

การคำนวณขนาดตัวอย่างสำหรับการศึกษาวิจัยเกี่ยวกับการตรวจวินิจฉัย

นฤมล สุธาใจ

ภาควิชาศัลยศาสตร์ออร์โธปิดิกส์และกายภาพบำบัด คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

Received: November 19, 2019

Revised: February 20, 2020

Accepted: March 18, 2020

บทคัดย่อ

วิธีการคำนวณขนาดตัวอย่างสำหรับการศึกษาวิจัยเกี่ยวกับการตรวจวินิจฉัยขึ้นอยู่กับวัตถุประสงค์หลักของงานวิจัย โดยความชุกของโรคเป็นปัจจัยสำคัญในการคำนวณขนาดตัวอย่างที่ต้องการใช้เพื่อการประมาณค่า/การทดสอบความไวและความจำเพาะ กรณีที่ความชุกของโรคเท่ากับ 0.50 ขนาดตัวอย่างที่ต้องการใช้เพื่อการประมาณค่า/การทดสอบความไวและความจำเพาะจะเหมือนกัน ในทางกลับกันถ้าความชุกของโรคไม่เท่ากับ 0.50 การคำนวณขนาดตัวอย่างควรใช้สูตรขนาดตัวอย่างเพื่อการประมาณค่า/การทดสอบความไวและความจำเพาะ บทความนี้ผู้เขียนอธิบายแนวคิดของการตรวจวินิจฉัย การคำนวณขนาดตัวอย่างเพื่อการประมาณค่า/การทดสอบความไวและความจำเพาะ พร้อมทั้งแสดงให้เห็นถึงการนำไปใช้ผ่านตัวอย่างและนำเสนอตารางขนาดตัวอย่างสำเร็จรูปเมื่อกำหนดค่าพารามิเตอร์ที่ต่างกัน

คำสำคัญ: ขนาดตัวอย่าง การตรวจวินิจฉัย ความไว ความจำเพาะ

ผู้สนับสนุนประสานงาน:

นฤมล สุธาใจ

ภาควิชาศัลยศาสตร์ออร์โธปิดิกส์และกายภาพบำบัด

คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

เลขที่ 2 ถนนวังหลัง แขวงศิริราช

เขตบางกอกน้อย กรุงเทพฯ 10700

อีเมล: narumol.sud@mahidol.ac.th

Sample size calculation for diagnostic test studies

Narumol Sudjai

Department of Orthopaedic Surgery, Faculty of Medicine, Siriraj Hospital, Mahidol University

Abstract

The method of sample size calculation for diagnostic test studies is based on the primary objective of the study. The prevalence of disease is an important factor to calculate the required sample size for estimating and testing sensitivity and specificity. In the case of the prevalence of disease, it was equal to 0.50 and the required sample size for estimating and testing sensitivity and specificity will be the same. On the other hand, if the prevalence of disease is not equal to 0.50, the sample size calculation would be use both sample size formulas for estimating and testing sensitivity and specificity. In this article, the author described a concept of diagnostic testing, for sample size calculation to estimate or test sensitivity and specificity, as well as illustrating the use through applied examples and presenting the sample size table for different parameters.

Keywords: sample size, diagnostic test, sensitivity, specificity

Corresponding Author:

