต้น = 1 / 0.34

= 3

หมายถึงการเปลี่ยนการรักษาผู้ป่วยโรคเบาหวานจากการรักษาปกติเป็นการรักษาด้วยอินซูลิน 3 ราย โดยเฉลี่ยแล้วจะทำให้เกิดผู้ป่วยเนื่องจากน้ำตาลในเลืดต่ำเพิ่มขึ้น 1 ราย

ตรงข้ามกับ NNT สำหรับค่า NNH นี้ยิ่งสูงจะยิ่งดี เพราะหมายถึงยาใหม่มีผลข้างเคียงน้อย คือต้องทำการรักษาผู้ป่วยจำนวนมาก จึงจะเพิ่มการมีผลข้างเคียงรายใหม่ 1 ราย

Efficacy, Effectiveness และ Efficiency

การประเมินผลการรักษาทางการแพทย์ มีศัพท์ 3 คำที่มีความหมายและที่ใช้แตกต่างกัน

1. Efficacy หมายถึง intervention ที่สนใจศึกษาได้ผลหรือไม่ โดยเป็นการพิจารณาผลที่เกิดในบริบทสถานการณ์ที่เป็นอุดมคติในการวิจัย มักใช้ในกรณีของผลการวิจัยที่มีการควบคุมองค์ประกอบต่าง ๆ ไว้เป็นอย่างดี เช่น ยาใหม่ตัวนี้ มี Efficacy ดี เมื่อพิจารณาจากผลการวิจัยแบบ RCT
2. Effectiveness หมายถึงผลของ intervention เช่นการรักษา ในสถานการณ์การให้บริการในชีวิตจริงหรือในคลินิกทั่วไปที่ไม่ได้มีการควบคุมองค์ประกอบต่าง ๆ เหมือนอย่างในการทำวิจัย เช่น ยาที่ได้ผลดีใน RCT อาจใช้ไม่ได้ผลในการรักษาพยาบาลตามปกติ ณ สถานพยาบาล ด้วยข้อจำกัดต่าง ๆ
3. Efficiency เป็นการพิจารณาผลของ intervention ในทางเศรษฐศาสตร์ ว่ามีความคุ้มค่า (cost effective) หรือไม่ โดยพิจารณาเทียบกับต้นทุน ค่าใช้จ่ายในการป้องกันหรือรักษาโรค intervention ที่มี Efficacy และ Effectiveness ดี จึงอาจไม่มี Efficiency ก็ได้ หากมีราคาแพงไม่คุ้มกับผลการรักษาที่ได้

ในภาษาไทย มีศัพท์ทั่วไปที่ไม่ได้ใช้เฉพาะทางการแพทย์ 2 คำ ซึ่งมีความหมายใกล้เคียงกับคำในภาษาอังกฤษ 3 คำ ข้างต้น คือคำว่า “ประสิทธิผล” ที่หมายถึงผลสำเร็จ ผลที่เกิดขึ้น และ “ประสิทธิภาพ” หมายถึงความสามารถที่ทำให้เกิดผลในการทำงาน ในกรณีของผลการรักษาหรือการป้องกันโรค การใช้คำว่าประสิทธิผล น่าจะจะตรงตามความหมายที่ต้องการสื่อมากกว่า เพราะเราต้องการเน้นที่ตัวผล ไม่ได้เน้นขั้นตอนกระบวนการ นอกจากนั้น คนไทยส่วนใหญ่มักแปลคำว่า Efficiency เป็นภาษาไทยว่าประสิทธิภาพ ด้วยเพราะฉะนั้นหากเราพิจารณาเฉพาะผลการรักษาโดยไม่ได้คำนึงถึงความคุ้มค่า หรือความประหยัด จึงไม่ควรใช้คำว่าประสิทธิภาพ

การประยุกต์ใช้ผลการศึกษาแบบ RCT ในการดูแลรักษาผู้ป่วยแต่ละราย

การนำผลการศึกษาแบบ RCT มาใช้ประกอบการดูแลรักษาผู้ป่วยแต่ละราย มีข้อควรคำนึงดังนี้ คือ มีผลการศึกษาแบบ RCT ที่น่าเชื่อถือว่าได้ผลในการรักษาผู้ป่วยลักษณะนั้น ๆ หรือไม่ การศึกษานั้นทำในผู้ป่วยในลักษณะที่คล้ายคลึงกับผู้ป่วยรายที่กำลังจะให้การรักษาหรือไม่ กรณีที่ลองให้การรักษาไปแล้ว ต้องมีการติดตามผลการรักษาอย่างใกล้ชิด เนื่องจากผลการรักษาใน RCT เป็นการเฉลี่ยจากผู้ป่วยหลายราย อาจไม่ได้ผลดีในผู้ป่วยรายนี้ก็ได้ และต้องพิจารณาถึงผลข้างเคียงที่เกิดจากการรักษาด้วยว่า มีมากกว่าประโยชน์ที่ได้จากการให้การรักษาหรือไม่

****

เอกสารอ้างอิง

1. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
2. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.
3. Leon Gordis. Epidemiology 4th edition. USA: 2008

****

แบบฝึกหัด

1. รูปแบบมาตรฐานในการวิเคราะห์ผลวิจัย ของการศึกษาแบบ Randomized controlled trials ได้แก่

    A. Interim analysis

    B. Survival analysis

    C. Explanatory analysis

    D. On-treatment analysis

    E. Intention-to-treat analysis

2. ข้อใดมีความหมายเดียวกับ Treatment efficacy

    A. Experimental event rate

    B. Control event rate

    C. Relative risk reduction

    D. Absolute risk reduction

    E. Number needed to treat

3. การศึกษาประสิทธิผลของยาป้องกันเส้นเลือดแดงอุดตันตัวใหม่ เทียบกับยามาตรฐาน ผู้วิจัยตรวจหาภาวะเลือดออกผิดปกติซึ่งเป็นผลข้างเคียงของยาด้วย พบว่าในผู้ป่วยที่ได้รับยามาตรฐาน 40 ราย มีผู้มีภาวะ

ดังกล่าว 4 ราย ขณะที่ผู้ได้รับยาใหม่ 80 ราย มีผู้มีภาวะดังกล่าว 12 ราย Relative risk increase ของภาวะเลือดออกผิดปกติเนื่องจากยาใหม่เทียบกับยาเดิม คิดเป็นร้อยละเท่าใด

A. 10

B. 15

C. 20

D. 50

E. 75

4. ยา A รักษานิ่วได้ผล 80% ยา B รักษานิ่วได้ผล 60% ให้คำนวณ Number needed to treat ของการเปลี่ยนการรักษาจากยา B ไปเป็นยา A

A. 5

B. 10

C. 0.2

D. 0.02

E. 20

1.11.1.9 Sensitivity, specificity, predictive values

การวินิจฉัยโรค (Diagnosis)

ก่อนที่แพทย์จะให้การดูแลรักษาผู้ป่วย ขั้นตอนสำคัญที่ต้องทำคือต้องให้การวินิจฉัย (Diagnosis) ว่าอาการ (Symptoms) ตลอดจนอาการแสดง (Signs) ที่นำผู้ป่วยมานั้น เป็นโรคหรือภาวะความผิดปกติอะไร โดยทั่วไปแพทย์จะให้วินิจฉัยโรคจากข้อมูลการเจ็บป่วย ประกอบกับความรู้และประสบการณ์ที่ได้มาจากการให้การดูแลรักษาผู้ป่วยรายอื่น ๆ ที่มีลักษณะคล้ายคลึงกัน ข้อมูลสำคัญอีกส่วนหนึ่งที่ช่วยให้แพทย์วินิจฉัยโรคได้อย่างถูกต้อง คือข้อมูลที่ได้จากชุดตรวจวินิจฉัย (Diagnostic tests) ต่าง ๆ

ชุดตรวจวินิจฉัยหมายรวมถึง เครื่องช่วยในการประเมินการเจ็บป่วย อาจเป็นผลการตรวจร่างกาย ผลคะแนนจากแบบสัมภาษณ์ เช่นแบบทดสอบภาวะซึมเศร้า หรือการประเมินลักษณะทางคลินิก เช่นแบบประเมินภาวการณ์ขาดออกซิเจนในเด็กทารกแรกเกิด Apgar's score หรือเป็นผลการตรวจทางห้องปฏิบัติการ (Laboratory tests) ต่าง ๆ เช่นการตรวจเลือดเพื่อหาภาวะการทำงานของตับ หรือภาพถ่ายรังสีทรวงอกเป็นต้น ข้อมูลที่ได้จากชุดตรวจวินิจฉัยเหล่านี้ จะช่วยแพทย์ในการวินิจฉัยโรค ช่วยพยากรณ์การดำเนินโรค และวางแผนการรักษาที่เหมาะสม สำหรับผู้ป่วยแต่ละรายต่อไป

การตรวจวินิจฉัยที่ดีที่สุด แม่นยำที่สุดเท่าที่จะเป็นไปได้สำหรับแต่ละโรค ณ ช่วงเวลาหนึ่ง ๆ เรียกว่าชุดตรวจมาตรฐานทอง (Gold standard diagnostic test) แต่เนื่องจากการตรวจที่ถือเป็น Gold standard เหล่านี้มักมีความยุ่งยากซับซ้อนในการตรวจ มีค่าใช้จ่ายสูง ผู้ป่วยอาจต้องเสี่ยงอันตรายหรือต้องเผชิญกับความเจ็บปวดทรมานระหว่างการตรวจ หรือใช้เวลานานกว่าจะทราบผล จึงทำให้มีการพัฒนาชุดตรวจวินิจฉัยอื่น ๆ ขึ้นมา เพื่อใช้แทน Gold standard ในสถานการณ์ที่เหมาะสม ตัวอย่างเช่น เราใช้ผลการตรวจเอกซเรย์คอมพิวเตอร์ช่องท้องในการวินิจฉัยมะเร็งตับ แทนการเจาะหน้าท้องเพื่อนำชิ้นเนื้อมาตรวจเนื่องจากอาจมีอันตราย หรือเราใช้การตรวจคลื่นไฟฟ้าหัวใจ (Electrocardiogram) แทนที่จะทำการสวนเส้นเลือดแดงใหญ่เพื่อฉายภาพรังสี ในการวินิจฉัยโรคกล้ามเนื้อหัวใจขาดเลือดไปเลี้ยง เป็นต้น

โดยทั่วไปผลตรวจของชุดตรวจวินิจฉัยมีความเป็นไปได้เพียง 2 อย่างคือ มีความผิดปกติหรือไม่มี บวกหรือลบ เป็นโรคหรือไม่เป็นโรค เช่นการตรวจเลือดเพื่อหาภาวะการติดเชื้อเอชไอวี มีผลที่เป็นไปได้สองอย่างคือ ผลบวกหมายถึงติดเชื้อ ผลลบหมายถึงไม่ติดเชื้อ หรือการตรวจอุจจาระเพื่อหาไข่พยาธิ ผลบวกคือพบไข่พยาธิ ผลลบคือไม่พบไข่พยาธิ เป็นต้น สำหรับข้อมูลของชุดตรวจวินิจฉัยที่เป็นข้อมูลบอกปริมาณ เช่นระดับน้ำตาลในเลือด นิยมหาค่าที่เป็นจุดตัด (cut off point) เพื่อแยกค่าผิดปกติออกจากค่าปกติเช่นเดียวกัน ซึ่งจะได้กล่าวถึงต่อไป

ความถูกต้องของผลการตรวจ (Accuracy of test result)

ในการศึกษาเกี่ยวกับคุณภาพของชุดตรวจวินิจฉัย ประชากรศึกษาสามารถแบ่งออกได้เป็น 2 กลุ่ม คือผู้ป่วยและผู้ไม่ป่วย โดยอ้างอิงผลตรวจของ Gold standard และในขณะเดียวกันก็แบ่งประชากรศึกษาออกเป็น 2 กลุ่ม ตามผลของชุดตรวจวินิจฉัยที่เราต้องการศึกษาหรือนำมาเปรียบเทียบกับความเป็นจริง คือชุดตรวจวินิจฉัยให้ผลบวก (Positive) หรือผิดปกติ (Abnormal) และผลลบ (Negative) หรือปกติ (Normal) ชุดตรวจวินิจฉัยที่ดี ต้องมีคุณสมบัติให้ผลบวกในผู้ป่วย เรียกว่าผลบวกจริง (True positive) และให้ผลลบในผู้ไม่ป่วย เรียกว่าผลลบจริง (True negative) ได้อย่างถูกต้องในสัดส่วนที่มากที่สุด ขณะที่ให้ผลผิดพลาด คือให้ผลลบในผู้ป่วย เรียกว่าผลลบลวง (False negative) และให้ผลบวกในผู้ไม่ป่วย เรียกว่าผลบวกลวง (False positive) ใน

สัดส่วนที่น้อยที่สุด ความถูกต้อง (Accuracy) ของชุดตรวจวินิจฉัย หาได้จากสัดส่วนของ ผลรวมของผลบวกจริงกับผลลบจริง หารด้วยจำนวนการตรวจทั้งหมด นิยมบอกเป็นค่าร้อยละ การจะทราบว่าชุดตรวจวินิจฉัยหนึ่ง ๆ มีความถูกต้องมากน้อยแค่ไหน จะต้องนำผลการตรวจที่ได้จาก Diagnostic test นั้น ๆ ไปเปรียบเทียบผลการตรวจของ Gold standard

Accuracy of diagnostic test = (True positive + True negative) / Total tests

ตารางแสดงความสัมพันธ์ระหว่างผลการตรวจของชุดตรวจวินิจฉัย กับการป่วยเป็นโรค

		Disease (Gold standard)	
		Present	Absent
Diagnostic Test	Positive	True positive	False positive
	Negative	False negative	True negative

ความไว (Sensitivity) และความจำเพาะ (Specificity)

คุณลักษณะที่สำคัญ 2 ประการ ที่เป็นเครื่องบอกถึงความถูกต้องของชุดตรวจวินิจฉัย และแพทย์ใช้ประกอบการพิจารณาเลือกนำไปใช้ ได้แก่ ความไว (Sensitivity) และความจำเพาะ (Specificity)

1. Sensitivity หมายถึง ความน่าจะเป็นของการที่ชุดตรวจวินิจฉัยจะให้ผลตรวจเป็นบวก จากผู้ที่ป่วยจริงด้วยโรคที่สนใจศึกษา (Gold stand เป็นบวก) ทั้งหมด ชุดตรวจวินิจฉัยที่มี Sensitivity สูง จะไม่พลาดในการตรวจพบผู้ป่วย จึงให้ผลลบลวง (False negative) น้อย เราควรเลือกใช้ชุดตรวจวินิจฉัยที่มี Sensitivity สูง ในกรณีที่โรคมีความรุนแรงสูงแต่สามารถให้การรักษาได้ เช่นวัณโรค หรือเป็นโรคติดต่อที่แพร่กระจายได้ เช่นซิฟิลิส นอกจากนั้นชุดตรวจวินิจฉัยที่มี Sensitivity สูง ยังเหมาะที่จะใช้ในขั้นตอนแรกๆ ของการตรวจวินิจฉัย ที่อาการและอาการแสดงของผู้ป่วยเข้าได้กับโรคหลาย ๆ โรค เพราะหากผลการตรวจของชุดตรวจวินิจฉัยที่มี sensitivity สูงของโรคที่สงสัยโรคใดออกมาเป็นลบ แพทย์จะได้ตัดโรคนั้น ๆ ออกไปได้ (Rule out) และพิจารณาถึงเฉพาะแต่โรคที่ชุดตรวจวินิจฉัยให้ผลบวกที่ยังเหลืออยู่ต่อไป

   Sensitivity of diagnostic test = True positive / Diseases

2. Specificity คือ ความน่าจะเป็นของการที่ชุดตรวจวินิจฉัยจะให้ผลตรวจเป็นลบ จากผู้ที่ไม่ได้ป่วยด้วยโรคที่สนใจศึกษา (Gold standard เป็นลบ) ทั้งหมด ชุดตรวจวินิจฉัยที่มี Specificity สูง จะไม่ให้ผลการตรวจที่ผิดพลาดโดยบอกว่าผู้ไม่ป่วยเป็นผู้ป่วย จึงมีผลบวกลวง (False positive) น้อย เหมาะที่จะใช้ในการตรวจยืนยันว่าผู้ป่วย ๆ เป็นโรคนั้นจริง (Rule in) โดยเฉพาะอย่างยิ่งในโรคที่หากวินิจฉัยผิดพลาดว่าเป็นโรคนั้นทั้ง ๆ ที่จริง ๆ แล้วไม่ได้เป็น จะทำอันตรายต่อผู้ป่วยทั้งทางด้านร่างกาย จิตใจ และเศรษฐกิจ เช่นโรคมะเร็งที่ต้องรับการรักษาด้วยการผ่าตัด หรือการฉายรังสี หรือโรคที่ยังไม่มีการรักษา เป็นต้น

   Specificity of diagnostic test = True negative / Non-diseases

Predictive Values

Sensitivity และ Specificity ของชุดตรวจวินิจฉัยเป็นคุณสมบัติเฉพาะตัวที่จะมีค่าคงที่เสมอ และใช้ประกอบการตัดสินใจว่าจะเลือกใช้ชุดตรวจวินิจฉัยใด อย่างไรก็ตามเมื่อแพทย์ตัดสินใจใช้ชุดตรวจวินิจฉัยหนึ่ง ๆ และได้ผลการตรวจกลับมาแล้ว ค่า Sensitivity และ Specificity ของชุดตรวจวินิจฉัยจะไม่มีประโยชน์ในการประกอบการตัดสินใจดูแลรักษาผู้ป่วยอีกต่อไป เพราะสิ่งที่แพทย์ต้องการทราบในขณะนี้ก็คือผลการตรวจที่ออกมาแล้วว่ามีโอกาสถูกต้องตามความสถานะความเจ็บป่วยจริงมากน้อยเพียงใดเรียกว่าค่า Predictive value of diagnostic test ซึ่งมี 2 ลักษณะ คือ

1. <u>Positive predictive value (PPV)</u> คือความน่าจะเป็นที่ผลบวกของชุดตรวจวินิจฉัยจะเป็นผลบวกจริง PPV คำนวณจาก สัดส่วนหรือร้อยละของจำนวนชุดตรวจวินิจฉัยที่ให้ผลบวกจริง หารด้วยจำนวนชุดตรวจวินิจฉัยที่ให้ผลเป็นบวกทั้งหมด

PPV = True positive / All positive

2. <u>Negative predictive value (NPV)</u> คือความน่าจะเป็นที่ผลลบของชุดตรวจวินิจฉัยจะเป็นผลลบจริง NPV คำนวณจาก สัดส่วนหรือร้อยละของจำนวนชุดตรวจวินิจฉัยที่ให้ผลลบจริง หารด้วยจำนวนชุดตรวจวินิจฉัยที่ให้ผลลบทั้งหมด

NPV = True negative / All negative

Predictive value ทั้งสองลักษณะ ไม่ได้เป็นผลจากคุณสมบัติเฉพาะของชุดตรวจวินิจฉัยเท่านั้น แต่จะมีค่าไม่คงที่และเปลี่ยนแปลงไปตามความชุกของโรค (Prevalence of disease) ในประชากรศึกษาด้วย หากใช้ในประชากรที่มีความชุกของโรคนั้นต่ำ ชุดตรวจวินิจฉัยจะให้ PPV ต่ำลง แต่ NPV สูงขึ้น และในทางกลับกันหากใช้ในประชากรที่มีความชุกของโรคสูง ชุดตรวจวินิจฉัยจะให้ PPV สูงขึ้น แต่ NPV ต่ำ

<u>*ตัวอย่าง*</u>

ในการศึกษาความสามารถของเทคนิคใหม่ทางห้องปฏิบัติการ เพื่อวินิจฉัยแยกผู้ติดเชื้อHIV ที่มี CD4+ cell น้อยกว่า 200 cells / cu.mm. (ผู้ป่วยเอดส์) ออกจากผู้ติดเชื้อรายอื่นๆ เปรียบเทียบกับเทคนิคที่เป็นมาตรฐานทอง (Gold standard) ได้ผลการศึกษาดังนี้

gold standard ได้ค่าน้อยกว่า 200 และเทคนิคใหม่ได้ค่าน้อยกว่า 200 จำนวน 80 ราย

gold standard ได้ค่าน้อยกว่า 200 แต่เทคนิคใหม่ได้ค่ามากกกว่าหรือเท่ากับ 200 จำนวน 20 ราย

gold standard ได้ค่ามากกว่า 200 แต่เทคนิคใหม่ได้ค่าน้อยกว่า 200 จำนวน 120 ราย

gold standard ได้ค่ามากกว่า 200 และเทคนิคใหม่ได้ค่ามากกกว่าหรือเท่ากับ 200 จำนวน 880 ราย

	AIDS (CD4+ cells < 200)	
	Present	Absent

Diagnostic Test	Positive	80	120
	Negative	20	880

Accuracy = *(True positive + True negative) / Total tests*

= *(80+880) / 1,100* = *0.83* = *83%*

Sensitivity = *True positive / Diseases* = *80 / (80+20)* = *0.8* = *80%*

Specificity = *True negative / Non-Disease* = *880 / (880+120) = 0.88= 88%*

PPV = *True positive / All positive* = *80 / (80+120)* = *0.40* = *40%*

NPV = *True negative / All negative* = *880 / (20+880)= 0.98* = *98%*

ตารางแสดงความสัมพันธ์ระหว่างการเจ็บป่วยจริงกับผลตรวจของชุดตรวจวินิจฉัย

		Disease (Gold standard)	
		Present	Absent
Diagnostic Test	Positive	a	b
	Negative	c	d

สรุปวิธีการคำนวนค่าต่าง ๆจากตาราง 2 X 2

Sensitivity	= a / (a+c)	Specificity	= d / (b+d)
PPV	= a / (a+b)	NPV	= d / (c+d)
Accuracy	= (a+d) / (a+b+c+d)	Prevalence	= (a+c) / (a+b+c+d)

Multiple tests

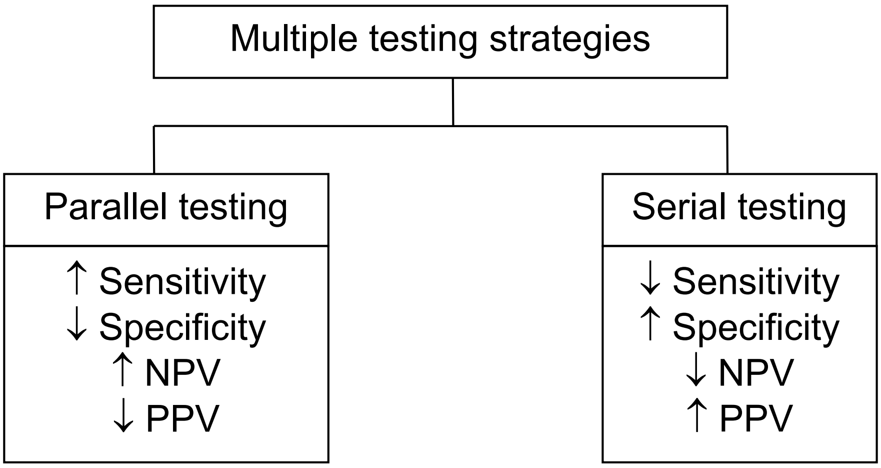
ในการตรวจวินิจฉัยโรค แพทย์มักต้องใช้ชุดตรวจวินิจฉัยหลาย ๆ แบบ เพื่อช่วยเพิ่มความถูกต้องของการวินิจฉัยโรค ซึ่งมีวิธีการ และหลักการในการเลือกใช้ 2 ลักษณะ คือ

1. Parallel testing หมายถึงการใช้ชุดตรวจวินิจฉัยตั้งแต่สองชุดขึ้นไปพร้อมกันในเวลาเดียว การแปลผลคือหากมีชุดใดชุดหนึ่งให้ผลบวกให้ถือว่าเป็นผลบวก แพทย์มักเลือกใช้หลักการนี้ในสถานการณ์ที่ต้องการผลตรวจในเวลาอันรวดเร็ว เช่นผู้ป่วยฉุกเฉิน หรือในหน่วยบริการเคลื่อนที่ ๆ ซึ่งการกลับไปตรวจซ้ำ หรือนัดผู้ป่วยกลับมาฟังผลทำได้ยาก

   เทคนิคการเลือกใช้ชุดตรวจวินิจฉัยแบบ Parallel tests ช่วยเพิ่ม Sensitivity และ NPV ของกระบวนการตรวจวินิจฉัย อย่างไรก็ตามในขณะเดียวกันก็จะไปลด Specificity และ PPV หมายถึงความน่าจะเป็นที่จะพลาดการวินิจฉัยโรค (False negative) จะลดลง แต่โอกาสวินิจฉัยว่าผู้ป่วยเป็นโรคทั้ง ๆ ที่ไม่ได้เป็น (False positive) จะเพิ่มขึ้น

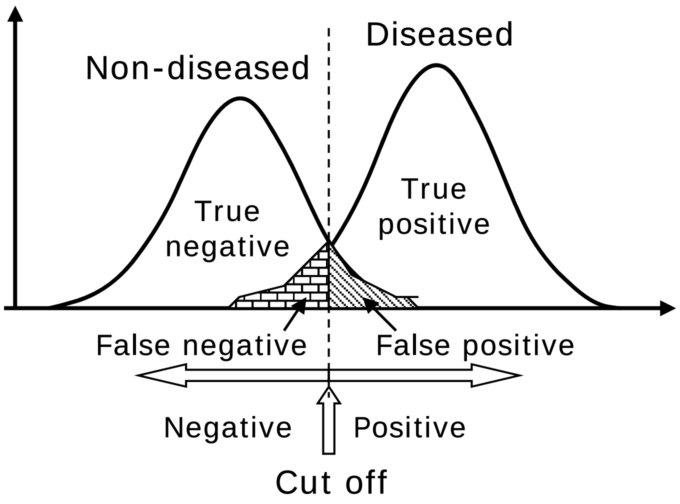
2. Serial testing หมายถึงการใช้ชุดตรวจวินิจฉัยทีละอย่าง การจะใช้ชุดต่อไปหรือไม่ขึ้นกับผลของชุดก่อนหน้า โดยหากชุดแรกให้ผลลบให้รายงานผลเป็นลบได้เลย และหยุดการตรวจแค่นี้ แต่หากชุดแรกให้ผลบวกจึงจะนำชุดตรวจชุดที่สอง ซึ่งเป็นอีกชุดหนึ่งที่ใช้เทคนิคการตรวจแตกต่างออกไปมาตรวจต่อ การจะรายงานว่าเป็นผลบวกก็ต่อเมื่อชุดตรวจทุกชุดที่ใช้ให้ผลบวก แพทย์มักเลือกใช้หลักการนี้ในการดูแลรักษาปกติที่ไม่ได้ต้องการทราบผลตรวจในเวลารวดเร็ว หรือการตรวจอื่นที่จะทำพร้อมกันในลักษณะของ Parallel testing ต้องใช้เวลามากและมีค่าใช้จ่ายสูง

ในทางกลับกัน Serial testing จะช่วยเพิ่ม Specificity และ PPV ของกระบวนการตรวจวินิจฉัย แต่ไปลด Sensitivity และ NPV ด้วยเทคนิคการเลือกชุดตรวจวินิจฉัยแบบนี้แพทย์จะมีความมั่นใจมากขึ้นว่าผู้ป่วย ๆ จริงตามที่ให้การวินิจฉัย อย่างไรก็ตามก็มีโอกาสพลาดการวินิจฉัยมากขึ้น

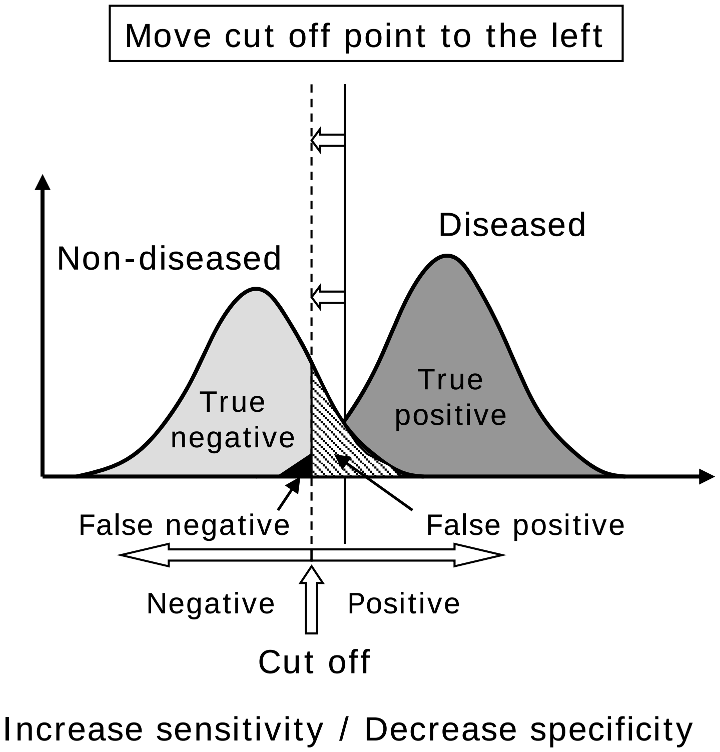


การกำหนดค่า Cut off

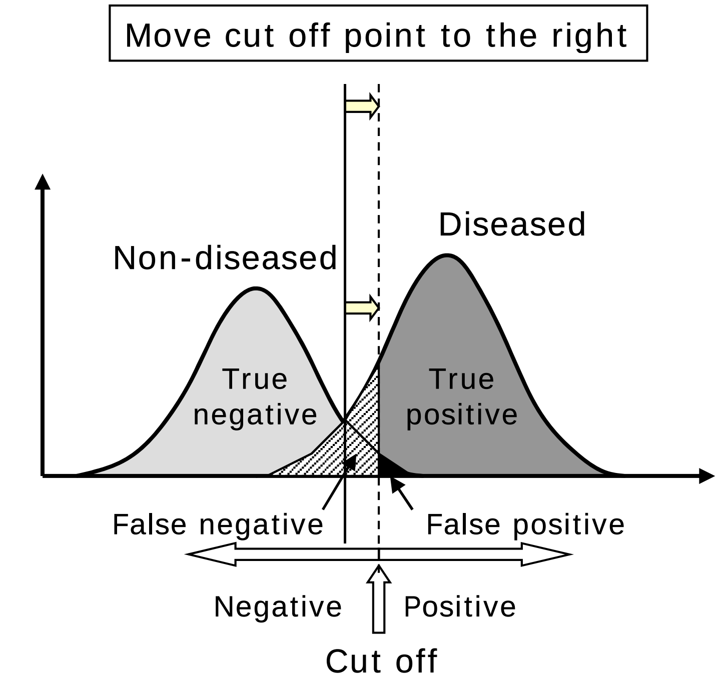
กรณีที่ชุดตรวจวินิจฉัยใช้วิธีวัดข้อมูลบอกปริมาณเป็นตัวเลขที่มีความต่อเนื่อง (Continuous data) แพทย์มักจะต้องการลดความซับซ้อนของข้อมูลลงให้เหลือความเป็นไปได้เพียง 2 อย่าง (Dichotomy) คือ ผิดปกติหรือปกติ เพื่อให้ง่ายต่อการแปลผลและการนำผลที่ได้ไปประกอบการตัดสินใจให้การดูแลรักษาผู้ป่วย ซึ่งทำได้โดยกำหนดค่าตัวเลขที่แบ่งขอบเขตของความปกติกับความผิดปกติขึ้นมาค่าหนึ่ง เรียกว่าค่า Cut off และเนื่องจากการกระจายของผลตรวจในผู้ป่วยและผู้ไม่ป่วย จะมีการทับซ้อนกันเสมอ กรณีที่ค่าของผลตรวจยิ่งสูงยิ่งมีโอกาสเป็นผู้ป่วยจริง เช่นระดับน้ำตาลในเลือดในการวินิจฉัยโรคเบาหวาน ผู้ที่ป่วยเป็นเบาหวานจริงจำนวนหนึ่งจะมีระดับน้ำตาลในเลือดต่ำกว่าผู้ที่ไม่ได้ป่วยเป็นเบาหวานที่มีระดับน้ำตาลค่อนข้างสูง การกำหนดค่า Cut off ณ จุดใดจุดหนึ่ง จะเป็นตัวกำหนดคุณลักษณะของชุดตรวจนั้น ๆ ว่าจะให้ค่า Sensitivity และ Specificity เท่าใด



การตั้งค่า Cut off ไว้ต่ำ (เลื่อนค่า Cut off มาทางซ้าย) จะทำให้ชุดตรวจวินิจฉัยมี Sensitivity สูงขึ้น แต่มี Specificity ต่ำลง หมายถึงที่ค่า Cut off นี้ จะมี False negative น้อยลง แต่ False positive มากขึ้น



ในทางกลับกัน หากเราเลื่อนค่า Cut off ให้สูงขึ้น (เลื่อนค่า Cut off มาทางขวา) จะทำให้ชุดตรวจวินิจฉัยมี Sensitivity ต่ำลง แต่มี Specificity สูงขึ้น คือจะมี False positive น้อยลง แต่ False negative มากขึ้น



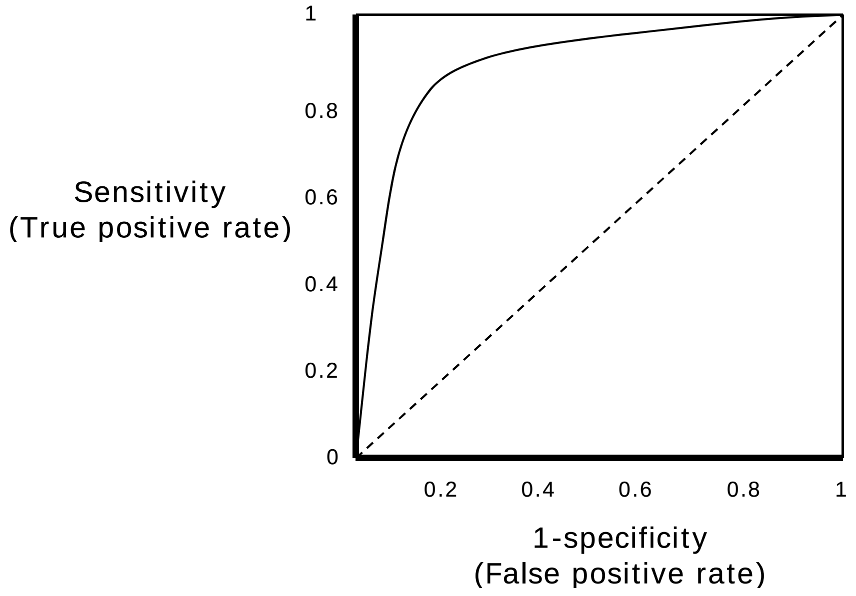
Receiver operator characteristic (ROC) curve

ROC curve เป็นแผนภูมิที่นิยมใช้ในการช่วยกำหนดค่า Cut off point ที่เหมาะสม สำหรับชุดตรวจวินิจฉัยแต่ละชนิด โดยนำค่า Sensitivity (True positive rate) กับ ค่า 1- specificity (False positive rate) ที่ชุดตรวจวินิจฉัยนั้น ๆ สามารถทำได้ ณ จุด Cut off ต่าง ๆ มา plot ในแกน Y และแกน X ตามลำดับ พื้นที่ใต้กราฟแสดงความน่าจะเป็นที่ชุดตรวจวินิจฉัยจะให้ผลที่ถูกต้องตามความจริง ชุดตรวจวินิจฉัยที่ดีจึงมีพื้นที่ใต้กราฟมาก และยอดโค้งของกราฟจะอยู่ใกล้มุมบนซ้ายมากกว่าชุดตรวจวินิจฉัยโรคเดียวกันที่ด้อยกว่า

กรณีที่ไม่ได้ต้องการชุดตรวจวินิจฉัยที่มี Sensitivity หรือ Specificity สูงเพียงอย่างใดอย่างหนึ่งเป็นการเฉพาะ จุด Cut off บนเส้นกราฟที่อยู่ใกล้กับมุมบนซ้ายมากที่สุดจะเป็นจุด Cut off ที่เหมาะสมที่สุด เนื่องจากเป็นจุดที่ชุดตรวจวินิจฉัยจะให้ Accuracy สูงที่สุด แต่หากต้องการชุดตรวจวินิจฉัยที่มี Sensitivity สูง ให้เลือก Cut off ที่อยู่ค่อนมาทางด้านขวาบนแกน X และหากต้องการชุดตรวจวินิจฉัยที่มี Specificity สูง ให้เลือก Cut off ที่อยู่ค่อนมาทางซ้ายของแกน X ตามลำดับ

นอกจากนั้นเรายังสามารถให้ ROC curve เปรียบเทียบคุณภาพของชุดตรวจวินิจฉัยหลาย ๆ แบบที่ใช้วินิจฉัยโรคเดียวกันได้อีกด้วย โดยชุดตรวจวินิจฉัยที่มีมีพื้นที่ใต้กราฟมากกว่า ซึ่งส่วนใหญ่จะมีปลายยอดส่วนโค้งอยู่ใกล้มุมซ้ายมากกว่าจะเป็นชุดตรวจวินิจฉัยที่ดีกว่า

Receiver operator characteristic (ROC) curve



****

เอกสารอ้างอิง

1. Knapp G Rebecca, Miller III M Clinton. Clinical epidemiology and biostatistics. Williams & Wilkins, Baltimore, Maryland. 1992.
2. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
3. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.

****

แบบฝึกหัด

1. ชุดตรวจวินิจฉัยที่มีคุณสมบัติดังข้อใดต่อไปนี้ เหมาะสมที่สุดสำหรับใช้ในการวินิจฉัยยืนยันว่าผู้ป่วยเป็นโรค Multiple sclerosis ซึ่งเป็นโรคร้ายแรง และไม่สามารถรักษาให้หายขาดได้

A. Sensitivity 99%, Specificity 80%

B. Sensitivity 95%, Specificity 85%

C. Sensitivity %90, Specificity 90%

D. Sensitivity 85%, Specificity 95%

E. Sensitivity 80%, Specificity 99%

2. Serial testing เพิ่มค่าใดของ Diagnostic tests

A. Negative predictive value

B. Sensitivity

C. Specificity

D. Accurary

E. Cut off

3. ค่าใดจะสูงขึ้น เมื่อความชุกของโรคในกลุ่มตัวอย่างต่ำลง

A. Accuracy

B. Sensitivity

C. Specificity

D. Negative predictive value

E. Positive predictive value

4. การตรวจคัดกรองวัณโรคเชิงรุกในชุมชน 1,000 ราย พบเสมหะบวก 99 ราย ในจำนวนนี้เป็นวัณโรคจริง 9 ราย ถ้าการตรวจเสมหะมี sensitivity ร้อยละ 45 ให้คำนวณหาความชุกของวัณโรคในชุมชนนี้

A. 0.9%

B. 2.0%

C. 4.5%

D. 9.1%

E. 9.9%

5. Receiver operator characteristic curve เป็นกราฟที่ plot ระหว่าง False positive กับค่าใด

A. Specificity

B. Sensitivity

C. Cut off

D. False negative

E. True negative

6. การตรวจคัดกรองโรค มีค่า Sensitivity, Specificity, PPV, และ prevalence เท่ากับร้อยละ 90, 80, 88, และ 56 ตามลำดับ ถามว่าหาก prevalence เพิ่ม ค่าใจจะเพิ่มขึ้นด้วย

A. Accuracy

B. Sensitivity

C. Specificity

D. PPV

E. NPV

1.11.2.1 Types of experimental studies

ในการศึกษาทางการแพทย์และระบาดวิทยา การวิจัยเชิงทดลอง (Experimental study, Intervention study) หมายถึงการศึกษาที่นักวิจัยเป็นผู้กำหนดว่าอาสาสมัครแต่ละรายจะได้รับ Exposure ต่อปัจจัยที่ต้องการศึกษาหรือไม่ ไม่ได้ปล่อยให้เกิดขึ้นเองตามธรรมชาติ ปัจจัยดังกล่าวเรียกว่าสิ่งทดลองหรือสิ่งแทรกแซง (Intervention) โดยการได้รับสิ่งทดลองนี้ต้องเป็นไปเพื่อวัตถุประสงค์ของการศึกษานั้น ๆ เท่านั้น เมื่อเทียบกับการศึกษาเชิงสังเกต การศึกษาเชิงทดลองให้ผลการศึกษาที่น่าเชื่อถือกว่า เนื่องจากผู้วิจัยสามารถควบคุมสภาพการณ์ต่าง ๆ ได้ ทำให้มีอคติในการสรุปผลการวิจัยน้อยกว่า การพิสูจน์ขั้นสุดท้ายว่าการรักษาโรค หรือการป้องกันโรคใหม่ ๆ ที่คิดขึ้นได้ผลหรือไม่ โดยทั่วไปต้องใช้การศึกษาเชิงทดลองเท่านั้น

ประเภทของการศึกษาเชิงทดลอง แบ่งตามประโยชน์ของสิ่งทดลอง

1. สิ่งทดลองเพื่อการรักษาโรค (Therapeutic trial)

   หรือ Secondary prevention trial หมายถึงการศึกษาเชิงทดลองที่อาสาสมัครเป็นผู้ป่วยด้วยโรคหรือปัญหาสุขภาพที่สนใจ มีวัตถุประสงค์เพื่อประเมินประสิทธิผลของ Intervention ที่ส่วนใหญ่เป็นแนวทางการรักษาใหม่ ๆ เช่นยาหรือเทคนิคการผ่าตัด หรือกระบวนการรักษาที่ใช้เพื่อป้องกันภาวะแทรกซ้อนของโรค เช่นการปรับเปลี่ยนรูปแบบการรับประทานอาหารเพื่อป้องกันการกลับเป็นซ้ำในผู้ป่วยโรคหัวใจขาดเลือด อาสาสมัครต้องได้รับการวินิจฉัยว่าป่วยด้วยโรคที่ต้องการศึกษา และต้องมาพบแพทย์ไม่ช้าเกินไป การดำเนินโรคอยู่ในระยะที่เหมาะสม ไม่ป่วยหนักหรือเบาเกินไป และเนื่องจากเป็นการวิจัยแบบนี้ อาสาสมัครเป็นผู้ป่วยและเกือบทั้งหมดกระทำในสถานพยาบาล จึงถูกเรียกเป็นการทั่วไปว่าการวิจัยเชิงทดลองทางคลินิก (Clinical trials) โดยรูปแบบการวิจัยที่ยอมรับกันว่าเป็นมาตรฐานและใช้บ่อยที่สุด คือ Randomized controlled trial หรือ RCT ซึ่งจะกล่าวถึงต่อไป

2. สิ่งทดลองเพื่อการป้องกันโรค (Preventive trials)

   หรือ Primary prevention trial เป็นการศึกษาเชิงทดลองอีกประเภทหนึ่ง ทำการวิจัยในอาสาสมัครที่แข็งแรงดี ยังไม่เจ็บป่วยเป็นโรคที่สนใจศึกษา มีวัตถุประสงค์เพื่อประเมินประสิทธิผลของ Intervention ว่าสามารถป้องกันการเกิดโรคได้จริงหรือไม่ การศึกษาแบบนี้โดยทั่วไปต้องการขนาดของกลุ่มตัวอย่างมากกว่าการศึกษาที่ Intervention ใช้เพื่อการรักษาโรคหรือป้องกันภาวะแทรกซ้อน เนื่องจากจำนวน Outcome ของการศึกษาซึ่งคือโอกาสการเกิดโรคในผู้ที่ยังไม่ป่วย ต่ำกว่าโอกาสการเกิดความเปลี่ยนแปลงของผลการรักษาโรคในผู้ที่ป่วยแล้ว Prevention trials ยังสามารถแบ่งออกได้เป็น 2 ลักษณะย่อย ๆ ตามหน่วยของการทดลอง ดังนี้ คือ

   2.1 *Field trial* หมายถึง Preventive trial ที่หน่วยของการทดลองเป็นบุคคลแต่ละคน ที่เรียกว่า Field trials เนื่องจากเป็นการศึกษาในประชากรที่ยังไม่ป่วยเป็นโรค จึงมักต้องออกไปศึกษานอกสถานพยาบาล เช่นที่ทำงาน ที่บ้าน หรือที่โรงเรียน ตัวอย่างของการศึกษาแบบนี้ ได้แก่ การศึกษาประสิทธิผลของวัคซีนป้องกันโรคต่าง ๆ

   2.2 *Community intervention trial* หมายถึง Preventive trial ที่หน่วยของการทดลองเป็นกลุ่มคน โดยทั่วไปหมายถึงการวิจัยเปรียบเทียบผลของมาตรการป้องกันโรคในชุมชนที่ได้รับสิ่งทดลอง กับชุมชนที่ไม่รับสิ่งทดลอง เนื่องจากด้วยลักษณะของการศึกษา จะเป็นการง่ายกว่าที่จะให้ Intervention และวัด Outcome ในระดับประชากร เช่นการศึกษาผลของการป้องกันฟันผุจากการเติมฟลูออไรด์ใน

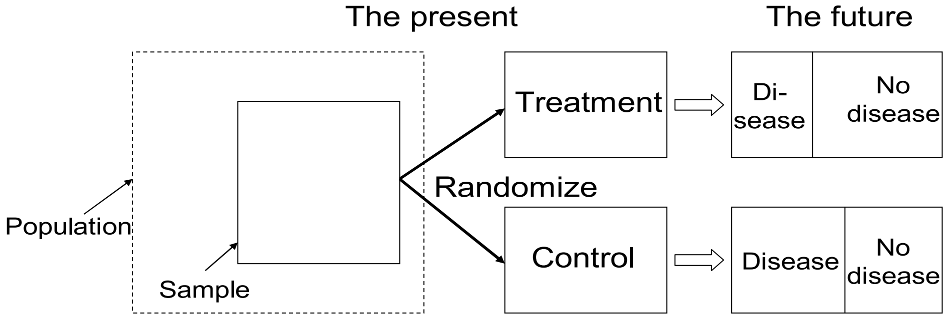
น้ำดื่ม อย่างไรก็ตาม สิ่งทดลองบางอย่าง เหมาะที่จะทำการศึกษาเปรียบเทียบในกลุ่มประชากรที่เล็กกว่าชุมชน เช่น การวิจัยเกี่ยวกับโภชนาการเหมาะที่จะทำในระดับครอบครัว การประเมินผลของการอุปกรณ์ป้องกันอันตรายจากการทำงานเหมาะที่จะเปรียบเทียบในระดับโรงงาน หรืออุปกรณ์ป้องกันการบาดเจ็บจากการเล่นกีฬา เหมาะที่จะเปรียบเทียบในระดับทีมกีฬาเป็นต้น Community intervention trial ที่ดี ควรมีการ Randomization เช่นเดียวกับ RCT และ Field trials แต่จะเป็นการสุ่มเลือกในระดับกลุ่มคน เรียกว่า Cluster randomization อย่างไรก็ตาม การปิดบัง (Blind) ไม่ให้อาสาสมัครทราบว่าตนเองอยู่ในกลุ่มที่ได้รับ Intervention หรืออยู่ในกลุ่มเปรียบเทียบมักทำได้ยาก เนื่องจากข้อจำกัดของสิ่งทดลองที่ไม่สามารถปิดบังได้

Randomized controlled trial (RCT)

คือรูปแบบของการศึกษาเชิงทดลองที่ถือเป็นมาตรฐานทอง (Gold standard) ในปัจจุบัน การวิจัยเพื่อหาประสิทธิผลของยาและวัคซีนส่วนใหญ่ใช้รูปแบบการศึกษาแบบนี้ มีลักษณะสำคัญคือ

1. อาสาสมัครจะถูกแบ่งเป็นอย่างน้อยสองกลุ่ม คือ กลุ่มที่ได้รับ Intervention เรียกว่ากลุ่มศึกษา (Treatment group หรือ Intervention group) และกลุ่มที่ไม่ได้รับ Intervention เรียกว่ากลุ่มควบคุมหรือกลุ่มเปรียบเทียบ (Control group)
2. การที่อาสาสมัครแต่ละรายจะได้อยู่ในกลุ่มใดต้องเป็นไปโดยกระบวนการสุ่ม (Randomization) ตามโอกาสความน่าจะเป็นตามธรรมชาติ ผู้ศึกษาไม่มีส่วนเลือกหรือกำหนดว่าอาสาสมัครรายใดจะได้อยู่กลุ่มไหน เป็นการลดอคติของการศึกษาโดยเฉพาะอย่างยิ่ง Confounding Randomization ถือเป็นคุณสมบัติที่สำคัญที่สุดของ RCT ที่ทำให้ผลการศึกษาน่าเชื่อถือกว่าการศึกษารูปแบบอื่น ๆ
3. ระหว่างทำการวิจัย ตลอดจนระหว่างการวัดผลการศึกษา ทั้งผู้วิจัย และอาสาสมัครต้องไม่ทราบว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มทดลองหรืออยู่ในกลุ่มเปรียบเทียบ เพื่อลดอคติ (bias) ในการศึกษา เรียกวิธีการนี้ว่า Double-blind technique บางครั้งจึงเรียก RCT ว่า Randomized double blinded trial

Randomized controlled trial



1. Select a sample from the population
2. Measure baseline variables
3. Randomize
4. Apply interventions (one should be a blinded placebo)
5. Follow up the cohorts
6. Measure outcome variables (blinded)

การจัดแบ่งประเภทของการศึกษาเชิงทดลองทำได้หลายแบบ ในที่นี้ขอนำเสนอ 3 ลักษณะคือ แบ่งตามประโยชน์ของสิ่งทดลองที่ใช้ในการวิจัย แบ่งตามรูปแบบงานวิจัย และแบ่งตามระยะของงานวิจัย

ประเภทของการศึกษาเชิงทดลอง แบ่งตามรูปแบบงานวิจัย

นอกเหนือจาก RCT ซึ่งถือเป็นรูปแบบการศึกษาเชิงทดลองมาตรฐานแล้ว ยังมีรูปแบบย่อยอื่น ๆ ของการศึกษาเชิงทดลองอีกหลายแบบ บางแบบเป็นส่วนขยายของ RCT และยังถือเป็น RCT อยู่ บางแบบเป็นเพียงการศึกษาเชิงทดลอง แต่ไม่ถือเป็น RCT เนื่องจากขาดลักษณะสำคัญของ RCT ข้างต้น ในที่นี้ขอกล่าวถึงรูปแบบสำคัญที่ใช้บ่อย ดังนี้

1. Run-in design

ในการศึกษารูปแบบนี้ ผู้วิจัยจะลองให้อาสาสมัครรับยาหลอก (Placebo) ไปเป็นระยะเวลาหนึ่งก่อน (ระยะ run in) ซึ่งโดยทั่วไปจะใช้เวลานานประมาณ 2-3 สัปดาห์ เพื่อสังเกตความร่วมมือ หลังจากนั้นเฉพาะอาสาสมัครที่รับประทานยาอย่างสม่ำเสมอและสามารถปฏิบัติตัวได้ตามที่โครงการกำหนด จึงจะถูกเชิญเข้าสู่การศึกษาในระยะเต็มรูปแบบ (full study) ประโยชน์คือเป็นการเพิ่มสัดส่วนของอาสาสมัครที่สามารถร่วมโครงการจนครบตามกำหนด และช่วยเพิ่มโอกาสในการตรวจพบความแตกต่างของผลการรักษาระหว่างกลุ่มทดลองกับกลุ่มเปรียบเทียบ (Statistical power)

2. Factorial design

เป็นรูปแบบการศึกษาที่ช่วยให้ผู้วิจัยสามารถตอบคำถามวิจัยได้ 2 ข้อ หรือมากกว่าพร้อมกันไปในการศึกษาเดียว จึงเป็นรูปแบบการวิจัยที่มีประสิทธิภาพสูง ช่วยประหยัดเวลา และค่าใช้จ่าย

ตัวอย่าง

คำถามวิจัยแรก ผลของยาแอสไพรินลดการเกิดโรคหัวใจขาดเลือด คำถามวิจัยที่สอง β-carotene ลดการเกิดมะเร็ง เราสามารถทำการศึกษาแบบ Factorial design เพื่อตอบคำถามวิจัยทั้งสองข้อพร้อมกัน โดยการสุ่มแบ่งอาสาสมัครออกเป็น 4 กลุ่ม ดังนี้

กลุ่มที่ 1 ได้รับ ยาแอสไพริน และ β-carotene
กลุ่มที่ 2 ได้รับ ยาแอสไพริน และ Placebo B
กลุ่มที่ 3 ได้รับ Placebo A และ β-carotene
กลุ่มที่ 4 ได้รับ Placebo A และ Placebo B

เพื่อตอบคำถามวิจัยแรก ผู้ศึกษาจะเปรียบเทียบ Outcome ของกลุ่ม 1+2 (Intervention) กับกลุ่ม 3+4 (Control) และเพื่อตอบคำถามวิจัยที่สอง ผู้ศึกษาจะเปรียบเทียบ Outcome ของกลุ่ม 1+3 (Intervention) กับ กลุ่ม 2+4 (Control)

ข้อจำกัดของการศึกษารูปแบบนี้คือ ตัวแปรต้นของแต่ละคำถามวิจัยต้องไม่มีความสัมพันธ์กับตัวแปรตามของอีกคำถามวิจัยหนึ่ง สำหรับตัวอย่างข้างต้นหมายถึง ยาแอสไพรินต้องไม่มีผลต่อการเกิดมะเร็งและ β-carotene ต้องไม่มีผลต่อการป่วยเป็นโรคหัวใจขาดเลือด มิฉะนั้นอาจได้ผลการศึกษาที่ผิดพลาดได้

3. Nonrandomized between-group designs

บางครั้งเรียกว่า Quasi-experiment เป็นการศึกษาเชิงทดลองที่มีกลุ่มทดลองและกลุ่มเปรียบเทียบเช่นเดียวกับ RCT แต่การกำหนดว่าอาสาสมัครจะอยู่ในกลุ่มทดลองหรือกลุ่มเปรียบเทียบไม่ได้เกิดจากการสุ่ม (non-randomized) ซึ่งอาจเนื่องด้วยข้อจำกัดบางประการ หรือเพื่อความสะดวก เช่นผู้วิจัยเลือกให้ Intervention ที่เป็นหลักสูตรเพศศึกษาแก่นักเรียนในบางโรงเรียนและไม่ให้ในบางโรงเรียน โดยไม่ได้สุ่มตัวอย่าง เนื่องจากความร่วมมือจากอาจารย์ใหญ่ของบางโรงเรียนมากกว่า หรือแพทย์ให้ผู้ป่วยเบาหวานเป็นผู้เลือกเองว่าจะอยู่ในกลุ่มศึกษาคือต้องฉีด Insulin วันละ 4 ครั้ง หรือจะอยู่ในกลุ่มเปรียบเทียบ คือฉีด Long acting Insulin เพียงวันละครั้ง เป็นต้น การศึกษารูปแบบนี้ถือว่าด้อยกว่า RCT เนื่องจากมีโอกาสเกิดเกิดอคติกับผลการศึกษามากกว่า

4. Time series design

เป็นการศึกษาที่มีอาสาสมัครเพียงกลุ่มเดียว โดยเป็นทั้งกลุ่มทดลองและกลุ่มเปรียบเทียบในตัวเอง แบ่งการศึกษาออกเป็น 2 ระยะ คือระยะที่อาสาสมัครได้รับสิ่งทดลองและระยะไม่ได้รับสิ่งทดลอง แล้ววัดผลที่คาดว่าจะเกิดในแต่ละระยะเปรียบเทียบกัน ข้อดีคือปัจจัยกวนที่อาจเกิดจากความแตกต่างกันของการมีอาสาสมัคร 2 กลุ่ม เช่น เพศ อายุ และปัจจัยทางพันธุกรรม จะถูกกำจัดออกไป และค่าใช้จ่ายในการศึกษาต่ำกว่า เหมาะสำหรับใช้ในกรณีที่สิ่งทดลองส่งผลในระยะเวลาสั้น และผลนั้นสามารถกลับคืนได้เมื่อหยุดให้สิ่งทดลอง เช่นการศึกษาผลของแอลกอฮอล์ต่อระดับของ HDL-Cholesterol รูปแบบงานวิจัยแบบนี้ยังมีการแบ่งย่อยออกไปอีก เช่นการมีการวัดหลาย ๆ ครั้งก่อน และหลังการให้ Intervention เพื่อให้แน่ใจว่าความเป็นแปลงของ Outcome เกิดจากตัว Intervention

ปัญหาที่สำคัญ 2 ประการของการศึกษาแบบนี้ได้แก่ ปัจจัยกวนเนื่องจากระยะเวลาที่ผ่านไป (Time dependent confounding variables) เช่น ผลของฤดูกาลที่เปลี่ยนแปลงต่อการดำเนินโรค ซึ่งอาจให้ผลที่แตกต่างกันระหว่างการศึกษาระยะแรกกับระยะหลัง และผลตกค้างของ Intervention ที่อาจหลงเหลืออยู่ (Carryover effect) ขณะที่อาสาสมัครถูกระบุว่าอยู่ในระยะเปรียบเทียบแล้ว เช่นผลของยาขับปัสสาวะต่อระดับความดันโลหิตจะยังคงอยู่แม้จะหยุดยาไปแล้วเป็นเวลานาน เป็นต้น

5. Crossover design

ลักษณะคล้ายคลึงกับ Time series designs คืออาสาสมัครแต่ละคนทำหน้าที่เป็นกลุ่มควบคุมของตนเองตามระยะเวลาต่าง ๆ แต่มีลักษณะต่างกันตรงที่อาสาสมัครจะถูกสุ่มแบ่งเป็นสองกลุ่ม โดยกลุ่มหนึ่งจะเป็นกลุ่มทดลองก่อน ขณะที่อีกกลุ่มหนึ่งจะเป็นกลุ่มเปรียบเทียบก่อน หลังจากนั้นจะสลับไปเป็นอีกกลุ่มหนึ่งเมื่อครบระยะแรกแล้ว ประโยชน์คือช่วยตัดปัญหาเรื่อง Time dependent confounding variables ที่เกิดใน Time series designs อย่างไรก็ตาม ปัญหาเรื่อง Carryover effect จะซับซ้อนขึ้น ทำให้ยากต่อการวิเคราะห์ข้อมูล

6. Natural experiment

ตำราบางเล่มไม่จัด Natural experiments เป็นการศึกษาเชิงทดลองแต่จัดเป็นการเชิงศึกษาเชิงสังเกต เนื่องจากผู้วิจัยไม่ได้เป็นผู้กำหนดให้มี Intervention ที่ต้องการศึกษาในกลุ่มตัวอย่างด้วยตนเอง แต่อาศัยสังเกตผลที่เกิดจากการกระทำของผู้อื่น ตัวอย่างเช่นการศึกษาผลของการบังคับใช้กฎหมายหมวกนิรภัยในผู้ขับขี่รถจักรยานยนต์กับอุบัติการณ์ของการเกิดการบาดเจ็บที่ศีรษะ หรือผลของรังสีจากระเบิดปรมาณูต่อสุขภาพของมนุษย์เป็นต้น ซึ่งหากพิจารณาตามรูปแบบการวิจัยแล้ว อันที่จริง Natural experiment เข้าได้กับ Time series design

ประเภทของการศึกษาเชิงทดลอง แบ่งตามระยะของการวิจัย (Phases of clinical trials)

การศึกษาเชิงทดลอง โดยเฉพาะอย่างยิ่ง การศึกษาเกี่ยวยารักษาโรคตัวใหม่ หรือวัคซีนป้องกันโรค แบ่งได้เป็น 4 ระยะ ดังนี้ คือ

- **Phases I** First experiment in man มีวัตถุประสงค์เพื่อตรวจสอบความปลอดภัยของการใช้ยาในคน และเพื่อหาขอบเขตของขนาดของยา (Dose range) ที่ปลอดภัย การทดลองจะทำในอาสาสมัครที่มีร่างกายปกติ ไม่มีกลุ่มควบคุม และใช้ตัวอย่างจำนวนน้อย มักไม่เกิน 50 ราย
- **Phases II** Small-scale clinical trial เป็นการศึกษาขยายจากการศึกษาในระยะที่หนึ่ง เพื่อทราบขนาดของยาที่เหมาะสม (PhaseIIA) และประสิทธิผลของยาในการรักษาโรค (Phase IIB) ในการวิจัยระยะนี้ อาสาสมัครจะเป็นผู้ป่วยด้วยโรคที่ต้องการศึกษา มีจำนวนมากกว่าการทดลองระยะที่หนึ่ง การทดลองระยะที่สองนี้อาจมีกลุ่มควบคุม หรือการ Blind หรือไม่ก็ได้
- **Phases III** Full-scale clinical trial เป็นการศึกษาขั้นสรุป ที่จะบอกถึงประสิทธิผลของยาในการใช้รักษาโรคและผลข้างเคียงที่พบบ่อย จำนวนอาสาสมัครต้องมากพอที่จะบอกความแตกต่างทางคลินิกอย่างมีนัยสำคัญทางสถิติระหว่างกลุ่มทดลองกับกลุ่มเปรียบเทียบ มักเป็นการทดลองแบบ RCT เมื่อสิ้นสุดการศึกษาในระยะนี้แล้ว หากยามีประสิทธิผลและมีความปลอดภัย คณะกรรมการอาหารและยา จะอนุมัติให้จดทะเบียนและออกวางขายได้ โดยทั่วไปเมื่อกล่าวถึงการทดลองทางคลินิกจะหมายถึงการศึกษาในระยะที่สามนี้
- **Phase IV** Post-marketing surveillance เนื่องจากการศึกษาระยะที่สามยังมีขนาดตัวอย่างไม่มากพอที่จะค้นพบผลข้างเคียงที่พบไม่บ่อย และระยะเวลาในการทดลองไม่นานพอที่จะพบผลข้างเคียงระยะยาว (Long term side effects) จึงมีความจำเป็นที่จะต้องติดตามผลการใช้ยาในกลุ่มประชากรทั่วไปอีกระยะหนึ่งว่ายามีความปลอดภัยจริงหรือไม่ และเพื่อทราบถึงประสิทธิภาพของยาในสถานการณ์การใช้จริง

****

เอกสารอ้างอิง

1. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
2. Rothman J Kenneth, Greenland Sander. Modern Epidemiology 2nd edition. Lippincott-Raven, 1998. Philadelphia, USA.
3. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.
4. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.