Narumol Sudjai

Department of Orthopaedic Surgery,

Faculty of Medicine, Siriraj Hospital,

Mahidol University

2 Wanglang Road, Bangkoknoi,

Bangkok 10700, Thailand

E-mail: narumol.sud@mahidol.ac.th

บทนำ

การศึกษาวิจัยเกี่ยวกับการตรวจวินิจฉัย (diagnostic test) มีวัตถุประสงค์เพื่อหาความแม่นยำ (accuracy) ของการตรวจวินิจฉัยแบบใหม่ ซึ่งการประเมินความแม่นยำของการตรวจวินิจฉัยจะพิจารณาจากค่าความไว (sensitivity) และความจำเพาะ (specificity)¹⁻³ สิ่งหนึ่งที่สำคัญที่ต้องพิจารณาอย่างระมัดระวังคือ ขนาดตัวอย่างที่ใช้ในการศึกษาต้องมีมากเพียงพอจึงจะนำไปสู่การประมาณค่าดังกล่าวข้างต้นที่มีความน่าเชื่อถือ⁴⁻⁵ และเนื่องจากสูตรการคำนวณขนาดตัวอย่างมีหลายสูตร ในการพิจารณาเลือกใช้สูตรใดนั้นขึ้นอยู่กับวัตถุประสงค์ของการวิจัย การวิเคราะห์เชิงสถิติ (statistical analysis) ที่จะใช้และรูปแบบการศึกษา (study design)⁶⁻¹³

ในบทความนี้ผู้นิพนธ์จะกล่าวถึงแนวคิดของการตรวจวินิจฉัย (diagnostic test) การคำนวณขนาดตัวอย่างที่ใช้ในการศึกษาเกี่ยวกับการตรวจวินิจฉัยโดยแบ่งประเภทของการคำนวณขนาดตัวอย่างตามวัตถุประสงค์ของงานวิจัยและนำเสนอตารางขนาดตัวอย่างเมื่อกำหนดค่าพารามิเตอร์ที่ต่างกัน

แนวคิดของการตรวจวินิจฉัย

เครื่องมือในการตรวจวินิจฉัยมีส่วนช่วยแพทย์ในการวินิจฉัยโรคเป็นอย่างมาก ซึ่งจะเป็นประโยชน์ในการวางแผนการรักษาผู้ป่วยต่อไปได้ แต่ก่อนที่จะนำเครื่องมือในการตรวจวินิจฉัยแบบใหม่ไปใช้ต้องมีการประเมินประสิทธิภาพของเครื่องมือว่ามีความแม่นยำ (accuracy) โดยเปรียบเทียบกับวิธีมาตรฐาน (gold standard) หรือการทดสอบอ้างอิง (reference test) ซึ่งเป็นที่ยอมรับ ดังตารางที่ 1

ตารางที่ 1 แสดงผลการทดสอบจากการตรวจวินิจฉัยแบบใหม่เปรียบเทียบกับวิธีการตรวจวินิจฉัยมาตรฐานหรือการทดสอบอ้างอิง

	การตรวจวินิจฉัยมาตรฐาน (หรือการทดสอบอ้างอิง)	
	เป็นโรค (disease)	ไม่เป็นโรค (non disease)
ผลการทดสอบจากการตรวจวินิจฉัยแบบใหม่		
ผลบวก (positive)	ผลบวกจริง (True positive; TP)	ผลบวกเท็จ (False positive; FP)
ผลลบ (negative)	ผลลบเท็จ (False negative; FN)	ผลลบจริง (True negative; TN)

ความแม่นยำของการวินิจฉัย (diagnostic accuracy) หมายถึง ผลการทดสอบจากการตรวจวินิจฉัยแบบใหม่ที่ได้มีค่าถูกต้องหรือใกล้เคียงกับค่าจริงมากที่สุด

การประเมินความแม่นยำของการวินิจฉัย

ความแม่นยำในการตรวจวินิจฉัยจะพิจารณาจากค่าความไวและค่าความจำเพาะดังนี้

1. ความไวเป็นคุณลักษณะของการตรวจวินิจฉัยที่แสดงถึงสัดส่วนของผลการทดสอบที่ให้ผล

เป็นบวกในผู้ป่วยที่เป็นโรค สำหรับวิธีการทดสอบที่มีความไวสูงจะมีโอกาสผิดพลาดน้อยมากในการค้นหาคนที่ป่วยเป็นโรคซึ่งจะเป็นประโยชน์ในการช่วยวินิจฉัยแยกโรคออกไป (rule out) และจะมีประโยชน์มากที่สุด เมื่อผลการทดสอบออกมาเป็นลบ¹⁴ แต่อย่างไรก็ตาม วิธีการทดสอบที่มีความไวสูง โอกาสที่จะมีคนที่ไม่ป่วยเป็นโรคปนเข้ามาได้ด้วยเช่นกัน ดังนั้น วิธีการทดสอบที่มีความไวสูงจึงเหมาะสำหรับการตรวจคัดกรองผู้ป่วย (screening test) โดยค่าความไวสามารถคำนวณได้ดังนี้