****

แบบฝึกหัด

1. แพทย์ศึกษาผลของยาลดความดันโลหิตตัวใหม่ โดยเปรียบเทียบความดันโลหิตของอาสาสมัครจำนวน 5 0 คน ก่อนและหลังได้รับยา การศึกษาข้างต้นจัดเป็นการศึกษารูปแบบใด

A. Run-in design

B. Factorial design

C. Crossover design

D. Quasi-experiment

E. Time series design

2. ประสิทธิผลของยาแต่ละตัว จะถูกศึกษาครั้งแรกใน Clinical trial ระยะใด

A. 0

B. 1

C. 2

D. 3

E. 4

3. การศึกษาแบบ Randomized controlled trial (RCT) มีโครงสร้างคล้ายกับการศึกษาแบบใดมากที่สุด?

A. Case series

B. Cross-sectional

C. Case-control

D. Retrospective cohort

E. Prospective cohort

4. ยา Nanosum ได้รับอนุญาตให้วางขาย ต่อมาพบว่าคนใช้ยาได้รับผลข้างเคียงและถูกระงับการขาย ถือว่าเป็นผลจากขั้นตอนใดของการวิจัยเชิงทดลอง

A. Phase I

B. Phase IIa

C. Phase IIb

D. Phase III

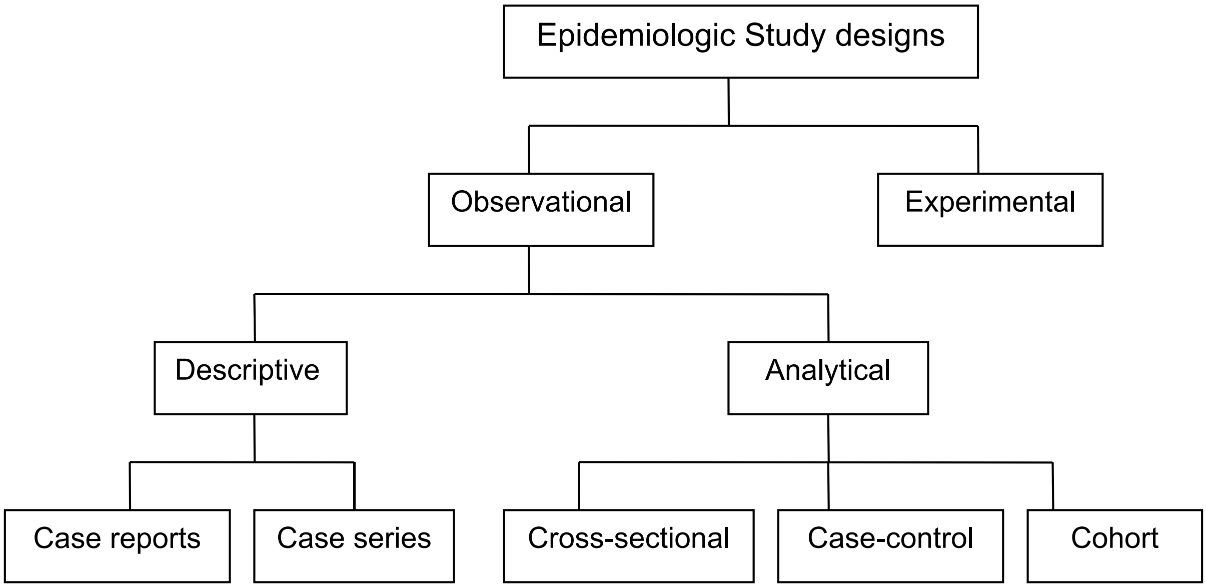
E. Phase IV

1.11.2.2 Types of observational studies

ผศ.นพ. เกรียงไกร ศรีธนวิบุญชัย

การจัดกลุ่มประเภทของศึกษาทางระบาดวิทยาทำได้หลายแบบ ในที่นี้จะเริ่มโดยการแบ่งการศึกษาทั้งหมดออกเป็น 2 กลุ่มใหญ่ ๆ คือการศึกษาเชิงสังเกต (Observational studies และการศึกษาเชิงทดลอง (Experimental studies)

แผนภูมิแสดงการแบ่งประเภทของการศึกษาทางระบาดวิทยา



การศึกษาเชิงทดลอง (Experimental study)

คือการวิจัยที่ผู้วิจัยเป็นคนควบคุมและกำหนดว่ากลุ่มตัวอย่างจะได้รับหรือไม่ได้รับปัจจัยที่เป็นสิ่งทดลอง (Intervention) หรือไม่ โดยทั่วไปจะแบ่งอาสาสมัครเป็น 2 กลุ่มโดยการสุ่ม (Random allocation) กลุ่มแรกเป็นกลุ่มทดลอง (Experimental group) กลุ่มนี้จะได้รับปัจจัยที่ต้องการศึกษา กลุ่มที่สองเป็นกลุ่มเปรียบเทียบ (Control group) กลุ่มนี้จะไม่ได้รับปัจจัยที่ต้องการศึกษา ปัจจัยนี้อาจเป็นยา วัคซีน กระบวนการป้องกันโรค หรืออื่น ๆ จากนั้น ผู้ศึกษาจะวัดผลลัพธ์ที่เกิดขึ้นในทั้งสองกลุ่มเปรียบเทียบกัน ว่ามีความแตกต่างกันหรือไม่ การศึกษาเชิงทดลองจึงเป็นการศึกษาเชิงวิเคราะห์เสมอ

การศึกษาเชิงทดลองเป็นการศึกษาที่ให้ผลการการศึกษาที่น่าเชื่อถือที่สุด อย่างไรก็ตาม เนื่องจากทำได้ยาก ค่าใช้จ่ายสูง และมีข้อจำกัดมากโดยเฉพาะอย่างยิ่งด้านจริยธรรม จึงสงวนไว้สำหรับเฉพาะในกรณีที่ปัญหาวิจัยไม่สามารถใช้การศึกษาโดยการสังเกต หรือเคยมีการศึกษาโดยวิธีการสังเกตแล้วแต่ยังไม่สามารถหาคำตอบที่ชัดเจนได้ การศึกษาโดยการทดลองที่ทำกันมาก และถือเป็นมาตรฐานทอง (Gold standard) ของการทำวิจัย คือการศึกษาที่เรียกว่า Randomized controlled trial

สำหรับเอกสารคำสอนฉบับนี้ จะเน้นรายละเอียดไปที่รูปแบบต่าง ๆ ของการศึกษาเชิงสังเกต ซึ่งเป็นประเภทของการวิจัยที่ทำกันมากในทางระบาดวิทยา

การศึกษาเชิงสังเกต (Observational studies)

คือการศึกษาที่ผู้วิจัยไม่ได้เป็นผู้เข้าไปกำหนดว่าประชากรศึกษา จะได้หรือไม่ได้รับปัจจัยที่สนใจ โดยปล่อยให้การได้รับหรือไม่ได้รับปัจจัยดังกล่าวเป็นไปตามธรรมชาติ ผู้วิจัยเพียงแต่สังเกตและวัดปรากฏการณ์ที่เกิดขึ้นเท่านั้น การศึกษาโดยการสังเกตยังมีข้อจำกัดทางด้านศีลธรรมน้อยกว่า และมีค่าใช้จ่ายในการทำการศึกษาต่ำกว่าการศึกษาโดยการทดลอง การศึกษาเชิงสังเกตสามารถแบ่งได้เป็น 2 ลักษณะใหญ่ ๆ คือ การศึกษาเชิงพรรณนา และการศึกษาเชิงวิเคราะห์

การศึกษาเชิงพรรณนา (Descriptive studies)

การศึกษาระบาดวิทยาเชิงพรรณนาเป็นการบรรยายลักษณะของการเจ็บป่วยในประชากร บอกถึงขนาดของปัญหาที่สนใจ โดยไม่มีการหาความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงกับการเกิดการเจ็บป่วย ในกรณีของโรคใหม่ที่ยังไม่ทราบสาเหตุหรือปัจจัยเสี่ยงที่ชัดเจน การศึกษาเชิงพรรณนามีประโยชน์ในการรวบรวมข้อมูลที่อาจนำไปสู่การสร้างสมมติฐานว่าปัจจัยใดอาจเป็นสาเหตุของการเกิดโรค และนำไปสู่การศึกษาเชิงวิเคราะห์ต่อไป การศึกษาระบาดวิทยาเชิงพรรณนามักดำเนินการก่อนการศึกษาเชิงวิเคราะห์และเป็นการศึกษาเชิงสังเกตเสมอ ลักษณะที่สำคัญ 3 ประการ ในการบรรยายลักษณะการเกิดโรคในการศึกษาเชิงพรรณนา ได้แก่

- <u>บุคคล (Person)</u> ลักษณะที่สำคัญได้แก่ อายุ เพศ สถานภาพสมรส เชื้อชาติ อาชีพ และฐานะทางสังคมเศรษฐกิจ
- <u>สถานที่ (Place)</u> ระบุสถานที่และการกระจายของโรค ซึ่งมีได้ตั้งแต่ระดับอาคาร ท้องถิ่น อำเภอ จังหวัด ประเทศ และระหว่างประเทศ
- <u>เวลา (Time)</u> ข้อมูลเกี่ยวกับการเกิดโรคที่สัมพันธ์กับเวลา มีความสำคัญช่วยบอกลักษณะการเกิดโรคในชุมชนว่ามีลักษณะเป็นแบบใด การเปลี่ยนแปลงเกี่ยวกับเวลา แบ่งออกได้เป็น 3 อย่าง คือ การเปลี่ยนแปลงระยะยาว (Secular trends) การเปลี่ยนแปลงเป็นรอบ (Cyclic fluctuations) และการเปลี่ยนแปลงระยะสั้น (Short-term fluctuations)

การศึกษาเชิงพรรณนาที่เป็นรูปแบบเฉพาะ และเกี่ยวข้องกับทางคลินิกมี 2 ลักษณะ คือ Case report และ Case series

1. ***Case report*** คือการนำเสนอรายละเอียดทางคลินิกของผู้ป่วยรายใดรายหนึ่งที่มีลักษณะเฉพาะน่าสนใจเป็นพิเศษ โดยอาจเป็นการเจ็บป่วยด้วยโรคใหม่ที่ไม่เคยพบมาก่อน หรือเป็นการเจ็บป่วยด้วยโรคเดิมแต่มีลักษณะพิเศษจากการเจ็บป่วยทั่วไป หรือเป็นการรายงานถึงการค้นพบการรักษาใหม่ที่ได้ผลดีในผู้ป่วย ตำราบางเล่มไม่จัด Case report เป็นการศึกษาทางระบาดวิทยาเนื่องจากไม่ได้ศึกษาในกลุ่มผู้ป่วยหรือระดับประชากร Case report ที่เป็นการทดลองการรักษาหรือการป้องกันโรคจัดเป็นการศึกษาเชิงทดลอง (Experimental study)
2. ***Case series*** คือการรายงานลักษณะการเจ็บป่วย การดำเนินโรค ของกลุ่มผู้ป่วยที่มีอาการเจ็บป่วยคล้ายคลึงกัน โดยผู้รายงานสันนิฐานว่าเกิดจากสาเหตุเดียวกัน ส่วนใหญ่จะเป็นกลุ่มอาการหรือโรคใหม่ ที่ไม่เคยปรากฏมาก่อน หรือเป็นโรคที่พบน้อยมาก (Rare diseases) โดยทั่วไปจะบอกถึงระยะเวลาที่พบผู้ป่วยด้วย ส่วนใหญ่ผู้ป่วยจะถูกรวบรวมมาจากที่เดียวกัน (สถานพยาบาลเดียวกัน) แต่อาจรวบรวมจากต่างสถานที่กันก็ได้ ลักษณะที่มีร่วมกันของผู้ป่วยใน Case series จะนำไปสู่การตั้งสมมติฐาน

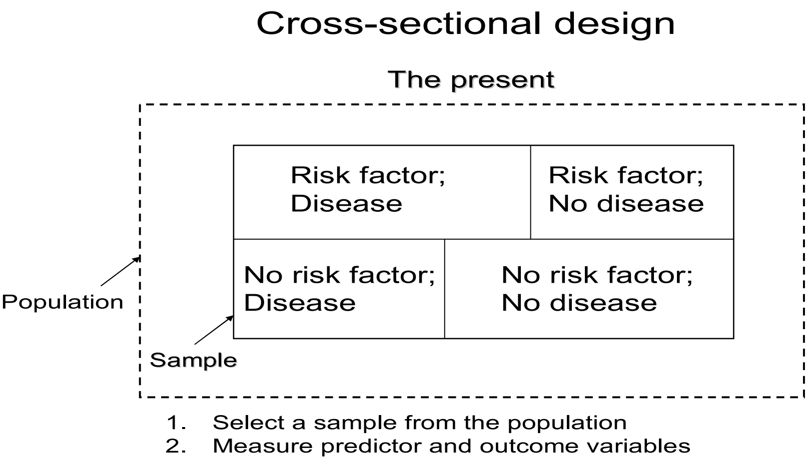
ถึงสาเหตุของโรคต่อไป เช่นเดียวกัน Case report กรณีที่เป็นการรายงาน Case series ที่เป็นการทดลองการรักษาหรือการป้องกันโรค Case series ดังกล่าวจะจัดเป็นการศึกษาเชิงทดลอง (Experimental study)

การศึกษาเชิงวิเคราะห์ (Analytical study)

การศึกษาระบาดวิทยาเชิงวิเคราะห์เป็นการทดสอบสมมติฐานเพื่อหาความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงกับการเกิดโรค มักดำเนินการหลังจากมีข้อมูลเชิงพรรณนามากเพียงพอแล้ว การศึกษาเชิงวิเคราะห์ยังสามารถแยกย่อยออกเป็นการศึกษาเชิงสังเกต (Observational study) และการศึกษาเชิงทดลอง (Experimental study)

1. ***Cross-sectional study*** เป็นการวัดการมีปัจจัยเสี่ยงและป่วยเป็นโรคในประชากรศึกษา ณ จุดใดจุดหนึ่งของเวลาไปพร้อม ๆ กัน หากพบว่าผู้ที่ป่วยเป็นโรคที่สนใจมีปัจจัยเสี่ยงที่สงสัยว่าจะเป็นสาเหตุของโรคสูงกว่ากลุ่มผู้ที่ไม่ป่วยก็มีแนวโน้มว่าปัจจัยเสี่ยงนั้นอาจเป็นสาเหตุของโรค การศึกษาแบบ Cross-sectional study เป็นการศึกษาที่ทำกันบ่อย เนื่องจากทำง่าย ใช้เวลาและค่าใช้จ่ายน้อย เนื่องจากเก็บข้อมูลเพียงครั้งเดียว บางครั้งเรียกว่าการศึกษาเชิงสำรวจ (Survey study) ข้อมูลอันเป็นประโยชน์ที่ได้จากการศึกษาแบบนี้ คือความชุกของโรค (Prevalence) อย่างไรก็ตาม ข้อมูลเชิงความสัมพันธ์ระหว่างปัจจัยเสี่ยงและการเกิดโรคเชื่อถือได้น้อยกว่าการวิจัยรูปแบบอื่น เนื่องจากการศึกษาแบบนี้วัดขนาดของปัจจัยเสี่ยงและการเป็นโรคในเวลาเดียวกัน ทำให้ไม่สามารถบอกได้ว่าปัจจัยเสี่ยงนั้น มีมาก่อนที่ผู้ป่วยจะป่วยเป็นโรคและเป็นสาเหตุของโรคจริงหรือไม่

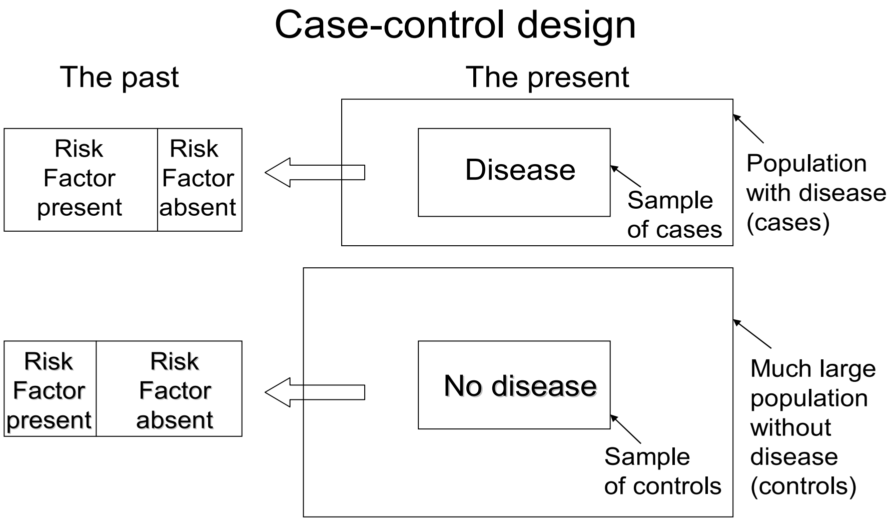
แผนภูมิแสดงขั้นตอนการศึกษาแบบ Cross-sectional studies



2. ***Case-control study*** โดยทั่วไปเป็นการศึกษาในสถานพยาบาล เริ่มจากมีกลุ่มผู้ป่วยด้วยโรคที่สนใจศึกษาหาสาเหตุจำนวนหนึ่ง จากนั้นผู้วิจัยจะหากลุ่มคนอีกกลุ่มหนึ่งซึ่งลักษณะอื่น ๆ คล้ายคลึงกับกลุ่มผู้ป่วยแต่ไม่ป่วยเป็นโรคดังกล่าวเรียกว่ากลุ่มเปรียบเทียบ (Control) มาเทียบ จากนั้นศึกษาประวัติการมีหรือไม่มีปัจจัยเสี่ยงที่สงสัยว่าอาจเป็นสาเหตุของโรคในกลุ่มคนทั้งสองกลุ่มในอดีต อาจโดยสอบถามหรือตรวจสอบจากประวัติ อัตราส่วนระหว่าง Odds ของการมีปัจจัยเสี่ยงของกลุ่มผู้ป่วยกับ Odds ของการมีปัจจัยเสี่ยงในกลุ่มเปรียบเทียบ เรียกว่า Odds Ratio (OR) ซึ่งบอกถึงความสัมพันธ์ระหว่างการมีปัจจัยเสี่ยงนั้น ๆ กับการเกิดโรค Case-control study เป็นการศึกษาที่เหมาะสำหรับการศึกษาสาเหตุของโรคที่พบไม่บ่อย (Rare diseases) มีประสิทธิภาพสูงเนื่องจากต้องการขนาดกลุ่มตัวอย่างน้อย ใช้เวลาตลอดจนค่าใช้จ่ายในการดำเนินการไม่มาก ข้อจำกัดที่สำคัญของการศึกษาแบบนี้ คือการอธิบายความสัมพันธ์เชิง

สาเหตุระหว่างปัจจัยเสี่ยงกับการเกิดโรค จะทำได้ได้ไม่ดีเท่ากับการศึกษาแบบ Cohort studies และมีโอกาสเกิดอคติ (Bias) ในการสรุปผลการวิจัยสูง เนื่องจากข้อมูลการมีหรือไม่มีปัจจัยเสี่ยงเป็นข้อมูลในอดีตที่อาจไม่ตรงตามความจริง นอกจากนั้นหากเลือกกลุ่มเปรียบเทียบไม่เหมาะสมก็อาจมีปัจจัยอื่น ๆ มากวนทำให้สรุปความสัมพันธ์ผิดพลาดได้

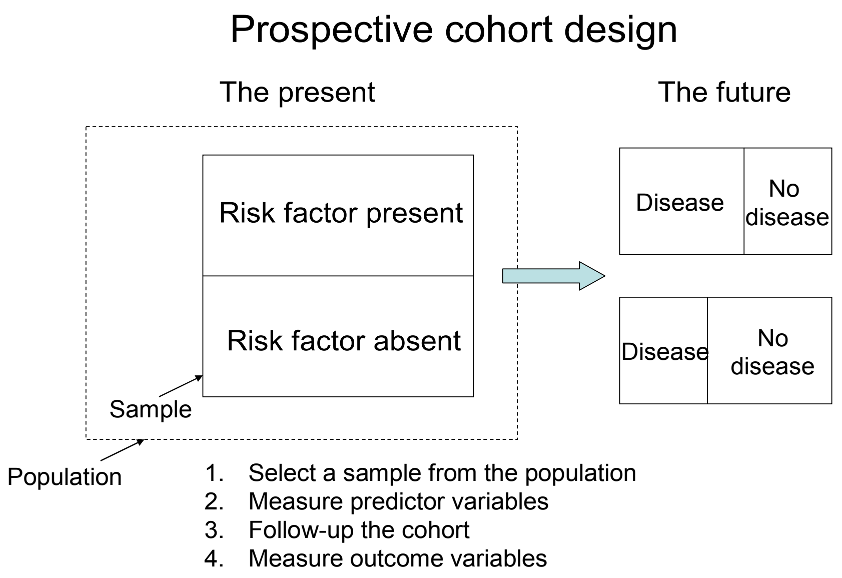
แผนภูมิแสดงขั้นตอนการศึกษาแบบ Case-control study



1. Select a sample from a population of people with the disease (cases)
2. Select a sample from a population at risk that is free of the disease (controls)
3. Measure predictor variables

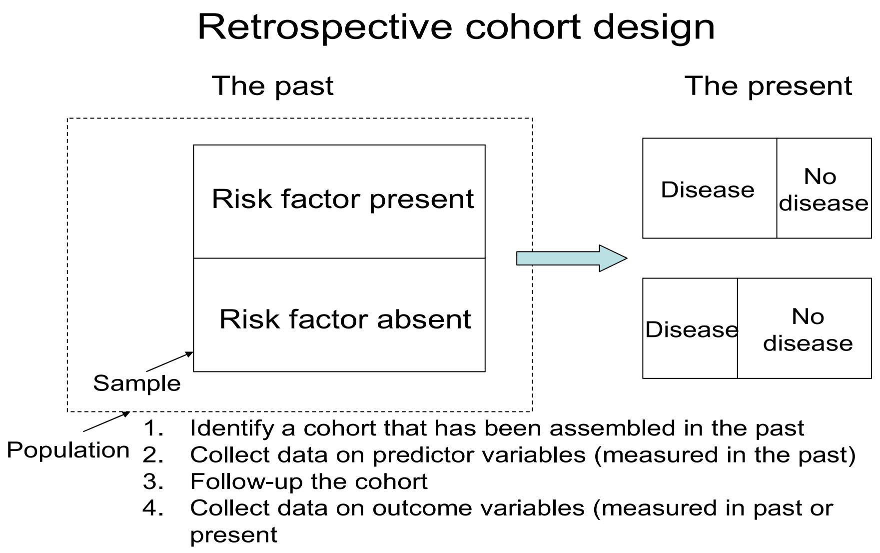
3. ***Cohort study*** การศึกษาแบบไปข้างหน้า (Prospective cohort) เป็นรูปแบบทั่วไปของการวิจัยประเภทนี้ เริ่มจากการสังเกตกลุ่มคนที่ยังไม่เป็นโรค วัดดูว่าใครบ้างที่ไม่มีปัจจัยเสี่ยงต่อการเกิดโรค และใครบ้างที่มีมีปัจจัยเสี่ยงต่อการเกิดโรค จากนั้นผู้ศึกษาจะเฝ้าติดตามประชากรกลุ่มนี้ไปในอนาคต แล้ววัดอุบัติการณ์ (Incidence) ของโรคในกลุ่มคนทั้งสองเปรียบเทียบกัน หากกลุ่มที่มีปัจจัยเสี่ยงมีอุบัติการณ์ของโรคสูงกว่ากลุ่มที่ไม่มีปัจจัยเสี่ยง หมายถึงปัจจัยนั้นมีความสัมพันธ์กับการเกิดโรค และอาจเป็นสาเหตุของโรค การศึกษาแบบ Cohort studies มีข้อดีในการแสดงหลักฐานที่หนักแน่นถึงความสัมพันธ์เชิงสาเหตุระหว่างปัจจัยเสี่ยงกับการเกิดโรค เนื่องจากเริ่มติดตามกลุ่มตัวอย่างตั้งแต่ยังไม่เป็นโรค ผลพลอยได้อีกอย่างหนึ่งคือทำให้ทราบอุบัติการณ์ (Incidence) ของโรคในประชากร และสามารถศึกษาการเกิดโรคหลาย ๆ โรคในเวลาเดียวกันได้ อย่างไรก็ตาม การศึกษาแบบนี้มีค่าใช้จ่ายสูงมาก ใช้เวลาในการศึกษานาน และต้องใช้กลุ่มตัวอย่างจำนวนมาก

แผนภูมิแสดงขั้นตอนการศึกษาแบบ Cohort study



การศึกษาแบบ Retrospective cohort เป็นรูปแบบเฉพาะของ Cohort study ซึ่งมีลักษณะคล้ายกับการศึกษาแบบ Prospective cohort เพียงแต่การรวบรวมกลุ่มคนที่ยังไม่เป็นโรค ตลอดจนการวัดว่าใครบ้างมีหรือไม่มีปัจจัยเสี่ยงต่อการเกิดโรคที่สนใจนั้นเป็นข้อมูลที่มีผู้บันทึกไว้แล้วในอดีต ผู้วิจัยมีหน้าที่วัดหรือหาข้อมูลอุบัติการณ์การเกิดโรคในกลุ่มตัวอย่างเดิมทั้งสองกลุ่มในปัจจุบันแล้วนำมาเปรียบเทียบกัน ข้อดีของการศึกษาแบบนี้ที่สำคัญที่สุดคือผู้วิจัยไม่ต้องเสียเวลาในการติดตามกลุ่มตัวอย่างเอง ซึ่งหากเป็นการศึกษาแบบ Prospective cohort อาจต้องใช้เวลานานเป็นสิบ ๆ ปี อย่างไรก็ตามการศึกษาแบบนี้ก็มีข้อจำกัดที่สำคัญคือ ข้อมูลที่ต้องการศึกษาต้องมีผู้เก็บรวบรวมตลอดจนบันทึกไว้ให้แล้ว และผู้วิจัยต้องยอมรับในคุณภาพ หรือความไม่ครบถ้วนของข้อมูล ซึ่งไม่สามารถปรับปรุงเปลี่ยนแปลงได้

แผนภูมิแสดงขั้นตอนการศึกษาแบบ Retrospective cohort studies



การศึกษาทางระบาดวิทยาลักษณะอื่น ๆ

1. **Correlational study (Ecological study)** คือการศึกษาทางระบาดวิทยาที่หน่วยที่ใช้ศึกษาเป็นกลุ่มประชากร ไม่ใช่อาสาสมัครแต่ละคนหรือผู้ป่วยแต่ละราย ตัวอย่างเช่น เพศชายมีอัตราการสูบบุหรี่สูงกว่าเพศหญิง ในขณะเดียวกันเพศชายก็เป็นมะเร็งปอดมากกว่าเพศหญิง ข้อมูลดังกล่าวช่วยสนับสนุนว่าการสูบบุหรี่เป็นสาเหตุของมะเร็งปอด หรือการศึกษาความสัมพันธ์ระหว่างคุณภาพอากาศของเมืองต่าง ๆ กับอัตราการป่วยเป็นโรคภูมิแพ้ทางเดินหายใจเป็นต้น อย่างไรก็ตามข้อมูลจากการศึกษาแบบนี้น่าเชื่อถือน้อยกว่า การศึกษา 2 แบบหลักข้างต้น และไม่สามารถใช้ในการสรุปความสัมพันธ์ได้ เป็นเพียงข้อมูลประกอบการนำไปสู่การศึกษาอื่น ๆเพิ่มเติม

2. **การเฝ้าระวังทางระบาดวิทยา (Surveillance)** การเฝ้าระวังทางระบาดวิทยาหมายถึงระบบติดตาม รวบรวม วิเคราะห์ และแปลผลข้อมูลสุขภาพอย่างต่อเนื่อง โดยมีวัตถุประสงค์เพื่อให้ทราบการระบาดของโรคแต่เนิ่น ๆ ทราบแนวโน้มของลักษณะการเกิด การกระจายของโรค ตลอดจนปัจจัยอื่น ๆ ที่เกี่ยวข้อง เพื่อนำไปใช้ในการวางแผน ดำเนินการ และประเมินผลมาตรการทางด้านสาธารณสุข นอกจากนั้นการเฝ้าระวังทางระบาดวิทยา ยังหมายรวมถึง การกระจายข้อมูลที่ได้จากระบบเฝ้าระวังให้กับผู้ที่จำเป็นต้องทราบเพื่อนำไปใช้ประโยชน์ด้วย การเฝ้าระวังทางระบาดวิทยาส่วนใหญ่จะกระทำโดยหน่วยงานภาครัฐที่มีหน้าที่รับผิดชอบในการป้องกันและควบคุมโรค

   วิธีการดำเนินงานเฝ้าระวังทางระบาดวิทยา อาจแบ่งได้เป็น 4 ลักษณะ คือ

   1. <u>Passive Surveillance</u> เป็นระบบเฝ้าระวังตามปกติที่ให้ผู้ให้บริการตามสถานบริการสุขภาพและเครือข่ายทางด้านสาธารณสุขในพื้นที่ รายงานโรคและภัยสุขภาพที่อยู่ในข่ายที่ต้องรายงาน ผ่านทางแบบฟอร์มรวบรวมข้อมูลที่มีอยู่เข้ามายังหน่วยงานที่รับผิดชอบ โดยมีระยะเวลาการส่งรายงานเป็นรายสัปดาห์ หรือรายเดือน ตามกำหนด

   2. <u>Active Surveillance</u> เป็นระบบเฝ้าระวังโรคที่เพิ่มขึ้นจากระบบปกติเพื่อให้ได้ข้อมูลมากขึ้น เป็นการค้นหาการเกิดโรคเชิงรุก เช่นกรณีโรคที่พบไม่บ่อย ที่เจ้าหน้าที่ออกติดตาม และเก็บข้อมูลในพื้นที่อย่างใกล้ชิด เพื่อค้นหาผู้ป่วยรายใหม่ ๆ เน้นในพื้นที่หรือกลุ่มประชากรที่เสี่ยงต่อการเกิดโรค การเฝ้าระวังแบบนี้จะได้ข้อมูลครบถ้วน และรวดเร็ว มากยิ่งขึ้น

   3. <u>Special Surveillance</u> เป็นระบบเฝ้าระวังที่สร้างขึ้นเพื่อให้ได้ข้อมูลเป็นกรณีพิเศษ ในการค้นหาการเกิดโรคใหม่ ๆ หรือโรคที่มีอยู่แล้ว แต่มีแนวโน้มการเกิดโรคมากผิดปกติไปจากเดิม หรือกรณีการเฝ้าระวังในภาวะที่เกิดภัยธรรมชาติ เช่นอุทกภัย ธรณีพิบัติภัย เป็นต้น

   4. <u>Sentinel Surveillance</u> เป็นระบบเฝ้าระวังในโรคหรือภาวะที่ต้องการข้อมูลที่มีความน่าเชื่อถือมากขึ้น และมีความรวดเร็ว โดยให้กลุ่มตัวอย่างกระจายตามลักษณะทางภูมิศาสตร์ ประชากร และอื่นๆ ที่เกี่ยวข้องให้มากที่สุด เช่นการเฝ้าระวังการติดเชื้อเอช ไอ วี ในผู้ติดยาเสพติด หรือผู้ให้บริการทางเพศ เป็นต้น

3. **การสอบสวนทางระบาดวิทยา (Epidemiologic Investigation)** หมายถึงมาตรการสาธารณสุขเชิงรุกเมื่อมีการระบาดของโรคเกิดขึ้น ส่วนใหญ่ใช้กับการระบาดอย่างเฉียบพลันของโรคติดเชื้อที่มีระยะฟักตัวสั้น และเกิดขึ้นในพื้นที่จำกัดเช่นโรคอุจจาระร่วงอย่างแรงเป็นต้น โดยทีมเจ้าหน้าที่จะออกสืบสวนสอบสวน หาต้นตอที่เป็นสาเหตุของการระบาดของโรค โดยอาศัยเทคนิคการศึกษาทางระบาดวิทยา และเมื่อทราบสาเหตุของโรคแล้วก็จะใช้มาตรการควบคุมโรคที่จำเป็นในการตัดวงจรการแพร่กระจายของเชื้อ

โรค จำกัดวงของปัญหาไม่ให้ขยายตัวออกไปจนส่งผลกระทบกับประชากรในพื้นที่อื่น รูปแบบการศึกษาทางระบาดวิทยาที่นิยมใช้ในการสอบสวนโรคได้แก่ Retrospective cohort และ Case-control

****

เอกสารอ้างอิง

1. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
2. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.

****

แบบฝึกหัด

1. การศึกษาเพื่อทดสอบสมมติฐานว่าการใช้ยาเม็ดคุมกำเนิดเป็นสาเหตุของการติดเชื้อในทางเดินปัสสาวะ ผู้วิจัยถามประวัติการใช้ยาเม็ดคุมกำเนิดในผู้ป่วย และผู้ที่ไม่ป่วย การศึกษาดังกล่าว จัดเป็นการศึกษารูปแบบใด

A. Cross-sectional

B. Case-control

C. Prospective cohort

D. Retrospective cohort

E. Randomized control trial

2. รูปแบบการศึกษาที่เหมาะสมที่สุดในการหาสาเหตุของโรคคือ

A. Cross-sectional study

B. Case-control study

C. Prospective cohort study

D. Randomized control trial

E. Systematic review

3. ข้อจำกัดที่สำคัญที่สุดของการศึกษาแบบ Retrospective cohort คือ

A ค่าใช้จ่ายสูง

B. จริยธรรมวิจัย

C. ใช้ระยะเวลานาน

D. คุณภาพของข้อมูลไม่ดี

E. ไม่บอกอุบัติการณ์ของโรค

4. นักวิชาการประจำสำนักงานสาธารณสุขอำเภอ เก็บรวบรวมข้อมูลทางระบาดวิทยาของการป่วยเป็นโรคตาแดงในพื้นที่รับผิดชอบ เป็นระยะเวลา 3 เดือน ตั้งแต่เริ่มเกิดปัญหาน้ำท่วม การกระทำดังกล่าวเข้าได้กับการเฝ้าระวังโรคแบบใด

A. Active

B. Passive

C. Special

D. Sentinel

E. Investigation

5. การศึกษาว่าการขริบหนังหุ้มปลายอวัยวะเพศชายช่วยลดโอกาสการติดเชื้อ HIV หรือไม่ ควรเลือกรูปแบบการศึกษาใด

A. Cohort

B. Ecological

C. Case-control

D. Experiment

E. Cross-sectional

1.11.2.3 Sampling and sample size

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

ประชากร และกลุ่มตัวอย่าง (Population and Sample)

ในการศึกษาเชิงปริมาณ ประชากร (Population) หมายถึง กลุ่มคนทั้งหมดที่ต้องการศึกษาและต้องการนำผลการวิจัยอ้างอิงไปถึงหรือนำไปใช้ ในทางการแพทย์มักเป็นบุคคลที่มีลักษณะเฉพาะ หรือผู้ป่วยด้วยโรคหรือภาวะสุขภาพที่สนใจ ประชากรมี 2 ลักษณะ คือประชากรนับได้ (Finite population) มีจำนวนจำกัด เช่นนักศึกษาแพทย์ คณะแพทยศาสตร์มหาวิทยาลัยเชียงใหม่ ปีการศึกษา 2556 และประชากรอนันต์ (Infinite population) ที่ผู้ศึกษาไม่ทราบจำนวนที่แน่นอน เช่นผู้ป่วยโรคเบาหวานในประเทศไทย ซึ่งการศึกษาทางคลินิกส่วนใหญ่ประชากรจะเป็นลักษณะหลัง การกำหนดขอบเขตของประชากรที่ชัดเจนมีความสำคัญอย่างมากต่อการดำเนินการออกแบบการศึกษาในขั้นตอนต่อ ๆ ไป ส่วนกลุ่มตัวอย่าง (Sample) หมายถึง ส่วนหนึ่งของประชากร ที่ถูกเลือกมาทำการศึกษาจริง

การวิจัยทางการแพทย์ไม่สามารถทำการศึกษาในประชากรทั้งหมดหรือในผู้ที่ป่วยด้วยโรคนั้น ๆ ทุกรายได้ด้วยข้อจำกัดหลายประการ อย่างไรก็ตามหากผู้ศึกษาใช้รูปแบบในการเลือกกลุ่มตัวอย่างจากประชากรอย่างเหมาะสมและด้วยขนาดตัวอย่างที่เพียงพอ ร่วมกับการดำเนินการวิจัยที่มีคุณภาพ ไม่มีอคติหรือมีน้อยที่สุด เราก็สามารถใช้หลักสถิติ ในการอนุมานผลการศึกษาที่ได้จากกลุ่มตัวอย่างนี้ ได้ใกล้เคียงกับความจริงในประชากร ความเข้าใจเกี่ยวกับวิธีการคัดเลือกกลุ่มตัวอย่างจากประชากรและการประมาณขนาดของกลุ่มตัวอย่าง จึงเป็นมีความสำคัญมาก

การประมาณขนาดกลุ่มตัวอย่าง (Sample size estimation)

1. <u>ความสำคัญของขนาดกลุ่มตัวอย่าง</u>

ในการศึกษาเชิงวิเคราะห์เพื่อหาความสัมพันธ์ระหว่างตัวแปรต่าง ๆ หรือเพื่อทดสอบสมมติฐาน ขนาดของกลุ่มตัวอย่างมีความสำคัญอย่างยิ่งต่อความแม่นยำของการสรุปผลการศึกษา การวิจัยที่มีขนาดตัวอย่างน้อยเกินไป ในทางสถิติแล้วจะมีอำนาจ (Statistical power) ไม่มากพอที่ตรวจพบความสัมพันธ์ดังกล่าวหากมีอยู่จริง ทำให้มีความน่าจะเป็นที่จะสรุปผลผิดพลาดชนิดผลลบลวง (Type II error, β error) สูง และผลการศึกษามีความไม่แน่นอนสูง (Confidence interval กว้าง) ขณะที่หากขนาดตัวอย่างมากเกินไปจะทำให้สิ้นเปลืองทรัพยากรที่ใช้ในการศึกษาโดยไม่จำเป็น

ขนาดกลุ่มตัวอย่างที่คำนวณได้ยังช่วยบอกถึงความเป็นไปได้ของการทำการศึกษา ตั้งแต่ระหว่างขั้นเตรียมการ เช่นหากค่าใช้จ่ายที่คาดว่าจะต้องใช้ในการทำวิจัยตามขนาดตัวอย่างที่คำนวณได้สูงกว่าทุนวิจัยที่มีมาก การวิจัยนั้นก็ยังไม่ควรเริ่มทำเพราะจะไม่สามารถศึกษาอาสาสมัครได้ครบตามจำนวนที่กำหนด ซึ่งจะนำไปสู่การไม่สามารถสรุปผลการวิจัยได้อย่างเหมาะสม หรือกรณีการศึกษาทางคลินิกในผู้ป่วยที่ป่วยด้วยโรคที่พบไม่บ่อย เราอาจคาดการณ์ได้ตั้งแต่ต้นถึงความเป็นไปได้ในหาอาสาสมัครซึ่งป่วยด้วยโรคนั้น ๆ ให้ครบตามจำนวนกลุ่มตัวอย่างที่คำนวณไว้ว่ามีมากน้อยเพียงใด นอกจากนั้นการเตรียมการต่าง ๆ สำหรับการวิจัย เช่นจำนวนบุคคลากร สถานที่ ระยะเวลา และค่าใช้จ่าย ยังขึ้นกับขนาดของกลุ่มตัวอย่างอีกด้วย

ข้อพึงระวังอีกประการหนึ่งคือ ต้องแยกนัยสำคัญทางสถิติ (Statistical significance) จากนัยสำคัญทางคลินิก (Clinical significance) เนื่องจากความแตกต่างอย่างมีนัยสำคัญทางสถิติซึ่งเกิดขึ้นได้เสมอหากตัวอย่างมีขนาดใหญ่พอ อาจไม่มีความสำคัญต่อการนำไปใช้ป้องกันแก้ไขปัญหาสุขภาพจริง เพราะความแตกต่างที่พบดังกล่าวมีขนาดน้อยเกินไป และในทางกลับกัน ความแตกต่างหรือความสัมพันธ์ที่ไม่มีนัยสำคัญทางสถิติบางอย่าง อาจมีความสำคัญทางระบาดวิทยาหรือทางคลินิกก็ได้ แพทย์จึงต้องพิจารณาขนาดของความแตกต่างหรือความสัมพันธ์ที่พบ เทียบเคียงกับลักษณะทางพยาธิวิทยา การดำเนินโรค ตลอดจนผลทางด้านการสาธารณสุขเสมอ การดูนัยสำคัญเพียงสถิติเพียงอย่างเดียว อาจทำให้มีการนำผลการศึกษาไปใช้อย่างไม่เหมาะสม เนื่องจากขนาดตัวอย่างที่เพิ่มขึ้นจะทำให้โอกาสพบการมีนัยสำคัญทางสถิติเพิ่มขึ้น

2. <u>ปัจจัยที่มีผลต่อการกำหนดขนาดกลุ่มตัวอย่าง (Factors affecting sample size)</u>

2.1 *Effect size* คือ ขนาดของความสัมพันธ์ ประสิทธิผลของ intervention หรือความแตกต่างที่น้อยที่สุดที่ผู้วิจัยต้องการตรวจพบ หากต้องการตรวจพบความแตกต่างขนาดเล็ก ๆ ก็ต้องใช้กลุ่มตัวอย่างจำนวนมาก (เปรียบได้กับมองหาสิ่งของขนาดเล็ก) แต่ถ้าความแตกต่างที่ต้องการตรวจพบมีขนาดใหญ่ (เปรียบได้กับการมองหาสิ่งของขนาดใหญ่) ก็จะตรวจพบได้ง่ายกว่า โดยไม่ต้องใช้กลุ่มตัวอย่างจำนวนมากนัก การกำหนด Effect size ต้องพิจารณาให้เหมาะสมกับลักษณะทางคลินิกของ Outcome หรือปัญหาสุขภาพนั้น ๆ ว่าความแตกต่างขนาดเท่าใดถือว่ามีความสำคัญทางคลินิก โดยอาจใช้ข้อมูลจากการวิจัยในอดีต การปรึกษาผู้เชี่ยวชาญ หรืออาจต้องลองทำการศึกษานำร่อง (Pilot study) เพื่อให้ได้ข้อมูลนี้มา

<u>*ตัวอย่าง*</u>

ผู้วิจัยวางแผนทำ RCT เพื่อหาประสิทธิผลของวัคซีนป้องกันโรคไข้หวัดใหญ่ในผู้สูงอายุ ถ้าต้องการทราบว่าวัคซีนจะสามารถลด incidence ลงจาก 3% เหลือ 1% ได้หรือไม่ จะต้องการขนาดตัวอย่างน้อยกว่า เมื่อเทียบกับการต้องการทราบว่าวัคซีนจะสามารถลด Incidence จาก 3% เหลือ 2% ได้หรือไม่

2.2 α *error* หรือ Type I error คือระดับของความน่าจะเป็นในการยอมรับความผิดพลาดของการสรุปผลการศึกษาแบบผลบวกลวง (False positive) เนื่องจากความบังเอิญ หมายถึงในความเป็นจริงไม่มีความแตกต่างหรือไม่มีความสัมพันธ์ระหว่างตัวแปร แต่ผู้วิจัยสรุปว่ามี ถ้ากำหนดความน่าจะเป็นนี้ไว้สูง ก็จะมีโอกาสตรวจพบนัยสำคัญทางสถิติได้ง่ายขึ้นทำให้ต้องใช้ขนาดของกลุ่มตัวอย่างน้อยลง แต่ขณะเดียวกันก็ต้องยอมรับโอกาสเกิดผลบวกลวงที่สูงขึ้นด้วย โดยทั่วไปถูกกำหนดไว้ที่ร้อยละ 5 หรือผู้วิจัยมีโอกาสสรุปผลการศึกษาผิดพลาดแบบนี้ 1 ใน 20

2.3 β *error* หรือ Type II error คือระดับของความน่าจะเป็นในการยอมรับความผิดพลาดของการสรุปผลการศึกษาแบบผลลบลวง (False negative) หมายถึงในความจริงมีความแตกต่างหรือมีความสัมพันธ์ระหว่างตัวแปรแต่ผู้วิจัยตรวจไม่พบ ปกติจะกำหนดในรูปของ Statistical power หมายถึงความน่าจะเป็นที่จะตรวจพบความแตกต่างหรือความสัมพันธ์ดังกล่าวหากมีอยู่จริง (True positive) ซึ่งมีค่าเท่ากับ 1 - β โดยทั่วไปกำหนด β error ไว้ที่ร้อยละ 20 หรือ Statistical Power = 80% การกำหนด Statistical Power ที่สูงขึ้น จะต้องการขนาดกลุ่มตัวอย่างเพิ่มขึ้น

จะเห็นว่าการกำหนดค่า α error ที่ 0.05 และ β error ที่ 0.20 ซึ่งยอมรับว่าเป็นมาตรฐานของงานวิจัยทั่วไป แสดงถึงการยอมรับความผิดพลาดในสรุปผลการวิจัยแบบผลลบลวงมากกว่าการเกิดผลบวกลวง

2.4 *Characteristics of the data* การประมาณขนาดของกลุ่มตัวอย่างที่เหมาะสม ยังขึ้นกับลักษณะของประชากรด้วย ถ้าประชากรส่วนใหญ่มีลักษณะคล้ายคลึงกัน (Homogeneity) ในตัวแปรที่ต้องการศึกษา กรณีนี้จะต้องการขนาดตัวอย่างน้อยกว่า เมื่อเทียบกันกับกรณีที่ประชากรมีลักษณะแตกต่างกัน (Heterogeneity) ในตัวแปรที่ต้องการศึกษา

2.5 *One-sided or two sided* ผู้วิจัยต้องกำหนดว่าจะหาความสัมพันธ์ไปในทิศทางเดียว (One-sided) เช่นในการทดลองประสิทธิผลของยาใหม่เทียบกับยามาตรฐานเดิม ผู้วิจัยสนใจเฉพาะว่ายาใหม่ให้ผลดีกว่ายาเดิมหรือไม่ ไม่สนใจกรณียาใหม่ให้ผลการรักษาแย่กว่ายาเดิม กรณีนี้จะต้องการขนาดตัวอย่างน้อยกว่าหากผู้วิจัยเผื่อความสนใจไว้ทั้งสองด้าน (Two-sided) โดยทั่วไปหากไม่มีเหตุผลที่ชัดเจน ยอมรับได้ ควรกำหนดขนาดตัวอย่างแบบ Two-sided เพราะในความเป็นจริงผลอาจเกิดได้ทั้งสองทาง คืออาจดีขึ้นหรืออาจแย่ลง

รูปแบบในการเลือกกลุ่มตัวอย่าง (Sampling strategies)

เทคนิคการเลือกกลุ่มตัวอย่างจากประชากรมีได้หลายลักษณะ การจะเลือกใช้วิธีใดขึ้นกับสถานการณ์ของการศึกษา โดยทั่วไปอาจแบ่งได้เป็น 2 กลุ่ม คือ การเลือกกลุ่มตัวอย่างโดยใช้หลักความน่าจะเป็น (Probability sampling) และ การเลือกกลุ่มตัวอย่างโดยไม่ใช้หลักความน่าจะเป็น (Non-probability sampling)

1. <u>Probability sampling</u>

หรือเรียกอีกชื่อหนึ่งว่า Random sampling เป็นเทคนิคการเลือกกลุ่มตัวอย่างโดยใช้หลักของความน่าจะเป็น (Probability) และหลักการสุ่ม (Randomness) โดยกลุ่มตัวอย่างจะถูกสุ่มเลือกจากประชากรอย่างทราบโอกาสความน่าจะเป็นที่จะถูกเลือก ทำให้ปราศจากอคติในการเลือกกลุ่มตัวอย่าง และมีน้ำหนักน่าเชื่อถือในทางสถิติ สามารถใช้หลักการสถิติเชิงอนุมาน (Inferential statistics) ในการอ้างอิงค่าสถิติที่ได้จากกลุ่มตัวอย่างไปยังค่าจริงในประชากรได้ดี ประกอบด้วยเทคนิคย่อย 5 แบบ คือ

1.1 Simple random sampling หรือการสุ่มตัวอย่างแบบง่าย เป็นวิธีพื้นฐานที่สุดของ Probability sampling ประชากรทุกคนจะมีโอกาสถูกสุ่มเลือกมาเป็นกลุ่มตัวอย่างเท่า ๆ กัน อาจทำโดยวิธีจับสลาก ใช้ตารางเลขสุ่ม หรือโปรแกรมคอมพิวเตอร์ช่วยในการสุ่มก็ได้ ข้อดีคือปราศจากอคติในการเลือกกลุ่มตัวอย่าง และไม่ต้องทราบลักษณะปลีกย่อยอื่นของประชากร อย่างไรก็ตามจะใช้วิธีนี้ได้ ต้องทราบประชากรทั้งหมด และประชากรควรมีจำนวนไม่มากนัก

1.2 Systematic random sampling หรือการสุ่มตัวอย่าง ๆ มีระบบ มีลักษณะคล้ายกับ Simple random sampling เพียงแต่ทำให้สะดวกและง่ายขึ้น เริ่มจากคำนวณหา Sampling interval โดยนำจำนวนประชากรทั้งหมดหารด้วยขนาดของตัวอย่างที่คำนวณได้ ตัวอย่างเช่น

จำนวนประชากร	= 1,000
ขนาดตัวอย่าง	= 100
Sampling interval	= 1,000 / 100 = 10

จากนั้นให้ทำการสุ่มหาเลขเริ่มต้น (Random start number) โดยการสุ่มจับสลากเลข 1 ถึง 10 มา 1 เลข สมมติว่าได้เลข 6 ก็ให้ไปดูรายชื่อประชากรในฐานข้อมูลที่เรียงลำดับไว้อยู่แล้วอย่างมีระบบ นำรายชื่อลำดับที่ 6 มาเป็นตัวอย่างรายแรก จากนั้นบวกด้วย Sampling interval คือ 10 จะได้กลุ่มตัวอย่างรายที่ 2 คือผู้ที่มีรายชื่อลำดับที่ 16 และเมื่อทำอย่างนี้ซ้ำไปเรื่อย ในที่สุดจะได้กลุ่มตัวอย่างครบ 100 ราย โดยไม่ต้องทำการสุ่มหรือจับสลากซ้ำทั้ง 100 ครั้ง เพียงแค่สุ่มเลขเริ่มต้นเพียงครั้งเดียวเท่านั้น ในการวิจัยทางคลินิกมักใช้บ่อยในลักษณะของการเลือกผู้ป่วยเข้าสู่การศึกษาทุก ๆ n คน เช่น ทุก 3 คน หรือทุก 4 คน ตามลำดับของการมารับการรักษา

1.3 *Stratified random sampling* หรือการสุ่มตัวอย่างตามระดับชั้น ขั้นแรกแบ่งประชากรออกเป็นกลุ่มย่อย (Strata) ตามตัวแปรต่าง ๆ ตามต้องการ เช่น เพศ อายุ เชื้อชาติ ฐานะทางเศรษฐกิจ หรือระยะของการดำเนินโรค จากนั้นทำการสุ่มเลือกตัวอย่างแบบง่าย จากประชากรแต่ละกลุ่มย่อยนั้นตามสัดส่วนที่เหมาะสม แล้วค่อยนำผู้ที่ถูกสุ่มได้มารวมกันเป็นกลุ่มตัวอย่างของการศึกษา ใช้ในกรณีที่กลุ่มย่อยแต่ละกลุ่มมีลักษณะที่แตกต่างกันมาก วิธีนี้จะช่วยลดความคลาดเคลื่อนในการเลือกกลุ่มตัวอย่าง ผู้วิจัยอาจเพิ่มโอกาสที่จะมีตัวแทนจากทุกกลุ่มย่อย กรณีที่กลุ่มนั้น ๆ มีจำนวนน้อยในประชากรได้ โดยการเพิ่มสัดส่วนของประชากรกลุ่มนั้นในกลุ่มตัวอย่าง (Over sampling)

1.4 *Clustered random sampling* หรือการสุ่มตัวอย่างแบบกลุ่ม ใช้ในกรณีประชากรมีการอยู่กันเป็นกลุ่มก้อนอยู่แล้วตามธรรมชาติ ส่วนใหญ่เป็นกลุ่มทางภูมิศาสตร์หรือสถานที่ เช่น โรงเรียน หมู่บ้าน ตำบล เรียกว่าอยู่กันเป็น Cluster ทำได้โดยทำการสุ่มตัวอย่างแบบง่ายให้ได้ Cluster มา แล้วทำการศึกษาเฉพาะใน Cluster ที่สุ่มได้ โดยการศึกษาประชากรทุกคนใน Cluster นั้น ๆ ประโยชน์ที่สำคัญ คือประหยัดค่าใช้จ่ายในการเดินทางไปเก็บข้อมูล และลดค่าใช้จ่ายด้านการบริหารจัดการ ข้อควรระวังคือ แต่ละ Cluster ควรมีลักษณะการกระจายของข้อมูลที่สนใจคล้ายคลึงกัน และทุก Cluster ควรสามารถเป็นตัวแทนที่ดีของประชากรได้ กรณีที่ Cluster ต่าง ๆ มีลักษณะของตัวแปรที่ต้องการศึกษาแตกต่างกันมาก การศึกษาเพียงบาง Cluster อาจทำให้ผลการศึกษาที่ได้คาดเคลื่อนจากความเป็นจริง

1.5 *Multi-stage sampling* หรือการสุ่มตัวอย่างแบบหลายขั้นตอน มักใช้ในการศึกษาทางระบาดวิทยาหรือการสำรวจทางสาธารณสุขขนาดใหญ่ที่ไม่ทราบขอบเขตที่แน่นอนของประชากร ทำโดยการสุ่มตัวอย่างโดยใช้เทคนิคหลายอย่างร่วมกัน และทำเป็นหลายขั้นตอน จนในที่สุดได้กลุ่มตัวอย่างที่ใช้ศึกษาจริง ส่วนใหญ่เริ่มจากการสุ่มเลือก Clusters ที่จะทำการศึกษา ซึ่งอาจจะทำหลายครั้งจากกลุ่มใหญ่ไปหากลุ่มย่อย ในขั้นตอนหลัง ๆ อาจใช้เทคนิค Stratified random sampling หรือ Simple random sampling ร่วมด้วย

2. <u>Non-probability sampling</u>

หมายถึงการได้มาซึ่งกลุ่มตัวอย่างโดยไม่ได้ใช้วิธีการสุ่มเลือก ทำให้ไม่สามารถทราบโอกาสทางสถิติที่ประชากรแต่ละคนจะถูกเลือกมาเป็นกลุ่มตัวอย่าง เทคนิคการเลือกกลุ่มตัวอย่างแบบนี้ใช้หลักสถิติเชิงอนุมานได้น้อย อย่างไรก็ตามแม้ว่าการเลือกตัวอย่างแบบใช้หลักความน่าจะเป็น จะมีข้อดีเหนือกว่าการเลือกกลุ่มตัวอย่างแบบไม่ใช้หลักความน่าจะเป็น แต่การทำการวิจัยในชีวิตจริงโดยเฉพาะอย่างยิ่งการศึกษาทางคลินิก มีข้อจำกัดหลายประการที่ทำให้ไม่สามารถใช้เทคนิคการเลือกกลุ่มตัวอย่างแบบแรกได้ เนื่องจากการศึกษาทางคลินิกส่วนใหญ่เป็นการศึกษาในผู้ป่วยเฉพาะโรคซึ่งไม่มีขอบเขตของประชากรที่แน่นอน หรือผู้ป่วยที่มีอาจมีจำนวนไม่มาก และไม่ได้เจ็บป่วยในเวลาใกล้เคียงกัน พร้อม ๆ กันให้ศึกษา นอกจากนั้นยังมีข้อจำกัดด้านเวลา ค่าใช้จ่าย และจริยธรรม ที่ทำให้การวิจัยทางคลินิกส่วนใหญ่จำเป็นต้อง