$$\text{Sensitivity (Se)} = \frac{TP}{TP + FN}$$

2. ความจำเพาะเป็นคุณลักษณะของการตรวจวินิจฉัยที่แสดงถึงสัดส่วนของผลการทดสอบที่ให้ผลเป็นลบในผู้ที่ไม่ป่วยเป็นโรค สำหรับวิธีการทดสอบที่มีความจำเพาะสูงจะมีโอกาสผิดพลาดน้อยมากในการวินิจฉัยคนที่ปกติว่าป่วยเป็นโรคซึ่งจะเป็น

ประโยชน์ในการยืนยันการเป็นโรค (confirm) หรือ rule in¹⁴ ดังนั้น การวินิจฉัยโรคที่มีผลกระทบร้ายแรง เช่น โรคมะเร็ง โรคเอดส์และโรคทางจิตเวช เป็นต้น จำเป็นอย่างยิ่งที่ต้องใช้วิธีการทดสอบที่มีความจำเพาะสูง โดยค่าความจำเพาะสามารถคำนวณได้ดังนี้

$$\text{Specificity (Sp)} = \frac{TN}{FP + TN}$$

นอกจากนี้ ค่าทำนายโรคและอัตราส่วนของความน่าจะเป็นของผลการทดสอบในผู้ที่เป็นโรคต่อผู้ที่ไม่เป็นโรคก็เป็นค่าที่สำคัญสามารถแสดงถึงประสิทธิภาพของการตรวจวินิจฉัย¹⁵ ที่ควรนำเสนอในรายงานผลการวิจัยด้วยเช่นกัน โดยแต่ละค่ามีรายละเอียดดังนี้

3. ค่าทำนายโรค (predictive values) แสดงถึงผลการทดสอบที่ให้ผลเป็นบวกนั้นผู้ป่วยมีโอกาส

ที่จะเป็นโรคจริงๆ เท่าไรหรือถ้าผลการทดสอบให้ผลเป็นลบผู้ป่วยมีโอกาสที่จะไม่เป็นโรคได้เท่าไร โดยค่าทำนายโรคหรือเรียกว่า post-test probabilities สามารถหาได้ดังนี้

1) ค่าทำนายโรคเมื่อผลการทดสอบเป็นบวก (positive predictive value: PPV) หรือความน่าจะเป็นของการเป็นโรคถ้าผลการทดสอบเป็นบวกคือ

$$\text{Positive predictive value (PPV)} = \frac{TP}{TP + FP}$$

2) ค่าทำนายโรคเมื่อผลการทดสอบเป็นลบ (negative predictive value: NPV) หรือ

ความน่าจะเป็นของการไม่เป็นโรคถ้าผลการทดสอบเป็นลบคือ

$$\text{Negative predictive value (NPV)} = \frac{TN}{TN + FN}$$

4. อัตราส่วนของความน่าจะเป็นของผลการทดสอบในผู้ที่เป็นโรคต่อผู้ที่ไม่เป็นโรค (likelihood ratio: LR) เป็นอีกวิธีหนึ่งที่สามารถแสดงถึงประสิทธิภาพของการตรวจวินิจฉัยแบบใหม่ได้ เช่น เมื่อผลการทดสอบเป็นบวกโอกาสที่ผู้ได้รับการตรวจ