ใช้การเลือกกลุ่มตัวอย่างแบบไม่ใช้หลักความน่าจะเป็น ซึ่งอาจแบ่งออกตามลักษณะของการได้มาซึ่งกลุ่มตัวอย่างดังนี้

2.1 *Consecutive sampling* เป็นเทคนิคการเลือกกลุ่มตัวอย่างแบบไม่ใช้หลักความน่าจะเป็นที่ใช้บ่อยที่สุดในการวิจัยทางคลินิกแบบ Clinical trials โดยการเชิญผู้ป่วยทุกรายที่มีคุณสมบัติเข้าข่ายตามที่กำหนดไว้ (Eligible criteria) เข้าเป็นอาสาสมัครของโครงการ ตามลำดับการมาติดต่อขอเป็นอาสาสมัครหรือการได้รับการวินิจฉัย และยุติการรับเข้าเมื่อได้จำนวนอาสาสมัครครบตามขนาดตัวอย่างหรือหมดกำหนดเวลาในการรับ ถือเป็นเทคนิคที่ดีที่สุดในบรรดาการเลือกตัวอย่างแบบไม่ใช้หลักความน่าจะเป็น เนื่องจากหากระยะเวลาในการรับอาสาสมัครเข้าสู่โครงการวิจัยนานพอสมควรจะได้กลุ่มตัวอย่างที่เป็นตัวแทนที่ค่อนข้างดีของประชากร

2.2 *Convenience sampling* คือการศึกษาในกลุ่มประชากรที่ผู้วิจัยสะดวกในการเข้าถึง หรือในกลุ่มตัวอย่างที่มีอยู่แล้ว เช่นการสอบถามความพึงพอใจของการรับบริการจากผู้ป่วยและญาติ ในช่วงเวลาที่กำหนด 5 วันทำการเป็นเวลา 1 สัปดาห์ หรืออาจารย์แพทย์ต้องการอาสาสมัครที่มีสุขภาพดีเพื่อการวิจัยทางคลินิกบางอย่างแล้วมีนักศึกษาแพทย์จำนวนหนึ่งอาสาสมัครเข้ามาเป็นต้น การเลือกตัวอย่างแบบนี้มีอคติในการเลือกกลุ่มตัวอย่างได้มาก แต่มีข้อดีคือ ทำง่าย ประหยัดค่าใช้จ่าย และเวลา มักใช้ในการศึกษาระยะแรก (Preliminary research) เพื่อให้ได้ประมาณการคร่าว ๆ ของค่าตัวแปรที่สนใจ ก่อนที่ทำการศึกษาอย่างละเอียดลึกซึ้งต่อไป Consecutive sampling และการเลือกตัวอย่างโดยบังเอิญ (Accidental sampling) อาจจัดเป็น Convenience sampling รูปแบบหนึ่ง

2.3 *Purposive sampling* หรืออีกชื่อหนึ่งคือ Judgmental sampling กรณีนี้ผู้วิจัยเจาะจงเลือกที่จะทำการศึกษาในกลุ่มตัวอย่างบางคนหรือบางกลุ่ม โดยใช้เหตุผลและวิจารณญาณของตนเองว่ามีความเหมาะสมในการนำมาเป็นกลุ่มตัวอย่างในการศึกษา เช่นเจาะจงเลือกศึกษาเกี่ยวกับพฤติกรรมทางเพศเฉพาะในโรงเรียนมัธยมที่ผู้บริหารให้ความร่วมมือ อีกกรณีหนึ่งคือเมื่อกลุ่มตัวอย่างมีจำนวนน้อย เช่นผู้ป่วยที่ป่วยด้วยโรคที่พบไม่บ่อย กรณีนี้ผู้วิจัยอาจต้องไปทำการศึกษาในโรงพยาบาล หรือสถาบันที่ดูแลผู้ป่วยเหล่านี้ อย่างไรก็ตามผลการศึกษาในสถานที่ ๆ เฉพาะเจาะจงนี้ อาจจะมีลักษณะที่แตกต่างจากผู้ป่วยโรคเดียวกัน แต่อยู่ในชุมชน

****

เอกสารอ้างอิง

1. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
2. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
3. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.

****

แบบฝึกหัด

1. เทคนิคการเลือกกลุ่มตัวอย่างแบบไม่ใช้หลักความน่าจะเป็น ที่ใช้บ่อยที่สุดในการวิจัยทางคลินิกคือ

A. Snowball sampling

B. Purposive sampling

C. Accidental sampling

D. Judgmental sampling

E. Consecutive sampling

2. ผู้วิจัยใช้ตารางเลขสุ่ม ๆเลือกกลุ่มตัวอย่างจากผู้มีกรุ๊ปเลือด A, B, AB, และ O กลุ่มละ เพื่อศึกษา ราย 25 เกี่ยวกับภูมิคุ้มกันโรค ถามว่าเป็นการสุ่มตัวอย่างแบบใด

A. Simple random

B. Systematic random

C. Stratified random

D. Clustered random

E. Multi-stage

3. ผู้วิจัยสุ่มเลือกพนักงานจำนวนเท่า ๆกัน จากแผนกต่าง ๆในบริษัท เพื่อเป็นตัวแทนในการตอบแบบสอบถาม จัดเป็นการสุ่มตัวอย่างแบบใด

A. Simple random

B. Systematic random

C. Stratified random

D. Clustered random

E. Multi-stage

1.11.2.4 Subject selection and exposure allocation

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

กลุ่มตัวอย่างที่ใช้ศึกษาในงานวิจัยทางคลินิก มีความสำคัญอย่างยิ่งต่อผลการศึกษา ตลอดจนการสรุปผลการศึกษาที่จะตามมา ความสำคัญอยู่ที่ ทำอย่างไรจึงจะได้มาซึ่งอาสาสมัครที่มีคุณลักษณะเหมาะสมและเป็นตัวแทนที่ดีของประชากรในเรื่องที่สนใจนั้น ๆขณะเดียวกันก็ไม่มีลักษณะที่เฉพาะเจาะจงเกินไปจนไม่สามารถนำผลไปใช้ในสถานการณ์จริงอื่น ๆ ได้ นอกจากนั้นอาสาสมัครยังต้องไม่มีข้อจำกัดอื่นที่อาจเป็นอุปสรรคต่อการร่วมในงานวิจัย และสำหรับการศึกษาแบบ Randomized controlled trials (RCT) จะมีการแบ่งกลุ่มตัวอย่างออกเป็นกลุ่มที่ได้รับสิ่งทดลองและกลุ่มเปรียบเทียบ โดยกระบวนการเชิงสุ่ม เพื่อให้เกิดอคติน้อยที่สุด

ประชากร และกลุ่มตัวอย่าง (Population and sample)

ประชากร (Population) คือกลุ่มบุคคลที่ระบุขอบเขตได้และมีลักษณะเพาะ ส่วนกลุ่มตัวอย่าง (Sample) เป็นส่วนหนึ่งของประชากรที่ผู้วิจัยเลือกมาจากประชากรเพื่อนำมาทำการศึกษา ในอุดมคติเพื่อให้การสรุปผลการศึกษาถูกต้องเสมอในทุกสถานการณ์เราต้องทำการศึกษาในประชากรทุกคนที่เราต้องการอนุมานผลการวิจัยไปถึง อย่างไรก็ตามในชีวิตจริงเป็นไม่สามารถกระทำได้ได้และไม่ควรทำ เนื่องจากประชากรดังกล่าวไม่สามารถแจงนับได้ นอกจากนั้นยังมีข้อจำกัดเรื่องงบประมาณ เวลาที่มีในการทำวิจัย และปัจจัยอื่น ๆ ในการศึกษาจริงจึงต้องมีการกำหนดประชากร และกลุ่มตัวอย่าง ซึ่งอาจแบ่งออกได้เป็น 3 ระดับ ดังนี้

1. <u>ประชากรเป้าหมาย (Target population, Reference population)</u>

   ในการทำวิจัยทางคลินิก ประชากรเป้าหมายมีความหมายใน 2 ลักษณะ ความหมายแรกหมายถึงประชากรในขอบเขตที่กว้างที่สุดที่ผู้วิจัยต้องการสรุป แปลผลการวิจัยไปถึง หรือต้องการนำผลการวิจัยที่ได้ไปใช้ สำหรับในอีกความหมายหนึ่งประชากรเป้าหมายเป็นการกำหนดขอบเขตของประชากรที่เหมาะสมสำหรับการนำมาใช้ในการศึกษา เป็นส่วนหนึ่งของตั้งเกณฑ์ในการเลือกกลุ่มตัวอย่าง (Eligibility criteria) ซึ่งจะกล่าวถึงต่อไป ประชากรเป้าหมายต้องมีลักษณะที่สอดคล้องกับรูปแบบ และวัตถุประสงค์ของวิจัย โดยทั่วไปใช้ลักษณะทางประชากรศาสตร์ (Demographic characteristics) และลักษณะทางคลินิก (Clinical characteristics) เป็นตัวกำหนด

   <u>*ตัวอย่าง*</u>

   ผู้หญิงวัยหลังหมดประจำเดือน

"ผู้หญิง"	*เป็นลักษณะทางประชากรศาสตร์*
"วัยหลังหมดประจำเดือน"	*เป็นลักษณะทางคลินิก*

   เด็กวัยรุ่นที่ป่วยเป็นโรคหืดหอบ

"เด็กวัยรุ่น	*เป็นลักษณะทางประชากรศาสตร์*
"ป่วยเป็นโรคหืดหอบ"	*เป็นลักษณะทางคลินิก*

2. <u>ประชากรที่เข้าถึงได้ (Accessible population, Sampled population)</u>

หมายถึงส่วนหนึ่งของประชากรเป้าหมายที่ผู้วิจัยสามารถเข้าถึงได้และสามารถนำมาใช้เป็นกรอบประชากรในการเลือกกลุ่มตัวอย่างในการศึกษาได้จริง กำหนดโดยขอบเขตทางด้านภูมิศาสตร์และระยะเวลา โดยกำหนดตามพื้นที่ที่โครงการวิจัยจะหาอาสาสมัครและระยะเวลาที่โครงการวิจัยกำหนดว่าจะรับอาสาสมัครเข้าโครงการ

ตัวอย่าง

ผู้หญิงวัยหลังหมดประจำเดือน ที่อาศัยอยู่ในอำเภอสันทรายจังหวัดเชียงใหม่ ในเดือนพฤศจิกายน 2555

"อาศัยอยู่ในอำเภอสันทรายจังหวัดเชียงใหม่" เป็นลักษณะทางภูมิศาสตร์

"ในเดือนพฤศจิกายน 2556" เป็นระยะเวลา

ผู้ป่วยเด็กวัยรุ่นที่ป่วยเป็นโรคหืดหอบ ที่ขึ้นทะเบียนเป็นผู้ป่วยรายใหม่ ณ โรงพยาบาลมหาราชนครเชียงใหม่ในปี 2556 และมีภูมิลำเนาอยู่ในภาคเหนือตอนบน

"มีภูมิลำเนาอยู่ในภาคเหนือตอนบน" เป็นลักษณะทางภูมิศาสตร์

"ขึ้นทะเบียนเป็นผู้ป่วยรายใหม่ ณ โรงพยาบาลมหาราชนครเชียงใหม่ในปี 2556" เป็นระยะเวลา

ประชากรที่เข้าถึงได้อาจแบ่งได้เป็น 2 ลักษณะ แบบแรกคือผู้ป่วยที่มารักษา ณ คลินิก หรือสถานพยาบาลที่ผู้วิจัยปฏิบัติหน้าที่อยู่ มักใช้ในการวิจัยทางคลินิก (Clinic based) เป็นกลุ่มประชากรที่ง่ายต่อการนำมาศึกษา เหมาะสำหรับการศึกษาเกี่ยวกับการรักษาโรคหรือการป้องกันภาวะแทรกซ้อนในผู้ที่ป่วยเป็นโรคที่สนใจศึกษาอยู่แล้ว ข้อควรระวังคือประชากรอาจมีความเฉพาะเจาะจงเกินไป โดยเฉพาะอย่างยิ่งหากเป็นการศึกษาที่ทำในศูนย์การแพทย์ขนาดใหญ่ เช่นโรงเรียนแพทย์ที่ผู้ป่วยมักถูกคัดกรองและส่งต่อมาหลายทอดทำให้อาจไม่เป็นตัวแทนที่ดีของผู้ป่วยโรคนั้น ๆ อีกแบบหนึ่งคือ การใช้ชุมชนเป็นฐานประชากร (Community based) ข้อดีคือเป็นตัวแทนที่ดีของประชากร เหมาะสำหรับการศึกษาเกี่ยวกับการส่งเสริมสุขภาพหรือการป้องกันโรค อย่างไรก็ตามไม่เหมาะกับการวิจัยเกี่ยวกับปัญหาสุขภาพที่พบไม่บ่อย เนื่องจากอาจไม่ได้ข้อมูลที่สนใจศึกษามากพอ นอกจากนั้นยังมีค่าใช้จ่ายในการได้มาซึ่งกลุ่มตัวอย่างสูงกว่า และยุ่งยากกว่าการใช้ฐานประชากรเป็นผู้ป่วยในคลินิก

ประชากรที่เข้าถึงได้ต้องมีจำนวนมากพอ คืออย่างน้อยต้องมากกว่าขนาดกลุ่มตัวอย่างที่คำนวณได้ และหากเป็นการศึกษาเชิงทดลอง ประชากรที่เข้าถึงได้ต้องมีลักษณะเฉพาะที่มีความเป็นไปได้ที่จะเกิดผลลัพธ์ทางคลินิก (Clinical end points) ในจำนวนครั้งที่เพียงสำหรับการเปรียบเทียบผลการทดลองในช่วงเวลาที่กำหนด

3. <u>กลุ่มตัวอย่าง หรือ อาสาสมัคร (Study samples, Participants, Subjects)</u>

หลังจากกำหนดประชากรที่เข้าถึงได้แล้ว จะมีการเลือกประชากรส่วนหนึ่งมาเป็นอาสาสมัครจริงของโครงการวิจัยซึ่งทำได้ 2 ลักษณะคือ กรณีที่ประชากรที่เข้าถึงได้มีขอบเขตชัดเจนและมีขนาดใหญ่กว่าขนาดของกลุ่มตัวอย่างที่คำนวณได้มาก ซึ่งส่วนใหญ่จะเป็นแบบ Community based ผู้วิจัยจะใช้เทคนิคการเลือกกลุ่มตัวอย่างแบบทราบโอกาสความน่าจะเป็น (Probability sampling) เพื่อให้ได้กลุ่มตัวอย่างที่เป็นตัวแทนที่ดีของประชากร อีกลักษณะหนึ่งคือกรณีที่การศึกษาเป็นแบบ Clinic based หรือประชากรเป้าหมายไม่ทราบขอบเขตชัดเจนหรือการศึกษาแบบ Randomized Controlled Trial ผู้วิจัยจะเชิญผู้ที่มีคุณสมบัติเข้าได้กับ Eligibility criteria และไม่มีข้อห้ามตาม Exclusion criteria เข้าเป็นอาสาสมัครของ

โครงการ โดยมักใช้การเลือกกลุ่มตัวอย่างแบบ Consecutive sampling จนได้อาสาสมัครตามจำนวนที่ต้องการ

ประชากรเป้าหมายและประชากรที่เข้าถึงได้อาจเป็นกลุ่มเดียวกันหรือไม่ก็ได้ขึ้นอยู่กับการตั้งขอบเขตการวิจัย สถิติเชิงอนุมานยอมให้ผู้วิจัยสรุปผลการศึกษาจากกลุ่มตัวอย่างไปยังประชากรที่เข้าถึงได้เท่านั้น (โดยมีข้อแม้ว่าต้องใช้เทคนิคการสุ่มตัวอย่างแบบทราบโอกาสความน่าจะเป็น) กรณีประชากรเป้าหมายและประชากรที่เข้าถึงได้มีความแตกต่างกัน ผู้วิจัยจะอ้างอิงผลการวิจัยไปยังประชากรเป้าหมายได้โดยมีเหตุผลอื่น ๆ ที่ไม่ใช่เหตุผลทางสถิติประกอบเท่านั้น เช่นเหตุผลว่าลักษณะของประชากรสองกลุ่มไม่ต่างกันหรือเหตุผลทางคลินิกอื่น ๆ เช่นเดียวกันหากไม่ได้ใช้เทคนิคการสุ่มตัวอย่างแบบสุ่มหรือทราบโอกาสความน่าจะเป็น เช่น Convenient sampling การสรุปผลการวิจัยจะไม่สามารถอ้างหลักสถิติเชิงอนุมานได้ และต้องสรุปผลด้วยความระมัดระวัง

เกณฑ์ในการรับอาสาสมัครในการวิจัยทางคลินิก (Eligibility criteria)

1. Inclusion criteria

คือเกณฑ์ที่ระบุถึงลักษณะเฉพาะ (Specification) ของประชากรซึ่งมีคุณสมบัติเหมาะสมในการเข้าร่วมโครงการวิจัยต่อไป และขณะเดียวกันก็จะเป็นตัวบอกว่าจะอนุมานผลการศึกษาไปยังใคร หรือกล่าวอีกลักษณะหนึ่งก็คือการกำหนดประชากรเป้าหมายตลอดจนประชากรที่เข้าถึงได้นั่นเอง ความยากของการกำหนด Inclusion criteria อยู่ที่ ต้องกำหนดคุณสมบัติของประชากรที่เหมาะสมกับบริบทของการวิจัยอย่างกว้าง ๆ เพื่อให้ผลการวิจัยสามารถนำไปใช้ได้ในสถานการณ์ทั่ว ๆ ไปให้มากที่สุด แต่ในขณะเดียวกันก็ต้องเป็นกลุ่มคนที่มีความเป็นไปได้ในการนำมาทำการศึกษาได้อย่างมีประสิทธิภาพ ไม่ยากลำบากหรือมีค่าใช้จ่ายสูงเกินไป

2. Exclusion criteria

เกณฑ์คัดออกหรือเกณฑ์ไม่รับ ระบุลักษณะของประชากรที่แม้จะเข้าได้กับ Inclusion criteria แต่หากนำเข้ามาศึกษาอาจทำให้มีความเสี่ยงที่จะได้ข้อมูลไม่ถูกต้องหรือไม่ครบถ้วน ซึ่งจะทำให้คุณภาพของงานวิจัยลดลง ที่สำคัญ ๆ ได้แก่ ผู้ที่มีแนวโน้มจะไม่สามารถสามารถมาให้ติดตามได้อย่างต่อเนื่อง (เช่นติดสารเสพติด หรือมีแผนจะย้ายไปอยู่ที่อื่นซึ่งไกลจากสถานที่ศึกษา) ผู้ที่ไม่สามารถให้ข้อมูล (มีปัญหาเรื่องการสื่อสาร) เกณฑ์คัดออกอื่น ๆ ได้แก่เหตุผลด้านจริยธรรมการวิจัย และปฏิเสธที่จะเข้าร่วมการวิจัย

ผู้วิจัยควรกำหนด Exclusion criteria เท่าที่จำเป็น เนื่องจากหากกำหนดไว้มากเกินไป จะทำให้มีผู้ที่ถูกคัดออกเป็นจำนวนมาก จนประชากรที่เหลือนำมาวิจัยไม่เป็นตัวแทนที่ดีพอของประชากร ไม่สามารถอนุมานผลการศึกษากลับไปยังประชากรได้ดีเท่าที่ควร และอาจทำให้ไม่สามารถหาอาสาสมัครได้ตามจำนวนที่ต้องการในระยะเวลาที่กำหนด

ตัวอย่าง

ตาราง แสดงความสัมพันธ์ระหว่างการกำหนด Inclusion และ Exclusion criteria กับ Target และ Accessible population ในการทำ RCT เพื่อหาประสิทธิผลของการขริบหนังหุ้มปลายอวัยวะเพศชายในการป้องกันการติดเชื้อ HIV จากการมีเพศสัมพันธ์

	ข้อพิจารณา	ตัวอย่าง
Inclusion criteria	*กำหนดลักษณะของประชากรที่เข้าได้กับปัญหาวิจัย และสามารถทำการวิจัยได้อย่างมีประสิทธิภาพ*	
▪ *Target population*	▪ *ลักษณะทางประชากร*	*เพศชาย, sexually active*
	▪ *ลักษณะทางคลินิก*	*HIV negative, ยังไม่เคยขริบฯ*
▪ *Accessible population*	▪ *ลักษณะทางภูมิศาสตร์*	*อาศัยอยู่ในพื้นที่วิจัย*
	▪ *ระยะเวลา*	*ไม่มีแผนจะย้ายออกภายในระยะเวลาวิจัย*
Exclusion criteria	*กำหนดลักษณะของประชากรที่ไม่สามารถนำมาศึกษาได้เนื่องจาก:*	
	▪ *มีแนวโน้มติดตามได้ยาก*	*ติดสารเสพติด*
	▪ *ข้อจำกัดทางคลินิก*	*เลือดหยุดยาก*
	▪ *ข้อจำกัดด้านจริยธรรม*	*มีข้อบ่งชี้ในการขริบฯ เพื่อการรักษา*

การสุ่มเลือกอาสาสมัคร (Randomization)

ในการศึกษาเชิงทดลอง วิธีการระบุว่าอาสาสมัครรายใดจะอยู่ในกลุ่มที่ได้รับสิ่งทดลอง (Intervention) หรืออยู่ในกลุ่มที่ไม่ได้รับสิ่งทดลอง (Control) คือกระบวนการที่เรียกว่า Exposure allocation หากดำเนินการโดยการสุ่ม คืออาสาสมัครแต่ละคนจะถูกจัดให้อยู่ในกลุ่มใด เกิดขึ้นโดยโชคหรือความบังเอิญเท่านั้นจะเรียกว่า Random allocation หมายถึงผู้วิจัยไม่มีส่วนชี้นำหรือเจาะจงในการเลือก ในกรณีของการเลือกตัวอย่างแบบ Consecutive sampling ผู้วิจัยต้องไม่สามารถคาดเดาได้ว่าอาสาสมัครรายต่อไปจะได้อยู่กลุ่มไหน

การแบ่งกลุ่มอาสาสมัครโดยการสุ่มเป็นคุณลักษณะที่สำคัญที่สุดของการศึกษาแบบ RCT ซึ่งหากทำอย่างปราศจากอคติและดำเนินการอย่างต่อเนื่องจนมีขนาดตัวอย่างที่มากพอ จะทำให้อาสาสมัครทั้งสองกลุ่มมีลักษณะพื้นฐาน (Basic characteristics) ตลอดจนปัจจัยกวน (Confounding variables) ต่าง ๆ ที่อาจจะมี ทั้งที่คาดหมายและไม่คาดหมาย ในสัดส่วนที่เท่ากันโดยปริยาย ความแตกต่างของผลการศึกษาในสองกลุ่มหากมีจึงเกิดจากการได้รับ หรือไม่ได้รับสิ่งที่ต้องการทดสอบเท่านั้น

วิธีการสุ่มที่ดี ต้องมีการวางแผนให้แน่ใจว่า ผลที่ได้เกิดโดยบังเอิญจริง ๆ การสุ่มปกติจะทำภายหลังจากอาสาสมัครที่ถูกเลือกมา (Sampling) ผ่านการคัดกรองด้วยเกณฑ์ Inclusion และ Exclusion criteria ตลอดจนผ่านกระบวนการให้คำยินยอม (Informed consent process) ในการเข้าร่วมเป็นอาสาสมัครของ

โครงการเรียบร้อยแล้ว วิธีการอาจทำได้ตั้งแต่การใช้ตารางเลขสุ่ม ใช้โปรแกรมคอมพิวเตอร์ การโทรศัพท์สอบถามผลการสุ่มจากศูนย์ข้อมูลงานวิจัย หรือการเปิดซองปิดผนึกที่กำหนดผลการสุ่มตามลำดับการรับอาสาสมัครเข้าสู่โครงการไว้ล่วงหน้า

****

เอกสารอ้างอิง

1. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
2. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.
3. Hennekens HC, Buring EJ. Epidemiology in medicine. Boston; Williams & Wilkins: 1987.
4. Danial WW. Biostatistics Basic concepts and Methodology for the Health Sciences. 9th Edition. Wiley@Sons, Inc, 2010

****

แบบฝึกหัด

1. ข้อใดใช้กำหนด Accessible population ในการวางแผนงานวิจัย

   A. ลักษณะทางประชากรศาสตร์
   B. ขอบเขตทางด้านภูมิศาสตร์
   C. การยินยอมเข้าร่วมโครงการ
   D. ลักษณะทางคลินิก
   E. งบประมาณ

2. ข้อใดเรียงลำดับในการดำเนินการวิจัยแบบ Randomized controlled trial ได้ถูกต้อง

   A. Sampling, Informed consent process, Eligibility criteria check, Randomization
   B. Sampling, Eligibility criteria check, Informed consent process, Randomization
   C. Eligibility criteria check, Informed consent process, Randomization, Sampling
   D. Eligibility criteria check, Sampling, Informed consent process, Randomization
   E. Eligibility criteria check, Informed consent process, Sampling, Randomization

1.11.2.5 Outcome assessment

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

ตัวแปรผลการวิจัย (Outcome variables, Clinical end points)

ในการวิจัยเชิงทดลองทางคลินิก (Clinical trials) ผู้วิจัยจะเปรียบเทียบผลการรักษาหรือผลการป้องกันโรคระหว่างอาสาสมัครกลุ่มที่ได้รับและไม่ได้รับ Intervention นิยมเรียกผลการรักษาหรือผลการป้องกันโรคนี้ว่า Clinical outcomes หรือ Clinical end points กรณีที่มี End points หลายอย่าง ตัวหรือกลุ่มที่สำคัญที่สุดจะเรียกว่า Primary end point(s) และที่สำคัญรอง ๆ ลงไปจะเรียกว่า Secondary end point(s)

การวัด Outcome ถือเป็นหัวใจสำคัญในการได้มาซึ่งผลการวิจัย แพทย์จึงควรทราบถึงประเภท หลักการ ตลอดจนเทคนิคที่ใช้ในการวัด Outcome เพื่อให้ผลการศึกษาที่ได้ถูกต้องตรงตามความเป็นจริง และมีอคติน้อยที่สุด

การแบ่งประเภทของ Outcomes ทางสถิติ

ในทางสถิติ เราอาจแบ่ง Health outcomes ออกได้เป็น 3 กลุ่ม ตามลักษณะของข้อมูล และเทคนิคในการวิเคราะห์ข้อมูล คือ

1. <u>ข้อมูลบอกปริมาณ (Quantitative outcome)</u> ได้แก่ข้อมูล (ตัวแปร) ที่วัดออกมาเป็นค่าตัวเลข (Interval หรือ Ratio scale) เช่น น้ำหนักแรกคลอดของทารก หรือ ระดับน้ำตาลในเลือดภายหลังการรักษา ข้อมูลลักษณะนี้มีข้อดีคือมีรายละเอียดมาก โอกาสตรวจพบความแตกต่างทางสถิติได้ง่ายกว่า (Statistical power สูง) ทำให้ไม่ต้องใช้ขนาดของกลุ่มตัวอย่างใหญ่มาก
2. <u>ข้อมูลบอกลักษณะ</u> (Qualitative outcome) ใช้มากในการวิจัยทางการแพทย์ ส่วนใหญ่จะเป็นข้อมูล (ตัวแปร) บอกลักษณะแบบ Dichotomous คือผลลัพธ์มีความเป็นได้เพียง 2 ลักษณะ เช่น เสียชีวิต/รอดชีวิต หาย/ไม่หาย ข้อมูลลักษณะนี้ Statistical power จะขึ้นกับจำนวนครั้งของการเกิดเหตุการณ์ที่สนใจมากกว่าจำนวนกลุ่มตัวอย่างทั้งหมด
3. <u>ข้อมูลเป็นระยะเวลา</u> (Time to event outcome) ในการวิจัยที่มีการติดตามผู้ป่วยหรืออาสาสมัครไปในอนาคต เช่น Cohort studies หรือ RCT ในกรณีการศึกษาการพยากรณ์โรค (Prognosis) เราอาจสนใจตัวแปรที่เป็นระยะเวลาที่ดำเนินไปสู่เหตุการณ์ที่สนใจ เรียกว่า Survival data ในการศึกษาลักษณะนี้ ผู้วิจัยประมาณโอกาสความน่าจะเป็นที่โดยเฉลี่ยแล้วผู้ป่วยลักษณะหนึ่ง ๆ จะเกิด Outcome ที่สนใจ ณ ช่วงเวลาต่าง ๆ การวิเคราะห์ข้อมูลลักษณะนี้มีวิธีการเฉพาะเรียกว่า Survival analysis และแม้ชื่อการวิเคราะห์จะหมายถึงศึกษาอัตราการเสียชีวิตหรือรอดชีวิต แต่ก็เป็นเทคนิคที่สามารถใช้กับ Outcomes อื่น ๆ ที่เป็น Dichotomous variables ได้ด้วย เช่นโอกาสการกลับเป็นซ้ำ (Relapse) เป็นต้น