จะเป็นโรคเป็นกี่เท่าของผู้ที่ไม่เป็นโรค เป็นต้น โดยค่า LR สามารถคำนวณได้ดังนี้²

1) อัตราส่วนความน่าจะเป็นที่ผลการทดสอบเป็นบวกในผู้ที่เป็นโรคต่อผู้ที่ไม่เป็นโรค (likelihood ratio for a positive result: LR+) คือ

$$\begin{aligned} LR+ &= \frac{\text{ความน่าจะเป็นที่ผลการทดสอบเป็นบวกในผู้ที่เป็นโรค}}{\text{ความน่าจะเป็นที่ผลการทดสอบเป็นบวกในผู้ที่ไม่เป็นโรค}} \\ &= \frac{\text{sensitivity}}{1 - \text{specificity}} \end{aligned}$$

ถ้า LR+ มีค่าสูงแสดงว่าการตรวจวินิจฉัยแบบใหม่นั้นเป็นการทดสอบที่ดี

2) อัตราส่วนความน่าจะเป็นที่ผลการทดสอบเป็นลบในผู้ที่ เป็นโรคต่อผู้ที่ไม่เป็นโรค (likelihood ratio for a negative result: LR-) คือ

$$LR- = \frac{\text{ความน่าจะเป็นที่ผลการทดสอบเป็นลบในผู้ที่ เป็นโรค}}{\text{ความน่าจะเป็นที่ผลการทดสอบเป็นลบในผู้ที่ไม่เป็นโรค}}$$

$$= \frac{1 - \text{sensitivity}}{\text{specificity}}$$

ถ้า LR- มีค่าต่ำมากแสดงว่าการตรวจวินิจฉัยแบบใหม่นั้นเป็นการทดสอบที่ดี

การประมาณค่าดังที่กล่าวมาทั้งหมดข้างต้น จะไม่มีความน่าเชื่อถือถ้าขนาดตัวอย่างที่ใช้ในการศึกษามีไม่เพียงพอ ดังนั้น การคำนวณขนาดตัวอย่าง จึงเป็นสิ่งสำคัญอย่างยิ่งที่ต้องนำมาพิจารณา⁴⁻⁵ โดยมีรายละเอียดดังนี้

ประเภทของการคำนวณขนาดตัวอย่าง

การคำนวณขนาดตัวอย่างสามารถแบ่งได้เป็น 2 ประเภท ตามวัตถุประสงค์ของการวิจัยคือ

1) การคำนวณขนาดตัวอย่างเพื่อการประมาณค่าพารามิเตอร์ในประชากร (parameter estimation) เช่น ค่าสัดส่วน (π), ความไว (Se) และความจำเพาะ (Sp) เป็นต้น

2) การคำนวณขนาดตัวอย่างเพื่อการทดสอบสมมติฐาน (hypothesis testing) ในสูตรการคำนวณขนาดตัวอย่างจะต้องกำหนดค่าระดับนัยสำคัญ (α) และค่า Power ($1 - \beta$) ซึ่งความหมายของค่า α และ β แสดงดังตารางที่ 2¹⁰⁻¹¹

ตารางที่ 2 ความผิดพลาดในการทดสอบสมมติฐานทางสถิติ

ผลการทดสอบสมมติฐาน	ความจริง	
	H_0 เป็นจริง	H_0 ไม่เป็นจริง
ไม่ปฏิเสธ H_0	$1 - \alpha$ (confidence)	β (type II error)
ปฏิเสธ H_0	α (type I error)	$1 - \beta$ (power)

การกำหนดค่าต่างๆ ที่จำเป็นต้องใช้ในการคำนวณขนาดตัวอย่าง

1) ความไวหรือความจำเพาะที่คาดการณ์ไว้ (anticipated sensitivity/specificity) ผู้วิจัยสามารถประมาณได้จากการศึกษาที่ผ่านมาหรือจากประสบการณ์ของผู้เชี่ยวชาญในสาขานั้นซึ่งโดยทั่วไปแล้วค่าความไวควรอยู่ในช่วง 0.70 ถึง 0.90 และค่าความจำเพาะไม่ควรต่ำกว่า 0.70^{13,16}

2) ค่าความคลาดเคลื่อนที่ยอมรับได้ (acceptable error: d) ในการประมาณค่าพารามิเตอร์ ผู้วิจัยต้องกำหนดให้ค่าความไวหรือความจำเพาะ $\pm d$ มีค่าอยู่ในช่วง 0 ถึง 1 เช่น ผู้วิจัยกำหนดค่าความคลาดเคลื่อนที่ยอมรับได้ในการประมาณค่าความไวไม่เกิน 10% ถ้าสมมติว่าค่าความไวจริงในประชากรเท่ากับ 0.80 ดังนั้น ผู้วิจัยเชื่อมั่น $(1 - \alpha)100\%$ ว่าค่าความไวในประชากรจะมีค่าอยู่ระหว่าง 0.70 ถึง

0.90 สำหรับการกำหนดค่า d นั้นสามารถทำได้ 2 แบบคือ กำหนดให้ d เป็นค่าคงที่ เช่น 0.05 เป็นต้น หรืออาจจะกำหนดให้ d เป็นค่าที่สัมพันธ์กับความไวหรือความจำเพาะ เช่น ให้ $d=10\%$ ของค่าความไวหรือความจำเพาะ ถ้าค่าความไวเท่ากับ 0.80 จะได้ว่า $d=0.10*0.80=0.08$ เป็นต้น^{7,13}

3) ค่าระดับนัยสำคัญ (α) โดยทั่วไปจะกำหนดให้มีค่าเท่ากับ 0.05 หรือ 0.01 นั้นหมายถึงผู้วิจัยต้องการประมาณค่าความไวหรือความจำเพาะด้วยความเชื่อมั่น 95% หรือ 99% ตามลำดับ ซึ่งการประมาณค่าด้วยความเชื่อมั่นสูงๆ ก็จำเป็นที่จะต้องใช้นขนาดตัวอย่างที่มากขึ้นตามไปด้วย ในทางปฏิบัติผู้วิจัยส่วนใหญ่จะกำหนดให้ค่า $\alpha=0.05$ ดังนั้น Z-score for type I error จึงเท่ากับ $Z_{1-\alpha/2}=1.96$

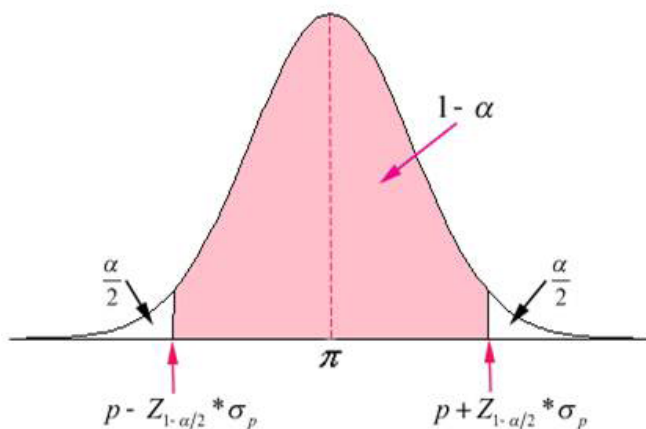
4) ความชุกของโรค (prevalence of disease: P) ผู้วิจัยสามารถประมาณได้จากการศึกษาในอดีตที่มีรายงานค่าความชุกของโรคที่สนใจจะศึกษา

5) อำนาจการทดสอบ (Power: $1-\beta$) ผู้วิจัยสามารถกำหนดได้เองตามความเหมาะสมของงานวิจัยในเรื่องนั้นๆ หรืออาจจะกำหนดโดยใช้ Rule of thumb คืออัตราส่วนต่ำสุดที่ยอมรับได้ของ β ต่อ α ควรจะประมาณ 4 ต่อ 1 ยกตัวอย่างเช่น ถ้ากำหนดค่า $\alpha=0.05$ ดังนั้น ค่า Power ควรใช้เท่ากับ 80%^{8,17} ในทางปฏิบัติผู้วิจัยส่วนใหญ่จะกำหนดให้ค่า Power = 80% เมื่อ $\alpha=0.05$