หลักการกำหนด และวัด Outcomes

1. <u>กำหนด Outcome variables ให้เหมาะสมสอดคล้องกับคำถามวิจัย</u> ซึ่งต้องเริ่มจากการตั้งคำถามวิจัยที่มีความชัดเจน ไม่เป็นลักษณะนามธรรมที่วัดผลได้ยาก การตั้งคำถามวิจัยอย่างเป็นรูปธรรมที่วัดผลได้ชัดเจน จะช่วยให้สามารถกำหนด Outcomes ที่ตรงกับวัตถุประสงค์หรือสมมติฐานการวิจัย และมีความชัดเจนว่าสิ่งที่ต้องการวัดคืออะไรและวัดอย่างไร Outcome variable ที่เลือกควรสามารถวัดได้ด้วยความถูกต้อง แม่นยำ โดยมีอคติน้อยที่สุด มีความไวและความจำเพาะสูง เช่นการวัดระดับน้ำตาลในเลือดตอนเช้าหลังงดอาหารหลังเที่ยงคืนอาจไม่สะท้อนความสม่ำเสมอในการควบคุมน้ำตาลได้ ผู้วิจัยจึงเลือก

ประเมินด้วยระดับน้ำตาลในเม็ดเลือดแดง Hemoglobin a1c (ซึ่งสามารถประเมินระดับน้ำตาลในช่วง 3 เดือนที่ผ่านมาได้ดีกว่า)

2. กำหนดวิธีวัด และ Outcome ที่มีลักษณะเป็น Objective มากกว่า Subjective การกำหนด Outcome variables ผู้ศึกษาควรหลีกเลี่ยงการเลือกตัวแปรรวมถึงวิธีวัดที่ต้องใช้วิจารณญาณ หรือความเห็นของผู้วัดประกอบมากเพราะอาจทำให้เกิดอคติในการวัดได้ ตัวแปรและวิธีการวัดแบบนี้ถือว่ามี Subjectivity สูง ส่วนใหญ่ต้องอาศัยการประเมิน ตีความ หรือให้คุณค่าโดยผู้สังเกต เช่นผู้วิจัยต้องสังเกตอากับกิริยาของอาสาสมัครเพื่อวัดภาวะเครียดซึ่งเป็นวิธีที่ไม่แน่นอน เพื่อลดอคติในการวัดผลการศึกษา ผู้วิจัยควรกำหนดตัวแปรและวิธีวัดผลที่มี Objectivity สูง คือตัวแปรที่ได้ข้อมูลจากการใช้เครื่องมือในการวัดมากกว่าผลจากการสังเกต เช่นใช้การวัดระดับ Adrenalin ในเลือด แทนการสังเกตอารมณ์ความรู้สึกในการวัดภาวะเครียด เป็นต้น

3. การสังเกต/วัด Outcomes อย่างต่อเนื่องและครบถ้วน การศึกษาเชิงทดลองส่วนใหญ่ต้องมีการติดตามวัด outcomes ที่สนใจไปตลอดช่วงเวลาที่ทำการศึกษา ในกรณีที่ระยะเวลาดังกล่าวสั้นมักไม่ค่อยมีปัญหาในการติดตามอาสาสมัคร แต่ในการวิจัยที่มีระยะเวลานาน หากมีอาสาสมัครที่ไม่สามารถติดตามได้ (Loss to follow up) จำนวนมาก อาจทำให้เกิดอคติในการสรุปผลการศึกษาได้ ถ้าผู้ที่ไม่สามารถติดตามได้ในกลุ่มศึกษา และกลุ่มเปรียบเทียบ มี Outcomes ที่แตกต่างกัน เพื่อลดโอกาสการเกิดอคติดังกล่าว ผู้วิจัยควรพยายามติดตามอาสาสมัครให้ได้ครบถ้วนมากที่สุดเท่าที่จะทำได้

4. ลดโอกาสการเกิดอคติในการวัด Outcome กรณีที่ผู้วิจัย และ/หรือ อาสาสมัครทราบว่าอาสาสมัครแต่ละคนอยู่ในกลุ่มศึกษาหรือกลุ่มเปรียบเทียบ และ Outcome มีลักษณะ Subjectivity สูง เช่น ความรุนแรงของโรคหรือการเกิดผลข้างเคียงของการรักษา ผู้วิจัยมักจะมองหา Outcome ในอาสาสมัครกลุ่มที่ได้รับ Intervention มากกว่าในกลุ่ม Control ซึ่งอาจนำไปสู่อคติในการวัด Outcome ได้ วิธีป้องกันปัญหานี้ คือควรทอดเวลาการรับรู้ดังกล่าวออกไปให้นานที่สุด เทคนิคที่ช่วยให้สิ่งนี้เกิดขึ้นได้ที่สำคัญได้แก่ การ Blind โดยใช้ Placebo ช่วย

Surrogate outcomes

ในการศึกษาทางคลินิก บ่อยครั้งที่ผลลัพธ์ที่แท้จริงทางคลินิก (Clinical outcome) ต้องใช้เวลานานมากในการเกิดหรือมีจำนวนเหตุการณ์น้อยมาก ซึ่งทำให้ต้องใช้ขนาดตัวอย่างมากเกินกว่าจะทำการศึกษาได้ หรือใช้เวลาในการศึกษานานมาก จึงมักมีการกำหนดตัวแปรที่จะใช้วัดผลการศึกษาขึ้นแทน Clinical outcome จริงขึ้น เรียกว่า Surrogate outcome หรือ Surrogate marker เช่นการวัดอัตราตายจากเส้นเลือดหัวใจอุดตันในผู้ป่วยที่รักษาด้วยยาลดไขมันในเลือดอาจใช้เวลาศึกษานานมาก ผู้วิจัยจึงเลือกวัดระดับไขมันในเลือดหลังการรักษาแทน เป็นต้น

Adverse outcomes

ในการวิจัยทางคลินิก โดยเฉพาะอย่างยิ่งการศึกษาเชิงทดลอง ที่มีการประเมินประสิทธิผลของวิธีการรักษาหรือการป้องกันโรค ผู้วิจัยต้องวางแผนในการวัดผลการศึกษา ให้ครอบคลุมผลด้านลบที่อาจเกิดจากตัว Intervention ด้วยเสมอ ผลด้านลบนี้เรียกว่า Adverse outcomes หรือ Adverse effects ประโยชน์ของการเก็บข้อมูลลักษณะนี้ก็เพื่อความปลอดภัยของอาสาสมัครระหว่างทำการศึกษา และเป็นข้อมูลประกอบการพิจารณาถึงการนำ Intervention ไปใช้จริงในผู้ป่วย หรือประชากรทั่วไปในอนาคต

เนื่องจากผลข้างเคียงด้านลบจากสิ่งทดลองมักเกิดไม่บ่อยและเราไม่ได้คำนวณขนาดของกลุ่มตัวอย่างให้ครอบคลุมกรณีนี้ไว้ในตอนต้น การเกิดเหตุการณ์ผิดปกติด้านลบเพียงรายเดียว ถ้ามีความรุนแรงมาก เช่น ทำให้อาสาสมัครมีความพิการหรือถึงขั้นเสียชีวิต จึงถือว่ามีความสำคัญที่จะต้องตรวจสอบเสมอ ไม่ว่าเป็นผลจากสิ่งทดลองที่ศึกษาหรือไม่ อย่างไรก็ตามบางครั้งผลด้านลบอาจตรวจไม่พบระหว่างการทำ RCT (Phase III clinical trial) เนื่องจากจำนวนอาสาสมัครและระยะเวลาที่ติดตามน้อยเกินไป จึงต้องอาศัยการศึกษาแบบอื่นในการทราบถึงผลเสียนี้ เช่น Case-control study ในระยะหลังจากที่ยาออกวางตลาดแล้ว (Phase IV clinical trial, post marketing surveillance) จึงจะพบหลักฐานว่ายามีผลข้างเคียงอันตรายจริง

Interim analysis

ในการศึกษาแบบ RCT บางครั้งผู้วิจัยจะกำหนดไว้ล่วงหน้าว่าจะมีการนำผลการศึกษามาวิเคราะห์เพื่อทดสอบสมมติฐานทางสถิติเป็นระยะ ๆ ระหว่างทำการศึกษา ว่าผลมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือยัง เรียกว่าการทำ Interim analysis หรือการกำหนด Early stopping rules หากการวิเคราะห์พบว่ากลุ่มที่ได้รับ Intervention มีผลการรักษาแตกต่างจากกลุ่ม Control เป็นที่ชัดเจนแล้ว ก็ไม่จำเป็นต้องติดตามอาสาสมัครต่อไปจนครบระยะเวลาที่กำหนดไว้เดิม ควรยุติการศึกษาก่อนกำหนดด้วยเหตุผลด้านจริยธรรมวิจัยในแง่ที่อาสาสมัครที่อยู่ในกลุ่มเปรียบเทียบจะมีโอกาสได้รับการรักษาโรคหรือการป้องกันโรคใหม่ที่ได้ผลดีนี้เร็วขึ้นโดยที่ไม่ต้องรอเวลาให้เนิ่นนานออกไปอีก นอกจากนั้นยังเป็นการลดค่าใช้จ่ายและระยะเวลาในการศึกษาลงอีกด้วย

อย่างไรก็ตาม บางครั้งผู้วิจัยอาจไม่ได้วางแผนการทำ Interim analysis ตั้งแต่ต้น แต่อาจพิจารณาดำเนินการภายหลัง หรือถูกสั่งให้ดำเนินการ โดยคณะกรรมการที่ควบคุมด้านจริยธรรมและความปลอดภัยของการวิจัย เหตุผลการทำ Interim analysis ในลักษณะนี้มักเนื่องมาจากมีผลการศึกษาเรื่องเดียวกันหรือใกล้เคียงกันที่ทำโดยคณะวิจัยอื่นออกมาในขณะนั้นว่าการรักษาหรือการป้องกันโรคที่กำลังศึกษาได้ผลแล้ว หรือการวิจัยมีแนวโน้มว่าจะไม่สามารถสรุปผลได้เนื่องจากเหตุการณ์ทางคลินิกที่เป็น Outcome เกิดน้อย

การปิดบังไม่ให้อาสาสมัครทราบว่าตนเองอยู่ในกลุ่มทดลอง หรือกลุ่มเปรียบเทียบ (Blinding, Masking)

การสุ่มเลือกอาสาสมัครที่กล่าวถึงข้างต้น ช่วยกำจัดตัวแปรกวนที่มี ณ ตอนเริ่มต้นการศึกษาเท่านั้น แต่ไม่ช่วยป้องกันอคติที่อาจเกิดขึ้นระหว่างการดำเนินการวิจัย และการประเมิน Outcome อคติเหล่านี้ ส่วนหนึ่งเกิดจากการที่ผู้ที่เกี่ยวข้องทราบว่าอาสาสมัครรายใดอยู่ในกลุ่มทดลองหรือกลุ่มเปรียบเทียบ ตัวอย่างเช่น หากแพทย์ทราบว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มใดก็จะมีแนวโน้มที่จะมองหา Outcome ที่ตนเองสนใจในอาสาสมัครที่ทราบว่าเป็นกลุ่มทดลองมากกว่าอาสาสมัครอีกกลุ่มหนึ่ง หรืออาสาสมัครที่ทราบว่าตนเองอยู่ในกลุ่มเปรียบเทียบอาจไม่กินยาที่ได้รับอย่างสม่ำเสมอเพราะคิดว่ากินไปก็ไม่มีประโยชน์ หรือไปแสวงหาการรักษาอื่นเพิ่มเติมที่คิดว่าจะทำให้การเจ็บป่วยของตนเองดีขึ้น ทำให้เกิดการได้รับสิ่งทดลองโดยไม่ตั้งใจ (Unintended intervention) หรือการเก็บรวบรวมข้อมูลที่มีอคติได้

มาตรการที่ช่วยลดปัญหาดังกล่าวคือการปิดบังไม่ให้ผู้เกี่ยวทราบว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มใด เรียกว่าการ Blind ซึ่งมี 3 ระดับ คือ

1. <u>Single-blind</u> หมายถึง เฉพาะตัวอาสาสมัครเท่านั้นที่ไม่ทราบว่าตนเองอยู่ในกลุ่มทดลองหรือกลุ่มเปรียบเทียบ

2. Double-blind ทั้งผู้รวบรวมข้อมูลและอาสาสมัครไม่ทราบว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มที่ได้รับ Intervention หรือไม่ ถือเป็นมาตรฐานการ Blind สำหรับการศึกษาแบบ RCT ทำได้โดยการกำหนดให้กระบวนการดูแลอาสาสมัครแต่ละรายมีลักษณะเหมือนกัน มีมาตรฐานเดียวกัน ผู้ให้การดูแลไม่สามารถสังเกตได้จากรูปแบบการให้บริการว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มใด สำหรับการวิจัยเพื่อศึกษาประสิทธิภาพของยาหรือวัคซีน ทำได้โดยให้เม็ดยาหรือวัคซีนที่ใช้ในกลุ่มเปรียบเทียบมีลักษณะทางกายภาพเหมือนกับยา หรือวัคซีนจริง เรียกว่า Placebo
3. Triple-blind เหมือนกับ Double-blind แต่มีที่เพิ่มขึ้นมาคือ ระหว่างการวิเคราะห์ข้อมูลผู้วิเคราะห์จะไม่ทราบว่าข้อมูลกลุ่มใดเป็นกลุ่มที่ได้รับ Intervention หรือไม่ได้รับ เพื่อลดอคติ

อย่างไรก็ตาม บางกรณีอาจไม่สามารถทำการ Blind ได้ เนื่องจากตัว Intervention เป็นกิจกรรมที่มองเห็นได้ชัด เช่นหัตถการผ่าตัด หลักสูตรการให้ความรู้ หรือการศึกษาแบบ Community Intervention trial ที่การศึกษามีกลุ่มตัวอย่างเป็นชุมชน นอกจากนั้นในการศึกษาเกี่ยวกับการรักษาโรค เมื่อการทดลองดำเนินไปถึงจุดที่ผลการรักษาเริ่มมีความแตกต่างกัน แพทย์และผู้ป่วยที่เข้าร่วมโครงการจะเริ่มสังเกตเห็นถึงความแตกต่างของผลการรักษาหรือผลข้างเคียงของยา และคาดเดาได้ว่าอาสาสมัครแต่ละรายอยู่ในกลุ่มไหน

เมื่อการศึกษาสิ้นสุด ผู้วิจัยจะทำการ Un-blind คือเปิดระหัสว่าอาสาสมัครแต่ละคนอยู่ในกลุ่มใด เพื่อประโยชน์ในการวิเคราะห์ข้อมูลและสรุปผลการศึกษา หากผลการรักษาในกลุ่มที่ได้รับสิ่งทดลองดีกว่ากลุ่มเปรียบเทียบอย่างมีนัยสำคัญ ถือเป็นข้อกำหนดทางด้านจริยธรรมที่ผู้วิจัยต้องให้การรักษานั้นแก่กลุ่มควบคุมด้วย เช่นหากการทดลองวัคซีนได้ผลว่าวัคซีนมีประสิทธิผลในการป้องกันโรค ผู้วิจัยต้องกลับไปให้วัคซีนกับอาสาสมัครที่อยู่ในกลุ่มเปรียบเทียบ การ Un-blind อีกลักษณะหนึ่ง คือการ Un-blind ก่อนกำหนด ในกรณีมีเหตุฉุกเฉินด้านความปลอดภัย ที่ทำให้จำเป็นต้องทราบว่าอาสาสมัครรายนั้น ๆ อยู่ในกลุ่มใด เพื่อประโยชน์ในการแก้ไขปัญหาทางด้านคลินิกที่อาจเกิดขึ้น การวิจัยทางคลินิกจึงต้องมีทีมข้อมูลที่เก็บระหัสของอาสาสมัครไว้และมีความพร้อมในการ Un-blind ตลอดเวลา สำหรับการวิจัยที่ไม่มีการ Blind จะเรียกว่า Open label study

****

เอกสารอ้างอิง

1. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
2. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.

****

แบบฝึกหัด

1. การวิจัยรูปแบบใด สามารถมี Health outcome แบบ Time to event ได้

A. Cohort

B. Case-control

C. Cross-sectional

D. Case series

E. Case report

2. ในกรณีทั่วไป อะไรเป็นผลที่ตามมาจากการทำ Interim analysis

A. ค่าใช้จ่ายที่เพิ่มขึ้น

B. การละเมิดจริยธรรมวิจัย

C. ระยะเวลาการทำวิจัยที่สั้นลง

D. การสรุปผลการวิจัยอย่างมีอคติ

E. การออกแบบงานวิจัยใหม่ที่มีคุณภาพ

3. การวัดอัตราตายจากเส้นเลือดหัวใจอุดตันในผู้ป่วยที่รักษาด้วยยาลดไขมันในเลือด อาจใช้เวลาศึกษานานมาก ผู้วิจัยจึงเลือกวัดระดับไขมันในเลือดหลังการรักษาแทน การวัดผลการศึกษาดังกล่าวเข้าได้กับข้อใดมากที่สุด

A. Time to event outcome

B. Qualitative outcome

C. Quantitative outcome

D. Surrogate outcome

E. Adverse outcome

1.11.2.6 Internal and external validity

การวิจัยทางการแพทย์และการสาธารณสุขที่เป็นการศึกษาเชิงวิเคราะห์ มีจุดมุ่งหมายสำคัญเพื่อทดสอบสมมติฐาน ประมาณค่า หรือหาความสัมพันธ์ทางสถิติระหว่างตัวแปรต้น (Independent variable) กับตัวแปรตาม (Outcome variable) โดยพยายามให้ได้ผลการศึกษาที่ถูกต้องความความเป็นจริง (valid) มากที่สุด ความถูกต้องดังกล่าวมี 2 ระดับคือ ความถูกต้องภายในและความถูกต้องภายนอก

ลำดับขั้นของความถูกต้องของสรุปผลการวิจัย

1. ความถูกต้องภายใน (Internal validity)

โดยทั่วไปหมายถึง การสรุปผลการวิจัยถูกต้องตามความเป็นจริงมากน้อยเพียงไร ที่เรียกว่าเป็น "ภายใน" เนื่องจากเป็นข้อสรุปที่ใช้อธิบายเฉพาะกับปรากฏการณ์ที่เกิดกับกลุ่มตัวอย่างจริงในการวิจัยเท่านั้น ความถูกต้องภายในของผลการศึกษาจึงขึ้นกับคุณภาพของงานวิจัยเอง ซึ่งหากการศึกษามีข้อผิดพลาดมาก ข้อสรุปที่ได้ก็จะคลาดเคลื่อนจากความเป็นจริงมากไปด้วย ข้อผิดพลาดของการทำวิจัยมีได้ตั้งแต่ระยะการออกแบบการวิจัย ระยะการดำเนินการเก็บรวบรวมข้อมูล และระยะการวิเคราะห์ข้อมูล

อย่างไรก็ตาม ตำราบางเล่มให้ความหมายของ Internal validity แตกต่างออกไป โดยให้ความหมายที่กว้างกว่า คือนอกจากหมายถึงความจริงที่เกิดขึ้นในการวิจัยแล้ว ยังหมายรวมถึงความสามารถในการสรุปว่าสิ่งที่พบในการทำวิจัยเป็นความจริงที่เกิดขึ้นในประชากรเป้าหมาย (Target population) ด้วยหรือไม่

ตัวอย่าง

การวิจัยแบบ RCT ผู้วิจัยสรุปว่าการขริบหนังหุ้มปลายอวัยวะเพศชาย ช่วยลดโอกาสเสี่ยงต่อการติดเชื้อ HIV ในอาสาสมัครของโครงการลงร้อยละ 51 เทียบกับผู้ไม่ได้ขริบฯ อย่างมีนัยสำคัญทางสถิติ ความถูกต้องภายในในที่นี้หมายถึง ผลสรุปของการวิจัยดังกล่าวถูกต้องตามความเป็นจริงที่เกิดขึ้นกับอาสาสมัครที่ร่วมในการวิจัยนี้มากน้อยเพียงไรนั่นเอง

ข้อผิดพลาดของการวิจัยซึ่งส่งผลต่อ Internal validity แบ่งได้เป็น 2 ลักษณะใหญ่ ๆ คือ ความคลาดเคลื่อนเชิงระบบ และความคลาดเคลื่อนเชิงสุ่ม

4. ความคลาดเคลื่อนเชิงระบบ (Systemic error) หรือเรียกอีกอย่างหนึ่งว่าอคติ (Bias) หมายถึงความคลาดเคลื่อนของขั้นตอนกระบวนการทำวิจัยที่ส่งผลให้เกิดความผิดพลาดของผลการศึกษาไปจากความเป็นจริงในทิศทางใดทิศทางหนึ่ง การเพิ่มขนาดกลุ่มตัวอย่างไม่ช่วยป้องกันหรือแก้ไขปัญหานี้ Bias อาจแบ่งออกได้เป็น 3 กลุ่มใหญ่ ๆ คือ

   2.1 *Selection bias* หมายถึงความคลาดเคลื่อนเนื่องจากอคติในการเลือกกลุ่มตัวอย่างจากประชากร โดยสัดส่วนของกลุ่มตัวอย่างต่าง ๆ ตามตาราง 2 X 2 ทางระบาดวิทยา ได้แก่ Exposed-disease, Non exposed-disease, Exposed-non disease, และ Non exposed-non disease) ที่ถูกเลือกมาทำการศึกษา ไม่สอดคล้องกับสัดส่วนจริงในประชากร

   ตัวอย่าง

   การศึกษาแบบ Cross-sectional study เพื่อหาความสัมพันธ์ระหว่างการสูบบุหรี่กับการเกิดโรคถุงลมปอดโป่งพอง ผู้วิจัยไม่พบความสัมพันธ์ระหว่างตัวแปรทั้งสอง ทั้ง ๆ ที่การสูบบุหรี่

เป็นสาเหตุของโรคถุงลมปอดโป่งพอง เนื่องจากผู้ป่วยที่สูบบุหรี่จำนวนหนึ่งได้เสียชีวิตไปก่อนแล้ว ไม่อยู่ให้สำรวจ (สัดส่วนของอาสาสมัครกลุ่ม Exposed-disease ที่ถูกเลือกมาจากประชากร ต่ำกว่ากลุ่มอื่น ๆ)

2.2 *Information bias* หมายถึงความคลาดเคลื่อนของผลการศึกษา เนื่องจากความแตกต่างในกระบวนการเก็บรวบรวมข้อมูล ทำให้เกิดความผิดพลาดหรืออคติในข้อมูลที่รวบรวมมาได้ ซึ่งอาจเกิดจากตัวผู้เก็บข้อมูลเอง จากรูปแบบวิธีการ หรือจากเครื่องมือที่ใช้ก็ได้

ตัวอย่าง

การศึกษาเกี่ยวกับระดับความดันโลหิตที่ใช้เครื่องวัดที่ไม่ได้มาตรฐาน โดยเครื่องวัดความดันวัดค่าความดันโลหิตได้สูงกว่าความเป็นจริงเสมอ

ในการศึกษาแบบ Case-control study เพื่อหาความสัมพันธ์ระหว่างประวัติการได้รับยาปฏิชีวนะในมารดาระหว่างตั้งครรภ์ มารดาที่มีบุตรป่วยเป็นโรค (Case) มีแนวโน้มจะจำการกินยาระหว่างตั้งครรภ์ได้มากกว่ามารดาที่ลูกปกติ (Control) ขณะที่มารดาที่บุตรปกติบางคนอาจกินยาตัวนั้น แต่จำไม่ได้

2.3 *Confounding* หมายถึงการสรุปความสัมพันธ์ระหว่างปัจจัยเสี่ยงที่สนใจศึกษา กับการเกิดโรคอย่างผิด ๆ เนื่องจากปัจจัยอื่น (Confounding factor) โดยตัวแปรที่จะเป็น Confounding factor ต้องมีความสัมพันธ์ทั้งกับปัจจัยเสี่ยงที่สนใจศึกษา และ Outcome อย่างไรก็ตาม ตำราบางเล่มไม่จัดความผิดพลาดของการสรุปผลการศึกษาแบบนี้ว่าเป็น Bias แต่จะแยกออกมาต่างหาก

ตัวอย่าง

การวิจัยพบว่าการดื่มกาแฟเป็นปัจจัยเสี่ยงของการป่วยด้วยโรคหลอดเลือดหัวใจ ซึ่งต่อมาพบว่าเป็นข้อสรุปที่ผิดพลาด เนื่องจากผู้ที่ดื่มกาแฟสูบบุหรี่มากกว่าผู้ที่ไม่ดื่มกาแฟ และการสูบบุหรี่เป็นสาเหตุของโรคหัวใจและหลอดเลือด (การสูบบุหรี่มีความสัมพันธ์กับทั้งการดื่มกาแฟ และโรคหลอดเลือดหัวใจ) ดังนั้นความสัมพันธ์ระหว่างการดื่มกาแฟกับการป่วยด้วยโรคหัวใจและหลอดเลือดอันที่จริงเกิดจากผลของการสูบบุหรี่ ซึ่งในที่นี้ถือเป็น Confounding factor

5. ความคลาดเคลื่อนเชิงสุ่ม (Random error) ขนาดตัวอย่างมีผลอย่างมากต่อความถูกต้องของผลการศึกษา หากจำนวนตัวอย่างน้อยเกินไป อาสาสมัครที่สุ่มได้อาจไม่เป็นตัวแทนที่ดีของประชากรที่เราอยากจะอนุมานผลการศึกษาไปถึง แม้เราจะเลือกกลุ่มตัวอย่างด้วยการสุ่มอย่างดีที่สุดแล้วก็ตาม เนื่องจากความบังเอิญ หรือความแปรปรวน (variation) ตามธรรมชาติ ขนาดของกลุ่มตัวอย่างที่เพิ่มขึ้นจะช่วยลด Random error ลงได้

2. ความถูกต้องภายนอก (External validity, Generalizability)

หมายถึง ผลการศึกษายังเป็นจริงมากน้อยเพียงใด ในประชากรกลุ่มอื่นนอกเหนือจากในอาสาสมัครโครงการ และในสถานการณ์อื่น ๆ ซึ่งสามารถแบ่งออกเป็นลำดับขั้นย่อย ๆ ได้อีก 2 ระดับ คือ

1) การอนุมานผลการศึกษาจากกลุ่มตัวอย่างไปยังประชากรที่เข้าถึงได้ การอนุมานในระดับนี้จะถูกต้องเพียงใดขึ้นกับวิธีการเลือกกลุ่มตัวอย่างเป็นสำคัญ หากใช้วิธีแบบทราบโอกาสความน่าจะเป็น กลุ่มตัวอย่างของศึกษามีจำนวนเพียงพอ และไม่มี Exclusion criteria ที่จะทำให้

ลักษณะของกลุ่มตัวอย่างแตกต่างจาก Accessible population มากเกินไป ผลการวิจัยจะมี External validity สูง

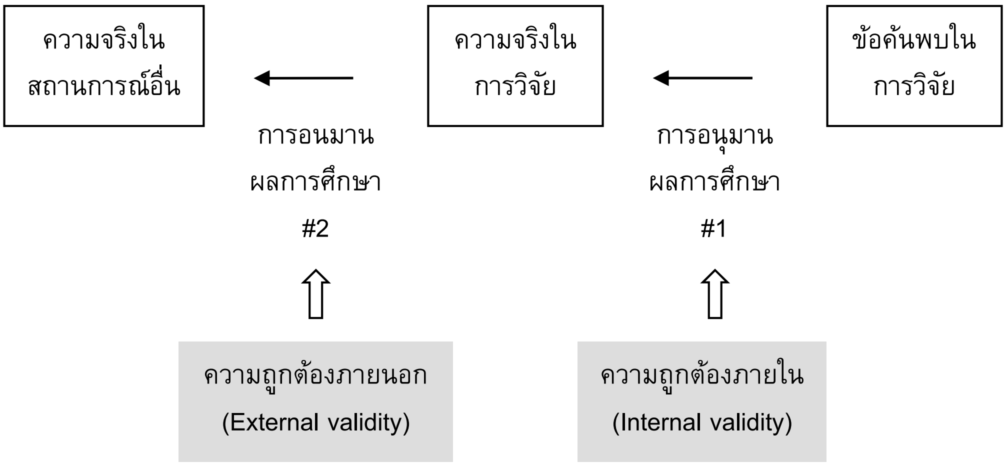
2) <u>การอนุมานผลการศึกษาจากประชากรที่เข้าถึงได้ไปยังประชากรเป้าหมาย</u> การอนุมานผลการศึกษาระดับนี้มักมีความถูกต้องแม่นยำน้อยกว่าการอนุมานแบบแรก ขึ้นกับธรรมชาติของคำถามวิจัยนั้น ๆ โดยทั่วไปผลการศึกษาของคำถามวิจัยที่เกี่ยวกับตัวแปรทางชีววิทยา เช่น การศึกษาที่เกี่ยวกับผลของยาหรือวัคซีนจะสามารถนำไปใช้ในประชากรลักษณะอื่นหรือในสถานการณ์อื่นได้กว้างขวางกว่า (External validity สูงกว่า) ความสัมพันธ์ของตัวแปรเชิงพฤติกรรมและสังคม เช่น การศึกษาผลของหลักสูตรการปรับพฤติกรรมสุขภาพ เป็นต้น

<u>*ตัวอย่าง*</u>

ผลการวิจัย	*External validity*
วัคซีนที่พัฒนาขึ้นใหม่มีประสิทธิผลในการป้องกันมะเร็งปากมดลูก	*สูง*
เทคนิคการผ่าตัดถุงน้ำดีแบบใหม่ช่วยลดผลข้างเคียงจากการผ่าตัดได้	*ปานกลาง*
หลักสูตรเพศศึกษาที่พัฒนาขึ้นโดยองค์กรพัฒนาเอกชนแห่งหนึ่ง ช่วยลดพฤติกรรมเสี่ยงทางเพศของนักเรียนมัธยมศึกษาในจังหวัดเชียงใหม่	*ต่ำ*

ปัจจัยสำคัญที่มีผลต่อความถูกต้องภายนอก คือผู้วิจัยต้องทำการศึกษาอย่างมีคุณภาพ เพื่อให้มีความถูกต้องภายในของผลการวิจัยสูงเป็นเบื้องต้นก่อน เลือกกลุ่มตัวอย่างที่มีลักษณะเป็นตัวแทนของประชากรทั่วไปได้ดี ไม่เลือกที่มีลักษณะเฉพาะมากเกินไป

แผนภูมิแสดงความสัมพันธ์ระหว่างความถูกต้องภายใน และความถูกต้องภายนอก ของการสรุปผลการวิจัย



****

เอกสารอ้างอิง

1. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
2. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.
3. Knapp G Rebecca, Miller III M Clinton. Clinical epidemiology and biostatistics. Williams & Wilkins, Baltimore, Maryland. 1992.

****

แบบฝึกหัด

1. ในการศึกษาแบบ Clinical trials การเพิ่มจำนวนอาสาสมัครช่วยลดความผิดพลาดในข้อใด

   A. Selection bias

   B. Information bias

   C. Confounding

   D. Random error

   E. Systemic error

2. เจ้าหน้าที่โครงการวิจัยวัด และลงผลความดันโลหิตของอาสาสมัครคลาดเคลื่อน โดยมักปัดเศษให้ลงท้ายด้วยเลขศูนย์ เช่น 80 mm.Hg., 90 mm.Hg, 100 mm.Hg” เหตุการณ์ข้างต้นเข้าได้กับข้อใดมากที่สุด

   A. Selection bias

   B. Information bias

   C. Confounding

   D. Random error

   E. Recall bias

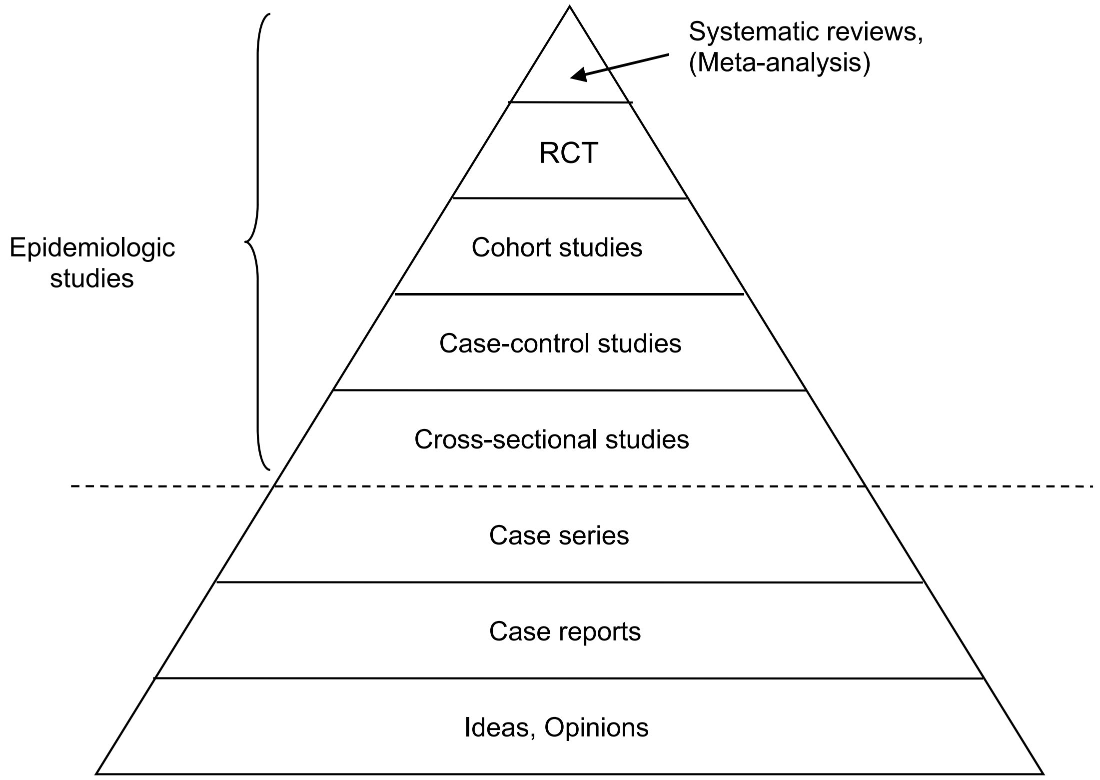
1.11.2.7 Level of evidence

ในอดีต แพทย์อาศัยความรู้ที่เล่าเรียนมาจากโรงเรียนแพทย์ การอ่านตำรับตำรา การปรึกษาแพทย์หรืออาจารย์แพทย์ที่อาวุโสกว่า ร่วมกับประสบการณ์ในการดูแลผู้ป่วยที่ผ่านมาของตนเอง เป็นเครื่องช่วยสำคัญในการตัดสินใจให้การดูแลรักษาผู้ป่วยแต่ละราย แต่ในยุคปัจจุบันที่ข้อมูลข่าวสารความรู้ทางการแพทย์ใหม่ ๆ ที่เกิดจากการทำวิจัยเพิ่มขึ้นอย่างรวดเร็วตลอดเวลา จึงเป็นการยากที่แพทย์แต่ละคนจะสามารถติดตามข้อมูลทุกอย่างได้ด้วยตนเอง ขณะที่ผู้ป่วยตลอดจนสาธารณชนก็เรียกร้องบริการที่มีคุณภาพและมีความเป็นปัจจุบันสูง จึงได้มีผู้คิดค้นศาสตร์ ตลอดจนเครื่องมือใหม่ ๆ ในการรวบรวม คัดสรร และให้คุณค่าแก่ข้อมูลองค์ความรู้เหล่านี้ เพื่ออำนวยความสะดวกให้แพทย์สามารถปฏิบัติหน้าที่ได้อย่างมีประสิทธิภาพมากขึ้น

ลำดับขั้นของหลักฐานทางการแพทย์ (Evidence level)

เอกสารวิชาการทางการแพทย์ที่เกี่ยวข้องกับเรื่องที่เราสนใจมีจำนวนมาก แต่ละชิ้นมีคุณค่าและน้ำหนักความน่าเชื่อถือที่จะนำมาใช้ประโยชน์ได้แตกต่างกันไป ขึ้นกับคุณภาพของงานวิจัย และรูปแบบการทำวิจัย โดยทั่วไปจะจัดให้ Systematic reviews มีน้ำหนักความน่าเชื่อถือสูงสุด รองลงมาได้แก่ Randomized controlled trials, Cohort studies, Case-control studies, และ Cross-sectional studies ตามลำดับ สำหรับในลำดับขั้นที่ต่ำกว่านั้นได้แก่ Case series, Case reports และ Ideas / Opinions ซึ่งอันที่จริงไม่จัดเป็นการวิจัยทางระบาดวิทยา แต่เป็นเพียงการรวบรวมข้อมูลผู้ป่วยและความคิดเห็นเท่านั้น เพื่อให้เห็นลำดับขั้นของหลักฐานทางการแพทย์ที่เป็นรูปธรรมชัดเจน นิยมใช้รูปปิรามิดแสดงดังด้านล่าง

แผนภูมิปิรามิดแสดงลำดับขั้นของหลักฐานทางการแพทย์



การแพทย์อิงหลักฐาน (Evidence-based medicine, EBM)

แม้แนวคิดที่เป็นรากฐานเกี่ยวกับเรื่องนี้จะมีมากนานแล้ว แต่ EBM ก็จัดว่าเป็นศาสตร์ทางการแพทย์แขนงที่ใหม่มาก คำว่า “Evidence-based medicine” เพิ่งจะมีการอ้างถึงเป็นครั้งแรกในเอกสารทางวิชาการเมื่อปี พ.ศ.2535 นี้เอง EBM เสนอแนะให้แพทย์ให้การวินิจฉัยตลอดจนเลือกแนวทางการดูแลรักษาผู้ป่วย โดยพิจารณาจากหลักฐานทางการแพทย์ที่มีคุณภาพและเชื่อถือได้ EBM ให้น้ำหนักน้อยต่อความเห็นของแพทย์ผู้เชี่ยวชาญหากไม่มีหลักฐานที่ดีประกอบ

ลักษณะที่ทำให้ EBM แตกต่างจากการวิจัยโดยทั่วไปคือ EBM เป็นกระบวนการที่ใช้เทคนิคทางวิทยาศาสตร์และทางสถิติที่เป็นมาตรฐานเดียวกันในการประเมินค่าผลการวิจัย และสรุปเป็นองค์ความรู้ที่ดีที่สุดสำหรับเรื่องนั้น ๆ โดยอาศัยกระบวนการทบทวนองค์ความรู้อย่างเป็นระบบ (Systematic review) และการจัดลำดับขั้นของหลักฐานทางการแพทย์ (Evidence level) เป็นเครื่องมือสำคัญ การนำหลักการของ EBM มาใช้อาจแบ่งได้เป็น 2 ระดับ คือ

1. <u>Evidence-based guidelines</u>

   หมายถึงการที่องค์กรหรือหน่วยงานต่าง ๆ นำสรุปองค์ความรู้ที่ได้จากหลักการ EBM มากำหนดแนวทาง หรือนโยบายในการป้องกันโรค คัดกรองโรค หรือให้การดูแลรักษาผู้ป่วยด้วยปัญหาสุขภาพต่าง ๆ ตัวอย่างเช่น แนวทางการดูแลรักษาผู้ป่วยที่ออกโดยสมาคมวิชาชีพแพทย์เฉพาะทาง และมาตรฐานการคัดกรองโรคของกระทรวงสาธารณสุข เป็นต้น

2. <u>Evidence-based individual decision making</u>

   หมายถึงการที่แพทย์แต่ละคน นำองค์ความรู้ที่ได้จากหลักการ EBM มาประกอบการตัดสินใจให้การดูแลรักษาผู้ป่วยแต่ละราย

การทบทวนองค์ความรู้อย่างเป็นระบบ (Systematic review)

หมายถึงการทบทวนวรรณกรรมเกี่ยวกับปัญหาสุขภาพเรื่องใดเรื่องหนึ่ง โดยผู้ทบทวนทำการสืบค้นผลการวิจัยที่เกี่ยวข้องกับเรื่องนั้นให้ได้ครบถ้วนมากที่สุด แล้วนำผลการศึกษาต่าง ๆ เหล่านั้นมาวิเคราะห์ และประมวลผลโดยใช้เทคนิคที่เป็นมาตรฐานเดียวกันเพื่อให้ได้ข้อสรุปที่เป็นองค์ความรู้ในภาพรวมต่อประเด็นนั้น ๆ EBM จึงจัด systematic review ให้อยู่ในลำดับขั้นสูงสุดของหลักฐานทางการแพทย์

เทคนิคการวิเคราะห์ข้อมูลเพื่อหาข้อสรุปอาจมีได้ทั้งเชิงคุณภาพและเชิงปริมาณ สำหรับรูปแบบเฉพาะของ Systematic review ที่เน้นวิธีการเชิงปริมาณ และอาศัยเทคนิคพิเศษทางระบาดวิทยาตลอดจนทางสถิติ ในการประมวลผลหาข้อสรุปของความสัมพันธ์ระหว่างปัจจัยที่สนใจ เรียกว่า Meta-analysis ซึ่งส่วนใหญ่เป็นการรวบรวมและวิเคราะห์ผลการวิจัยแบบ Randomized control trials หลาย ๆ อัน ที่ตอบคำถามวิจัยเดียวกัน ให้ได้ข้อสรุปเป็นตัวเลขทางสถิติที่เป็นหนึ่งเดียว ภายใต้วิธีการนี้ก่อนอื่นผลการศึกษาทั้งหมดจะถูกนำมาทดสอบทางสถิติก่อน ว่ามีลักษณะที่แตกต่างกันมากเกินกว่าจะนำมาวิเคราะห์รวมกันหรือไม่ (Test for heterogeneity) จากนั้นผลการศึกษาจากแต่ละการวิจัยจะถูกนำมาคำนวณโดยถ่วงน้ำหนักไม่เท่ากัน โดยการศึกษาขนาดใหญ่ที่มีกลุ่มตัวอย่างมาก และผลการศึกษามีค่าช่วงความเชื่อมั่น (Confidence interval) แคบจะได้รับการให้น้ำหนักในการนำมาใช้คำนวณดัชนีรวม มากกว่าการศึกษาขนาดเล็กที่มีกลุ่มตัวอย่างน้อยและผลการศึกษามีช่วงความเชื่อมั่นกว้าง

Cochrane Collaboration เกิดจากการรวมตัวกันของนักวิชาการ และผู้เชี่ยวชาญทั่วโลก เพื่อผลิตผลงานวิชาการในลักษณะของ Systematic review และนำออกเผยแพร่สู่สาธารณะ ปัจจุบัน Cochrane Database

of Systematic Reviews ซึ่งเป็นฐานข้อมูลของหน่วยงานนี้ สามารถเข้าสืบค้นได้ผ่านทางเครือข่ายอินเตอร์เนต ที่ http://www.thecochranelibrary.com

ข้อจำกัดของ EBM

EBM อาศัยการรวบรวมองค์ความรู้จากเอกสารวิชาการที่ได้รับการตีพิมพ์แล้ว นำมากลั่นกรองและสรุปเป็นข้อมูลเพื่อการนำไปใช้ ความถูกต้องจึงขึ้นอยู่กับความครบถ้วนในการรวบรวมองค์ความรู้ที่มีอยู่จริงในเรื่องนั้น ๆ อย่างไรก็ตามที่ผ่านมา มีแนวโน้มว่าการวิจัยที่ผลการศึกษาไม่เป็นไปตามสมมติฐาน มักจะไม่ได้รับการตีพิมพ์เผยแพร่ ทำให้องค์ความรู้ที่รวบรวมมาเพื่อการสรุปอาจมีอคติ (Publication bias)

คุณภาพของ Systematic review แต่ละเรื่อง ขึ้นกับคุณภาพของการวิจัยเกี่ยวกับปัญหาสุขภาพนั้น ๆ แต่ในหลายกรณี พบว่าอาจไม่มีผลการวิจัยที่มีคุณภาพสูงอย่าง RCT ที่เกี่ยวกับปัญหานั้น เนื่องจากเหตุผลหลายประการเช่น ค่าใช้จ่ายในการทำวิจัยสูง หรือลักษณะของตัวโรคเองที่พบไม่บ่อยทำให้ไม่สามารถมีขนาดกลุ่มตัวอย่างที่มากพอ หรือมีข้อจำกัดทางด้านจริยธรรมทำให้ไม่สามารถทำการศึกษาเชิงทดลองในมนุษย์ได้ ผู้ทำ Systematic reviews จึงต้องทำการสรุปทบทวนภายใต้ข้อจำกัดตามปริมาณและคุณภาพของข้อมูลงานวิจัยที่มี

นอกจากนั้นยังมีผู้กล่าวว่า EBM ทำให้แพทย์ให้การดูแลรักษาโดยอิงหลักฐานทางการแพทย์มากเกินไป ตัดสินใจตามข้อมูลตัวเลข ขาดศิลปะการดูแลผู้ป่วย ไม่ตอบสนองความต้องการทางด้านอารมณ์ความรู้สึกและสังคมของผู้ป่วย ไม่ได้ดูแลรักษาผู้ป่วยแต่ละรายอย่างองค์รวมอย่างที่ควรจะเป็น ซึ่งรวมถึงการไม่ได้พิจารณาปัจจัยแวดล้อมอื่น ๆ นอกเหนือจากข้อมูลทางคลินิกด้วย

Level of evidence

หลายหน่วยงานได้จัดลำดับขั้นของหลักฐานทางการแพทย์ สำหรับข้อมูลทางการแพทย์ลักษณะต่าง ๆ ตารางด้านล่างแสดง Level of evidence ของข้อมูลด้านการรักษาพยาบาลของมหาวิทยาลัย Oxford ประเทศอังกฤษ

Oxford Evidence Based Medicine levels for Treatment

Level	Description
1a	Systematic reviews (with homogeneity) of randomized controlled trials
1a-	Systematic reviews of randomized trials displaying worrisome heterogeneity
1b	Individual randomized controlled trials (with narrow confidence interval)
1b-	Individual randomized controlled trials (with wide confidence interval)
1c	All or none randomized controlled trials
2a	Systematic reviews (with homogeneity) of cohort studies
2a-	Systematic reviews of cohort studies displaying worrisome heterogeneity
2b	Individual cohort study or low quality randomized controlled trials (<80% follow-up)
2b-	Individual cohort study or low quality randomized controlled trials (<80% follow-up / wide confidence interval)

Level	Description
2c	“Outcome” research; ecological studies
3a	Systematic review (with homogeneity) of case-control studies
3a-	Systematic review of case-control studies with worrisome heterogeneity
3b	Individual case-control study
4	Case-series (and poor quality cohort and case-control studies)
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles”

****

เอกสารอ้างอิง

1. Hulley BS, Cummings RS. Designing Clinical Research an Epidemiologic Approach. Baltimore; Williams & Wilkins: 1988.
2. Fletcher HR, Fletcher WS, Wagner HE. Clinical epidemiology the essentials. 3rd ed. Baltimore; Williams&Wilkins; 1996.

****

แบบฝึกหัด

1. ข้อใดต่อไปนี้จัดเป็นการศึกษาทางระบาดวิทยา

   A. Meta-analysis

   B. Case series

   C. Case report

   D. Expert opinion

   E. Qualitative study

2. หลักฐานทางการแพทย์จากการวิจัยเชิงสังเกตรูปแบบใด ให้ผลการศึกษาน่าเชื่อถือที่สุด

   A. Meta-analysis of RCT

   B. Randomized controlled trial

   C. Cohort study

   D. Case-control study

   E. Cross-sectional study

3. แพทย์ต้องการทำวิจัยเพื่อผลของยาตัวหนึ่งที่ใช้ในการรักษาโรค อยากทราบว่าวิธีการศึกษาใดเหมาะสมที่สุด

   A. Cohort study

   B. Case - control

   C. Randomized control trial

   D. Meta-analysis

   E. Cross - sectional study

4. แพทย์ต้องการเลือกใช้ยาชนิดหนึ่งเพื่อการรักษาโรค หลักฐานจากการศึกษารูปแบบใดน่าเชื่อถือที่สุด

   A. Meta-analysis

   B. Expert opinion

   C. Observational study

   D. Consensus conference

   E. Randomized control trial

1.11.3.1 Confidence Interval

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

การกระจายของค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่ม (Sampling Distribution)

ในการทำวิจัยแต่ละครั้ง เราคาดหวังว่าค่าสถิติ (Statistics) ซึ่งเป็นผลการวิจัยที่ได้จากกลุ่มตัวอย่าง จะเท่ากับหรือใกล้เคียงกับค่าที่เป็นจริงในประชากร (Parameters) ตัวอย่างเช่น การสุ่มสำรวจเมื่อปี พ.ศ.2545 ข้อมูลจากกลุ่มตัวอย่างพบว่าเด็กวัยรุ่นไทยมีระดับ IQ เฉลี่ยเท่ากับ 88 จากการศึกษานี้เราอนุมานว่าเด็กวัยรุ่นไทยทั้งประเทศน่าจะมีระดับ IQ เฉลี่ยเท่ากับ 88 เช่นเดียวกัน อย่างไรก็ตามเนื่องจากธรรมชาติของการสุ่มตัวอย่างซึ่งมีความไม่แน่นอนตามธรรมชาติ (Random variation) ทำให้กลุ่มตัวอย่างที่เราใช้ศึกษาอาจไม่เป็นตัวแทนที่ดีที่สุด และอาจให้ค่าสถิติที่เบี่ยงเป็นไปจากความเป็นจริง จากตัวอย่างข้างต้นหากทำการสำรวจในลักษณะเดียวกันซ้ำหลาย ๆ ครั้ง เราจะได้ค่า IQ เฉลี่ยที่แตกต่างกันไปเสมอ ซึ่งอาจเท่ากับ น้อยกว่า หรือมากกว่า 88 ก็ได้ อย่างไรก็ตามเนื่องจากในชีวิตจริงเราทำการศึกษาเพียงครั้งเดียว และได้ค่าสถิติมาเพียงค่าเดียวเท่านั้น เพื่ออธิบายความไม่แน่นอนของผลการศึกษาดังกล่าววิชาสถิติจึงได้สมมติจินตนาการถึงสถานการณ์การสุ่มตัวอย่างขนาดเท่า ๆ กันจากประชากร โดยทำหลาย ๆ ครั้ง มากเท่าที่จะเป็นไปได้ แล้วพิจารณาการกระจายของค่าสถิติที่ได้จากกลุ่มตัวอย่างแต่ละกลุ่ม เรียกประชากรกลุ่มใหม่นี้ว่าการกระจายของค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่ม (Sampling distribution of means หรือ Population of sample means)

ในทางสถิติ การกระจายของค่าสถิติที่ได้จากการสุ่มตัวอย่างจากประชากรมีคุณสมบัติดังนี้

1. ค่าเฉลี่ยของค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่ม จะเท่ากับค่าจริงในประชากร
2. ค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่มจะมีลักษณะการกระจายแบบโค้งปกติ (Normal distribution) ถ้าประชากรมีการกระจายแบบปกติ หรือขนาดตัวอย่างมากพอ
3. ค่าสถิติที่ใช้ในการวัดการกระจายของค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่มเรียกว่าค่าความผิดพลาดมาตรฐานของค่าสถิติของตัวอย่าง (Standard Error)

จากตัวอย่างข้างต้น อธิบายได้ว่า IQ เฉลี่ยเท่ากับ 88 ซึ่งเป็นค่าสถิติที่ได้จากการศึกษาครั้งนี้ เป็นเพียงหนึ่งในหลาย ๆ ความเป็นไปได้ของค่าสถิติของกลุ่มตัวอย่างที่ถูกสุ่มมาจากประชากรจริง ซึ่งประชากรใหม่ที่ประกอบด้วยค่าสถิติของแต่ละกลุ่มตัวอย่างที่เรียกว่า Sampling distribution of means นี้ จะมีการกระจายตัวแบบโค้งปกติ

ค่าผิดพลาดมาตรฐานของค่าสถิติของตัวอย่าง (Standard errors)

Standard error เป็นค่าที่ใช้บอกลักษณะการกระจายของค่าสถิติของกลุ่มตัวอย่าง ค่า Standard error ยิ่งมากหมายถึงมีการกระจายสูง ค่า Standard error น้อยหมายถึงมีกระจายของค่าสถิติต่ำ เนื่องจากในชีวิตจริงเราทำการวิจัยเพียงครั้งเดียว ได้ค่าสถิติเพียงค่าเดียว Standard error จึงเป็นเครื่องบอกถึงโอกาสผิดพลาดของผลการศึกษาเนื่องจากความไม่แน่นอนของการสุ่มตัวอย่างว่ามีมากน้อยเพียงใดนั่นเอง ขนาดของ Standard error ขึ้นกับขนาดของกลุ่มตัวอย่างและเป็นเครื่องมือสำคัญในการคำนวณช่วงความเชื่อมั่น (Confidence interval) ซึ่งจะกล่าวถึงต่อไป

Standard error (se) สามารถคำนวณได้โดยสูตร

$$se = \frac{\sigma}{\sqrt{n}}$$

se	=	ค่าผิดพลาดมาตรฐานของค่าเฉลี่ย
σ	=	ค่าเบี่ยงเบนมาตรฐานของประชากร
n	=	ขนาดตัวอย่าง

จากตัวอย่างข้างต้น สมมติขนาดของกลุ่มตัวอย่างเท่ากับ 458 และประชากรมีค่าเบี่ยงเบนมาตรฐานเท่ากับ 2.5 เราสามารถคำนวณ Standard error ของ IQ เฉลี่ยของเด็กวัยรุ่นจากการศึกษาโดยสูตรคำนวนดังกล่าวดังนี้

$$se = \frac{2.5}{\sqrt{458}}$$

$$= 0.12$$

กรณีที่ค่าสถิติเป็นสัดส่วน ค่า Standard error สามารถคำนวณได้โดยใช้สูตร

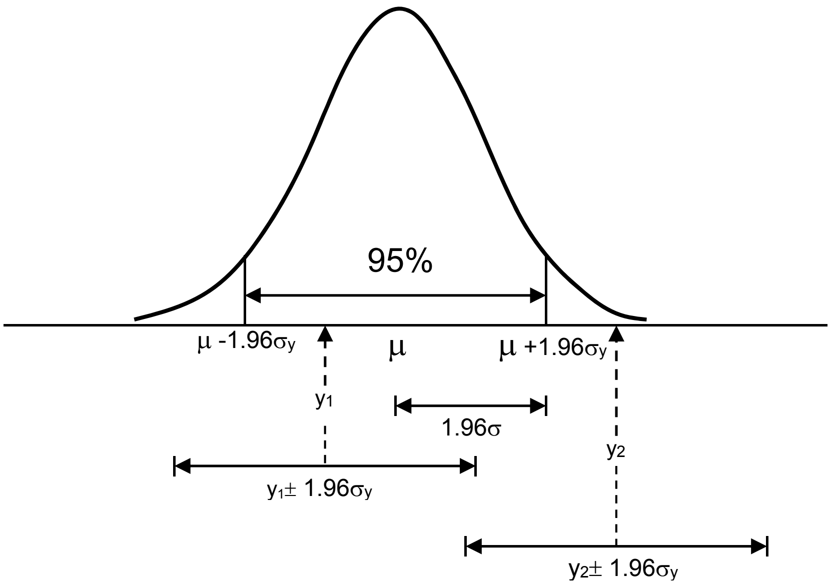
$$se = \sqrt{\frac{p * q}{n}}$$

‘ se	=	ค่าผิดพลาดมาตรฐาของสัดส่วน
p	=	ค่าสัดส่วน
q	=	(1 - p)
n	=	จำนวนตัวอย่าง

ช่วงความเชื่อมั่น (Confidence Intervals, CI)

ดังที่กล่าวในตอนต้นว่าค่าสถิติที่ได้จากกลุ่มตัวอย่างที่สุ่มมาเพื่อทำการวิจัยเป็นเพียงค่าเดียวที่เรามีอยู่ และสามารถใช้ประมาณค่าจริงในประชากรได้ดีที่สุด และค่าสถิติของกลุ่มตัวอย่างที่ได้จากการสุ่มจะมีลักษณะการกระจายแบบโค้งปกติ จากคุณสมบัติดังกล่าวทำให้เราสามารถกำหนดระดับของความแน่นอนเที่ยงตรงของผลการศึกษาได้ ในลักษณะของการกำหนดช่วงความเชื่อมั่น ซึ่งมีค่าสถิติที่ได้จากการวิจัยเป็นค่ากลางแล้วบวกลบด้วย Standard error เพื่อให้ได้ขอบเขตด้านสูง และด้านต่ำของช่วงความเชื่อมั่นตามลำดับ

แผนภูมิประกอบเพื่อความเข้าใจเกี่ยวกับ Confidence interval



จากแผนภูมิซึ่งแสดงการกระจายแบบปกติของความเป็นไปได้ทั้งหมดของค่าสถิติของกลุ่มตัวอย่าง (Sampling distribution of means) เมื่อพิจารณาตัวอย่างการวิจัยเรื่อง IQ ในเด็กวัยรุ่นไทย เนื่องจากเราทำวิจัยเพียงครั้งเดียว เราจึงไม่อาจทราบได้ว่า ผลการศึกษาซึ่งได้ค่า IQ เฉลี่ย = 88 นั้น อยู่ ณ ตำแหน่งใดบนความเป็นไปได้ทั้งหมดของค่าเฉลี่ยของกลุ่มตัวอย่าง

กรณี y_1 ที่ค่าเฉลี่ยของผลการศึกษาห่างจากค่าเฉลี่ยของประชากร μ น้อยกว่า 1.96se ช่วงค่า $y_1 \pm$ 1.96se จะมีค่า μ รวมอยู่ด้วย

กรณี y_2 ที่ค่าเฉลี่ยของผลการศึกษาห่างจากค่าเฉลี่ยของประชากร μ มากกว่าหรือน้อยกว่า 1.96se ช่วงค่า $y_2 \pm$ 1.96se จะไม่มีค่า μ รวมอยู่ด้วย

เนื่องจากค่าเฉลี่ยของกลุ่มตัวอย่างที่เป็นไปได้ทั้งหมดมีการกระจายแบบปกติ เพราะฉะนั้นความน่าจะเป็นที่ค่าเฉลี่ยของผลการศึกษาในที่นี้คือ IQ = 88 จะเป็นเหมือนกรณี y_1 คืออยู่ห่างจากค่าเฉลี่ยของประชากร μ ไม่เกิน $\pm$ 1.96se จึงเท่ากับร้อยละ 95 ทำให้เรามีความมั่นใจร้อยละ 95 ว่าค่าเฉลี่ยของ IQ ในประชากรจริงจะอยู่ในช่วงค่าเฉลี่ยของผลการศึกษา $\pm$ 1.96se อีกนัยหนึ่งเราสามารถประมาณค่า 95% CI ของทุกการศึกษาได้ตั้งแต่ก่อนจะได้ผลการศึกษาว่า จะเท่ากับ ผลการศึกษา $\pm$ 1.96se เสมอ

ระดับความเชื่อมั่นจะกำหนดเท่าใดก็ได้ แต่โดยทั่วไปทางการแพทย์กำหนดไว้ที่ร้อยละ 95 หมายถึง ผู้ศึกษามีความมั่นใจร้อยละ 95 ว่าค่าจริงในประชากรจะอยู่ในช่วงความเชื่อมั่นดังกล่าว หรือมีโอกาสความเป็นไปได้ร้อยละ 5 ที่ช่วงความเชื่อมั่นดังกล่าวจะไม่ครอบคลุมค่าจริงในประชากร และเมื่อพิจารณาพื้นที่ภายใต้โค้งปกติ จะพบว่าพื้นที่ร้อยละ 95 จะครอบคลุมค่าสถิติตั้งแต่ -1.96se ของค่าเฉลี่ยจนถึง 1.96se ของค่าเฉลี่ย ค่า 1.96 นี้เรียกว่า Confidence coefficient ซึ่งจะแปรเปลี่ยนไปตามระดับความเชื่อมั่นที่ต้องการ โดยจะสูงขึ้นหากต้องการระดับความเชื่อมั่นมากขึ้น และต่ำลงหากต้องการระดับความเชื่อมั่นน้อยลดลง เช่นที่ 99%CI และ 90%CI ค่า Confidence coefficient จะเท่ากับ 3.25 และ 1.64 ตามลำดับ ค่า CI จะกว้างขึ้นหากขนาดกลุ่มตัวอย่างลดลงและกำหนดระดับความเชื่อมั่นมากขึ้น ในทางตรงข้าม ค่า CI จะแคบลงหากขนาดกลุ่มตัวอย่างมากขึ้นและกำหนดระดับความเชื่อมั่นน้อยลง

สูตรคำนวณ 95%CI

CI = point estimate ± (confidence coefficient X standard error)

โดย Point estimate หมายถึงค่าสถิติของผลการศึกษา

จากตัวอย่างข้างต้นเกี่ยวกับการสำรวจIQ ในเด็กวัยรุ่นไทย เราสามารถคำนวน 95% Confidence interval (95% CI) ได้ดังนี้ คือ

95%CI = 88 – (0.12 x 1.96) ถึง 88 + (0.12 x 1.96)

= 88 – 0.24 ถึง 88 + 0.24

= 87.76 ถึง 88.24

ในวารสารทางแพทย์นิยมเขียนในรูปแบบดังนี้ คือ

IQ = 88, 95%CI 87.7 – 88.3

หมายถึง ผลการศึกษาค่าเฉลี่ย IQ ของเด็กวัยรุ่นไทยมีค่าเท่ากับ 88 โดยมีความมั่นใจ (โอกาสความน่าจะเป็น) ร้อยละ 95 ว่า IQ เฉลี่ยจริงของเด็กวัยรุ่นไทยทั้งประเทศ จะมีค่าอยู่ระหว่าง 87.7 – 88.3

การประยุกต์ใช้ค่าช่วงความเชื่อมั่น

เราสามารถให้ค่าCI ประมาณค่าจริงในประชากรทั้งในการศึกษาเชิงพรรณนาและการศึกษาเชิงวิเคราะห์ กรณีของการศึกษาเชิงวิเคราะห์ แม้ค่า CI จะได้มาจากการประมาณการ (Estimation) ส่วน ค่า p ได้มาจากการทดสอบสมมติฐาน (Hypothesis testing) แต่ค่าสถิติทั้งสองจะมีความสัมพันธ์กันอย่างใกล้ชิดเสมอ โดยหากค่า p น้อยกว่า 0.05 (มีนัยสำคัญทางสถิติ) ช่วงค่า 95%CI ที่คำนวณได้จะไม่มีค่าที่กำหนดไว้ใน Ho (Null value) รวมอยู่ด้วย เช่นในการศึกษาแบบ RCT ค่า CI ของ Relative risk ที่ไม่พาดผ่าน 1 หมายถึงการรักษานั้นมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ หรือในการศึกษาแบบ Case-control study ค่า Odds ratio ที่ไม่พาดผ่าน 1 หมายถึงปัจจัยเสี่ยงนั้นมีความสัมพันธ์กับการเกิดโรคอย่างมีนัยสำคัญทางสถิติ

ตารางแสดงตัวอย่างการแปลผลค่า CI ในสถานการณ์ต่าง ๆ

รูปแบบการศึกษา	ค่าสถิติ + CI	Null value	การแปลผล
ส่วนสูงเฉลี่ยของเด็กหญิงอายุ 10 ปี ที่อาศัยในชนบทน้อยกว่าส่วนสูงของเด็กหญิงอายุ 10 ปี ที่อาศัยในเมือง	Mean different = 1.8 cm, 95%CI 0.6 – 3.0 cm.	0	• ผู้วิจัยมีความมั่นใจร้อยละ 95 ว่า ส่วนสูงของเด็กหญิงอายุ 10 ปี ที่อาศัยในเมืองมากกว่าส่วนสูงของเด็กหญิงที่อาศัยในชนบทระหว่าง 0.6 – 3.0 cm. ซึ่งช่วงความเชื่อมั่นไม่รวม null value (0)

รูปแบบการศึกษา	ค่าสถิติ + CI	Null value	การแปลผล
			• สรุปว่า เด็กหญิงอายุ 10 ปี ที่อาศัยในชนบทเตี้ยกว่าเด็กหญิงอายุ 10 ปี ที่อาศัยในเมืองอย่างมีนัยสำคัญทางสถิติ
การสำรวจในโรงงานทำเฟอร์นิเจอร์พบว่าผู้ที่ทำงานในแผนกเลื่อยไม้มีภาวะหูเสื่อมมากกว่าผู้ที่ทำงานแผนกทาสี	OR = 2.1, 95%CI 1.8 – 2.4	1	• ผู้วิจัยมีความมั่นใจร้อยละ 95 ว่า ผู้ที่ทำงานในแผนกเลื่อยไม้มีโอกาสหูเสื่อมมากกว่าผู้ที่ทำงานแผนกทาสีระหว่าง 1.8 – 2.4 เท่า ซึ่งช่วงความเชื่อมั่นไม่รวม null value (1) • สรุป ผู้ทำงานในแผนกเลื่อยไม้มีโอกาสเกิดภาวะหูเสื่อมมากกว่าผู้ที่ทำงานในแผนกทาสีอย่างมีนัยสำคัญทางสถิติ
อุบัติการณ์ป่วยเป็นไข้หวัดใหญ่ในผู้ได้รับวัคซีนจริงต่ำกว่าผู้ได้รับวัคซีนหลอก	RR = 0.8, 95%CI 0.7 – 1.1	1	• ผู้วิจัยมีความมั่นใจร้อยละ 95 ว่า อุบัติการณ์ของการป่วยเป็นไข้หวัดใหญ่ในผู้ได้รับวัคซีนจริงเทียบกับผู้ได้รับวัคซีนหลอกอยู่ระหว่าง 0.7 – 1.1 เท่า ซึ่งช่วงความเชื่อมั่นรวม null value (1) • สรุป ผู้ได้รับวัคซีนจริงป่วยเป็นไข้หวัดใหญ่ไม่แตกต่างจากผู้ไม่ได้รับวัคซีน

ค่า CI ให้ข้อมูลที่เป็นประโยชน์ครอบคลุมกว่าค่า p เพราะในขณะที่ค่า p บอกเพียงว่า ผลการศึกษามีนัยสำคัญทางสถิติหรือไม่ ค่า CI นอกจากจะบอกว่ามีนัยสำคัญทางสถิติแล้ว ยังอธิบายถึงความไม่แน่นอนของผลการศึกษาอีกด้วยว่ามีมากน้อยเพียงใด ฉะนั้นในทางปฏิบัติ การเขียนรายงานควรระบุค่า p ร่วมกับค่า CI หรือถ้าจะเสนอเพียงอย่างค่าเดียว ให้เลือกเสนอค่า CI

****

เอกสารอ้างอิง

1. Altman GD. Practical statistics for medical research. London: Chapman & Hall; 1991.
2. Knapp GR, Miller MC. Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins; 1992.

****

แบบฝึกหัด

1. การศึกษาเชิงทดลองเพื่อหาประสิทธิผลของวัคซีนชนิดหนึ่งเทียบกับ Placebo ได้ค่า RR = 0.54, p = 0.03 ข้อใดเป็นConfidence interval ที่เป็นไปได้

   A. 0.62 – 0.98
   B. 1.02 – 1.36
   C. 0.10 – 0.98
   D. 0.54 – 0.98
   E. 0.10 – 0.54

2. การแสดงผลการศึกษาด้วยค่า Confidence interval ดีกว่า p value อย่างไร

   A. ถูกต้องกว่า
   B. แม่นยำกว่า
   C. อคติน้อยกว่า
   D. เข้าใจง่ายกว่า
   E. ให้ข้อมูลมากกว่า

3. การศึกษาในโรงงานทำเฟอร์นิเจอร์ พบว่าผู้ที่ทำงานในแผนกเลื่อยไม้มีภาวะหูเสื่อมมากกว่าผู้ที่ทำงานแผนกทาสีอย่างมีนัยสำคัญทางสถิติ OR = 2.1, %95CI 1.8 – 2.4 ข้อใดต่อไปนี้อธิบายผลการศึกษาข้างต้นได้ถูกต้อง

   A. ผู้วิจัยมั่นใจ %95ว่า OR จริงในประชากร จะมีค่าระหว่าง 1.8 – 2.4
   B. ผู้วิจัยมั่นใจ %95ว่า OR ของการศึกษา จะมีค่าระหว่าง 1.8 – 2.4
   C. ผู้วิจัยมั่นใจ %95ว่า OR จริงในประชากร 2.1 =
   D. %95ของ OR จริงในประชากร จะมีค่าระหว่าง 1.8 – 2.4
   E. %95ของ OR จริงในประชากร มีค่า 2.1 =

1.11.3.2 Statistical significance and Type I error

1.11.3.3 Statistical power and Type II error

ผศ.นพ.เกรียงไกร ศรีธนวิบุญชัย

การทดสอบสมมติฐาน (Hypothesis testing)

ในการศึกษาเชิงวิเคราะห์ ผู้วิจัยจะคาดการณ์ลักษณะของความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ไว้ล่วงหน้า เรียกว่าการตั้งสมมติฐานการวิจัย (Research hypothesis) ตัวอย่างเช่น แพทย์ต้องการเปรียบเทียบผลการรักษาโรคด้วยยา A ซึ่งเป็นยาชนิดใหม่ กับยา B ซึ่งเป็นยามาตรฐานที่ใช้รักษาโรคนั้นอยู่เดิม จึงตั้งสมมติฐานไว้ตั้งแต่ก่อนเริ่มทำการศึกษาว่า ยา A มีประสิทธิผลในการรักษาโรคดีกว่ายา B ผลการศึกษาที่จะออกมาอาจสอดคล้องกับสมมติฐานนี้หรือไม่ก็ได้ขึ้นกับผลการรักษาของยาทั้งสองชนิดที่เกิดกับกลุ่มตัวอย่างที่เลือกมาทำการวิจัย กระบวนการตรวจสอบทางสถิติว่าสมมติฐานดังกล่าวมีโอกาสเป็นจริงมากน้อยเพียงใด เรียกว่าการทดสอบสมมติฐาน

ก่อนจะทำการทดสอบสมมติฐาน ต้องมีการแปลงสมมติฐานการวิจัยให้เป็นสมมติฐานทางสถิติ (Statistical hypothesis) ก่อน เพื่อประกอบการทำความเข้าใจผลการศึกษา ซึ่งมี 2 ลักษณะ คือ

1. <u>Null hypothesis (Ho)</u> คือ สมมติฐานตั้งต้นที่ตั้งขึ้นเพื่อหาหลักฐานมาหักล้าง มักตั้งขึ้นโดยแสดงความไม่แตกต่างกันของคุณสมบัติของประชากร จากตัวอย่างข้างต้น อาจตั้ง Ho ว่า

   “ยา A ให้ผลการรักษาโรคไม่แตกต่างจากยา B”

   ผู้วิจัยจะใช้ผลการศึกษามาพิจารณาว่า โอกาสความน่าจะเป็นที่จะเกิดความแตกต่างกันของผลการรักษาดังกล่าวน้อยเกินกว่าที่จะเกิดจากความบังเอิญเพียงอย่างเดียวหรือไม่ ซึ่งจะนำไปสู่การปฏิเสธ Ho หรือไม่

2. <u>Alternative hypothesis (Ha)</u> คือสมมติฐานที่แสดงความแตกต่างกันของคุณสมบัติของประชากร ตั้งขึ้นเพื่อรองรับกรณีที่เราปฏิเสธ Ho นั่นคือถ้าปฏิเสธ Ho จะมายอมรับ Ha แทน จากตัวอย่างข้างต้นอาจตั้งว่า

   “ยา A ให้ผลการรักษาโรคแตกต่างจากยา B”

   หากโอกาสความน่าจะเป็นที่จะเกิดความแตกต่างกันของผลการรักษาดังกล่าวน้อยกว่าที่จะเกิดจากความบังเอิญแต่เพียงอย่างเดียว ผู้วิจัยจะปฏิเสธ Ho และมายอมรับ Ha การตั้ง Ha ทำได้ 2 ลักษณะ คือ แบบ 2-tailed หมายถึงการตั้งสมมติฐานโดยไม่กำหนดทิศทางของความสัมพันธ์ และแบบ 1-tailed หมายถึงการตั้งสมมติฐานที่มีการกำหนดทิศทางของความสัมพันธ์ โดยทั่วไปการตั้งสมมติฐานแบบ 2-tailed จะถือว่าอนุรักษ์นิยม (Conservative) กว่า ได้รับการยอมรับมากกว่า และมีโอกาสปฏิเสธ Ho น้อยกว่าแบบ 1-tailed

   2-tailed “ยา A ให้ผลการรักษาโรค<u>แตกต่าง</u>จากยา B”

   1-tailed “ยา A ให้ผลการรักษา<u>ดีกว่า</u>ยา B”

p-value

ข้อมูลที่เป็นผลการวิจัยจะถูกนำมาคำนวณ โดยใช้การทดสอบทางสถิติ (Test statistic) ที่เหมาะสม เพื่อหาความน่าจะเป็นที่เราจะตรวจพบความแตกต่างดังกล่าวหาก Ho เป็นจริง เรียกความน่าจะเป็นนี้ว่าค่า p ค่า p ยิ่งน้อย โอกาสที่จะตรวจพบความแตกต่างดังกล่าวเนื่องจากความบังเอิญเพียงอย่างเดียวก็จะยิ่งน้อยไปด้วย

ความน่าจะเป็นที่น้อยที่สุด ที่ใช้เป็นตัวกำหนดว่าผู้วิจัยจะปฏิเสธ Ho และยอมรับ Ha เรียกว่าระดับนัยสำคัญทางสถิติ (α) ค่านี้ไม่มีข้อกำหนดตายตัวว่าต้องเป็นเท่าใด ขึ้นกับผู้วิจัยว่าจะยอมรับโอกาสการสรุปผลผิดพลาดแบบผลบวกลวงได้มากน้อยแค่ไหน อย่างไรก็ตามการวิจัยทางการแพทย์โดยทั่วไปจะกำหนดไว้ที่ร้อยละ 5 หรือความน่าจะเป็น 0.05 หากค่า p ที่คำนวณได้ น้อยกว่า 0.05 หมายถึงมีหลักฐานเพียงพอที่จะปฏิเสธ Ho ซึ่งจะเป็นการยอมรับ Ha โดยปริยาย และหากค่า p มากกว่าหรือเท่ากับ 0.05 หมายถึงไม่มีหลักฐานเพียงพอที่จะปฏิเสธ Ho

จากตัวอย่างข้างต้น สมมติว่าผลการทดสอบทางสถิติ ได้ค่า p = 0.02 เราจะสรุปผลการศึกษาว่ามีหลักฐานเพียงพอที่จะปฏิเสธ Ho และยอมรับ Ha คือ ยา A ให้ผลการรักษาโรคแตกต่างจากยา B อย่างมีนัยสำคัญทางสถิติ เนื่องจาก ค่า p ที่คำนวณได้น้อยกว่า 0.05 หรืออีกนัยหนึ่ง ความแตกต่างของผลการรักษาระหว่างยาทั้งสองตัวที่พบจากการศึกษานี้ มีโอกาสน้อยเกินกว่าที่จะเกิดจากความบังเอิญแต่เพียงอย่างเดียว

ข้อผิดพลาดในการสรุปผลการศึกษา

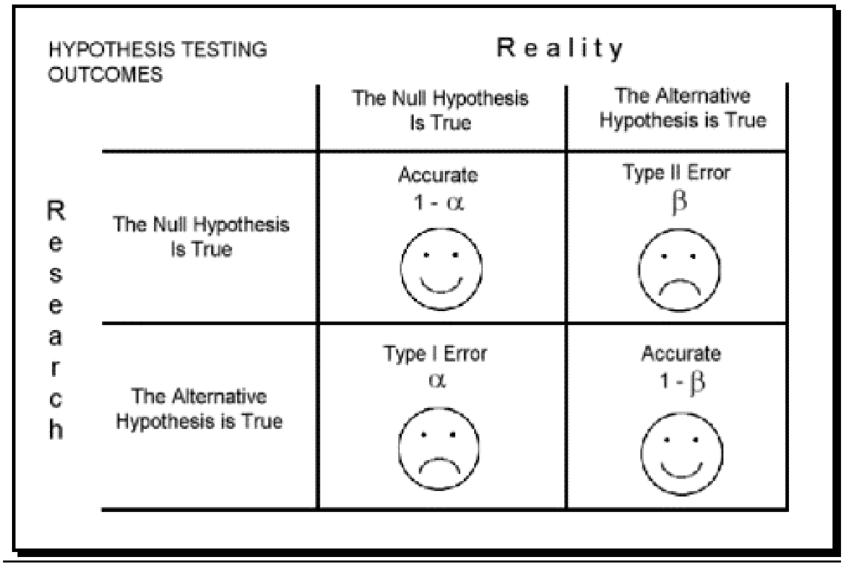
การใช้ค่าความน่าจะเป็นที่ได้จากการทดสอบทางสถิติหรือ ค่า P ในการตัดสินสรุปผลการศึกษา อาจนำไปสู่ข้อผิดพลาดที่อาจเกิดขึ้นได้ใน 2 ลักษณะ คือ

1. ***Type I error*** หรืออีกชื่อหนึ่ง α error คือ ผลบวกลวง (False positive) หมายถึง การสรุปผลการศึกษาผิดพลาด โดยปฏิเสธ Ho ทั้งที่ Ho เป็นจริง จากตัวอย่างข้างต้นคือในความเป็นจริง ยา A และ ยา B ให้ผลการรักษาในประชากรไม่ต่างกัน แต่ผู้วิจัยสรุปผลการศึกษาผิดเป็นว่ายาทั้งสองให้ผลการรักษาต่างกัน ความน่าจะเป็นของความผิดพลาดแบบนี้จะเท่ากับค่า p ที่คำนวณได้

2. ***Type II error*** หรืออีกชื่อหนึ่ง β error คือ ผลลบลวง (False negative) หมายถึงการสรุปผลการศึกษาผิดพลาด โดยยอบรับ Ho ทั้งที่ Ho ไม่เป็นจริง จากตัวอย่างข้างต้น คือในความเป็นจริง ยา A และ ยา B ให้ผลการรักษาแตกต่างกัน แต่ผู้วิจัยสรุปว่ายาทั้งสองให้ผลการรักษาไม่ต่างกัน ความน่าจะเป็นของความผิดพลาดแบบนี้ขึ้นกับความแตกต่างของผลการรักษา (Effect size) ว่ามีมากเท่าใดและขนาดของกลุ่มตัวอย่าง ค่าความผิดพลาดแบบ Type II error ไม่มีค่าตายตัว แต่ส่วนใหญ่กำหนดไว้ที่ 0.2 หรือร้อยละ 20 จะเห็นว่า โดยทั่วไปผู้วิจัยยอมรับความผิดพลาดแบบ Type II error มากกว่าแบบ Type I error คือจะไม่รายงานว่าพบความสัมพันธ์ถ้าไม่แน่ใจจริง ๆ ส่วนจะตรวจไม่พบความสัมพันธ์บ้างก็ยังไม่เป็นไร

   ในทางปฏิบัติ เรามักพิจารณาโอกาสที่จะตรวจพบความแตกต่างตาม Effect size ที่ต้องการ เรียกอีกอย่างหนึ่งว่าอำนาจของการทดสอบทางสถิติ (Statistical power) ซึ่งมีค่าเท่ากับ 1 - β ในกรณีที่กำหนด β error ไว้ที่ 0.2 Statistical power จะเท่ากับ 0.8 โดยทั่วไปเราสามารถกำหนดค่า β error และ Statistical power ล่วงหน้าได้โดยการกำหนดขนาดของกลุ่มตัวอย่าง

แผนภูมิแสดงความถูกต้องของข้อสรุปผลการวิจัยในลักษณะต่าง ๆ



Type I และ Type II error เป็นเพียงโอกาสในทางทฤษฎีที่จะเกิดความผิดพลาดในการสรุปผลการวิจัย แต่ในความเป็นจริงเราทราบเพียงว่าเนื่องจากความคลาดเคลื่อนตามธรรมชาติของการคัดเลือกกลุ่มตัวอย่าง โดยเฉลี่ยแล้วหากกำหนดค่า α error ไว้ที่ร้อยละ 5 ในการทำการศึกษา 20 ครั้ง จะมีการสรุปผลการศึกษาผิดพลาดแบบผลบวกลวง (Type I error) 1 ครั้ง เช่นเดียวกันหากกำหนด β error ไว้ที่ร้อยละ 20 ในการทำการศึกษาแบบเดียวกัน 5 ครั้ง จะมีการสรุปผลการศึกษาผิดพลาดแบบผลลบลวง (Type II error) 1 ครั้ง เพียงแต่เราไม่สามารถบอกได้ว่าการวิจัยครั้งที่เราทำนั้นจะเป็นครั้งที่มีการสรุปผลการศึกษาผิดพลาดหรือไม่ เนื่องจากเลือกกลุ่มอาสาสมัครจากประชากรมาทำการศึกษาเป็นปรากฏการณ์สุ่ม และไม่มีข้อมูลจริงในระดับประชากรมาเปรียบเทียบ

ข้อพิจารณาเกี่ยวกับค่า P values

ค่า p ที่น้อยมาก ๆ บอกถึงความเป็นไปได้ที่น้อยมากที่ผลการศึกษาที่พบจะเกิดเนื่องจากความบังเอิญแต่เพียงอย่างเดียวหาก Ho เป็นจริง อย่างไรก็ตามไม่ได้บอกขนาดของความสัมพันธ์ว่ามากเสมอไป ตัวอย่างเช่นการทดสอบสมมติฐานเพื่อหาความแตกต่างระหว่างยาใหม่กับยามาตรฐานเดิมในการรักษาโรคเบาหวาน สมมติได้ค่า $p < 0.0001$ ก็ไม่ได้หมายความว่ายาใหม่จะลดระดับน้ำตาลในเลือดลงได้มากกว่ายามาตรฐานเดิมหลาย mg.% ในกรณีนี้ ค่า $p < 0.0001$ บอกเพียงว่า โอกาสของการพบความแตกต่างดังกล่าว (จะมากหรือน้อยขึ้นกับขนาดของกลุ่มตัวอย่าง) มีโอกาสเกิดเนื่องจากความบังเอิญแต่เพียงอย่างเดียวน้อยมากเท่านั้น

อีกประการหนึ่ง การพิจารณาผลการวิจัยทางการแพทย์เพียงว่ามีนัยสำคัญทางสถิติ ($p < 0.05$) หรือไม่มีนัยสำคัญทางสถิติ (p > หรือเท่ากับ = 0.05) แล้วนำไปใช้ประกอบการตัดสินใจทางคลินิกเลยเป็นเรื่องอันตราย เนื่องจากผลการวิจัยที่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ มีแนวโน้มจะได้รับการตีพิมพ์ลงวารสารวิชาการต่าง ๆ มากกว่าการวิจัยที่ไม่พบความแตกต่าง (Publication bias) จึงอาจเป็นไปได้ว่า มีงานวิจัยเรื่องเดียวกันอีกเป็นจำนวนมากที่ไม่พบความสัมพันธ์และไม่ได้รับการตีพิมพ์ ผลการวิจัยที่มีนัยสำคัญ

ทางสถิติและได้รับการตีพิมพ์ ดังกล่าว จึงมีโอกาสเป็นผลการวิจัยที่เกิดเนื่องจากความบังเอิญ ไม่ใช่ความจริงมากขึ้น

ข้อควรระวังอีกประการหนึ่งเกี่ยวกับการแปลผลค่า p คือ สำหรับผลการศึกษาที่ไม่มีนัยสำคัญทางสถิติ อาจมีความสำคัญทางคลินิกก็ได้ Freiman และคณะ ทำการวิเคราะห์รายงานการวิจัยที่ไม่มีนัยสำคัญทางสถิติ จำนวน 71 เรื่อง ที่มีค่า $p > 0.1$ โดยการแปลงค่า P เป็นค่า Confidence interval พบว่า จำนวนเกือบครึ่งหนึ่งของผลการศึกษาดังกล่าวมีความแตกต่างที่แสดงถึงผลการรักษาที่ดีขึ้นถึงร้อยละ 50 ที่การศึกษาเหล่านี้มี ค่า p สูง จนไม่สามารถตรวจพบความแตกต่างทางสถิติได้ เนื่องจากมีขนาดตัวอย่างของการศึกษาน้อยเกินไป และในทางตรงกันข้าม ผลการศึกษาที่มีนัยสำคัญทางสถิติ คือ ค่า p ต่ำ ๆ เท่าที่ต้องการ สามารถทำให้เกิดขึ้นได้เสมอ หากขนาดของกลุ่มตัวอย่างมากพอ เนื่องจากการเพิ่ม statistical power จะทำให้สามารถตรวจพบความแตกต่างที่มีขนาดเล็กมากและไม่มีความสำคัญทางคลินิก นี่เป็นอีกเหตุผลหนึ่ง ที่ปัจจุบันเรานิยมแสดงผลการวิเคราะห์ทางสถิติด้วยค่า Confidence interval มากกว่าค่า p

****

เอกสารอ้างอิง

1. Altman GD. Practical statistics for medical research. London: Chapman & Hall; 1991.
2. Knapp GR, Miller MC. Clinical epidemiology and biostatistics. Baltimore: Williams & Wilkins; 1992.

****

แบบฝึกหัด

1. การศึกษาผู้ป่วยโรคแผลในกระเพาะอาหารที่ได้รับการรักษาจนอาการหายดีแล้วจำนวนหนึ่ง พบว่าผู้ป่วยที่มีภาวะเครียด มีโอกาสเกิดอาการกำเริบของโรคซ้ำ มากกว่าผู้ป่วยที่ไม่มีภาวะเครียด 3.67 เท่า อย่างมีนัยสำคัญทางสถิติ ข้อใดต่อไปนี้สอดคล้องกับข้อความข้างต้นมากที่สุด?

A. 95%CI = 0.98 – 8.46, p = 0.068

B. 95%CI = 1.78 – 6.44, p = 0.033

C. 95%CI = 0.98 – 8.46, p < 0.05

D. 95%CI = 1.78 – 6.44, p > 0.05

E. 95%CI = 0.98 – 8.46, p > 0.05

2. การศึกษาเพื่อหาความสัมพันธ์ระหว่างการดื่มไวน์กับการป้องกันโรคหัวใจและหลอดเลือด ผู้วิจัยกำหนด Type I error ไว้ที่ร้อยละ 5 ธและกำหนด Type II error ไว้ที่ร้อยละ 20 หากผลการวิจัยไม่พบความแตกต่าง การวิจัยนี้มีโอกาสสรุปผลคลาดเคลื่อนจากความเป็นจริงร้อยละเท่าใด

A. 25

B. 20

C. 15

D. 10

E. 5

3. ผลการทดสอบทางสถิติระหว่างผลการรักษาด้วยยาใหม่เทียบกับยาเก่า ได้ p = 0.04 ข้อใดสรุปได้ถูกต้อง

A. ถ้ายาใหม่ไม่ได้ดีกว่ายาเก่า โอกาสที่จะพบว่ายาใหม่ดีกว่ายาเก่าด้วยความบังเอิญ มีร้อยละ 4

B. ถ้ายาใหม่ดีกว่ายาเก่า โอกาสที่จะพบว่ายาใหม่ดีกว่ายาเก่าด้วยความบังเอิญ มีร้อยละ 4

C. ยาใหม่ดีกว่ายาเก่า การศึกษาครั้งนี้พบความแตกต่าง แต่ไม่มีนัยสำคัญทางสถิติ

D. ถ้าจะพิสูจน์ให้ได้ว่าการรักษาแบบใหม่ ดีกว่าการรักษาแบบเก่า จะเพิ่มจำนวนอาสาสมัครให้มากกว่านี้

E. การศึกษานี้ไม่มีอำนาจเพียงพอที่จะตรวจสอบว่าการรักษาด้วยวิธีใหม่นี้ดีกว่าวิธีเดิม

4. จากผลการศึกษาพบว่า ชายที่มีการขริบหนังหุ้มปลายอวัยวะเพศ มีอัตราการติดเชื้อ HIV 0.66 ราย/100 คน/ปี ในขณะที่ชายที่ไม่ได้ขริบหนังหุ้มปลายอวัยวะเพศ มีอัตราการติดเชื้อ HIV 1.33 ราย/100 คน/ปี p = 0.002 ค่าสถิติใดสอดคล้องกับข้อมูลข้างต้น

A. Type I error = 0.02

B. Type II error = 0.002

C. Alpha = 0.002

D. Beta = 0.002

E. Delta = 0.002