การหาขนาดตัวอย่างสำหรับการศึกษเกี่ยวกับการตรวจวินิจฉัยที่ผลลัพธ์ของการทดสอบมีค่าได้เพียง 2 ค่า (binary outcome)

1. วิธีการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าความไวและความจำเพาะ มีวัตถุประสงค์เพื่อให้ได้ค่าประมาณที่มีความแม่นยำ ในการรายงานผลการศึกษารายงานค่าที่คำนวณได้จากข้อมูลตัวอย่างและค่า $(1-\alpha)100\%$ Confidence interval (CI)¹⁵ ยกตัวอย่างเช่น ความไวของ cone beam computed tomography ในการตรวจพบการหักของกระดูก scaphoid มีค่าเท่ากับ 93% (95% CI, 89 - 96) โดยที่มาของสูตรการคำนวณขนาดตัวอย่างมีดังนี้

กำหนดให้ค่าสัดส่วนของประชากร (the population proportion) แทนด้วย π และค่าสัดส่วนของตัวอย่าง (the sample proportion) แทนด้วย p โดยที่การแจกแจงค่าตัวอย่าง (the sampling distribution) ของ p มีการแจกแจงใกล้เคียงการแจกแจงปกติ (normal distribution) ด้วยค่าเฉลี่ย: $E(p)=\pi$ และความแปรปรวน: $Var(p)=\sigma_p^2=\pi(1-\pi)/n$ ซึ่งการแจกแจงค่าตัวอย่างแสดงดังรูปที่ 1



รูปที่ 1 แสดงการแจกแจงค่าตัวอย่างของ p

การประมาณค่าสัดส่วนของประชากรสามารถหาได้ดังนี้¹⁸⁻¹⁹

เมื่อตัวอย่างมีขนาดใหญ่ (หรือ $np \geq 5$) และ $n(1-p) \geq 5$ แล้ว จะได้ว่า

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \approx \sqrt{\frac{p(1-p)}{n}}$$

ดังนั้น ช่วงเชื่อมั่นของ π คือ

$$P(p - Z_{1-\alpha/2} * \sigma_p \leq \pi \leq p + Z_{1-\alpha/2} * \sigma_p) = 1 - \alpha$$

$$\text{หรือ } (1-\alpha)100\% \text{ CI ของ } \pi = p \pm Z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} \quad (1)$$

หลังจากนั้นให้ $(1-\alpha)100\% \text{ CI}$ ของ $\pi = p \pm d$

$$\text{ซึ่งจะได้ว่า } d = Z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$$\text{ดังนั้น } n = Z_{1-\alpha/2}^2 * \left(\frac{p(1-p)}{d^2} \right) \quad (2)$$

เมื่อ n คือขนาดตัวอย่างที่ต้องการใช้เพื่อประมาณค่าสัดส่วน

P คือค่าสัดส่วนที่คาดการณ์ไว้

$Z_{1-\alpha/2}$ คือ Z-score for type I error เช่น กำหนดให้ $\alpha = 0.05$ ดังนั้น

$Z_{0.975} = 1.96$ และ d คือค่าความคลาดเคลื่อนที่ยอมรับได้

ซึ่งจะได้สูตรสำหรับการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าสัดส่วนดังสูตร (2) และจะสังเกตได้ว่าค่า d นี้บ่งบอกถึงความเที่ยง (precision) ในการประมาณค่าพารามิเตอร์ π ถ้าค่า d ยิ่งน้อยก็จะส่งผลให้ขนาดตัวอย่างที่ใช้ในการศึกษาเพิ่มมากขึ้น

ในการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าความไวและความจำเพาะจึงประยุกต์มาจากสูตร (2) ดังนี้

$$n_{se} = \frac{Z_{1-\alpha/2}^2 \times Se \times (1 - Se)}{d^2} \quad (3)$$

$$n_{sp} = \frac{Z_{1-\alpha/2}^2 \times Sp \times (1 - Sp)}{d^2} \quad (4)$$

เมื่อ n_{se} คือขนาดตัวอย่างที่ต้องการใช้เพื่อประมาณค่าความไว

n_{sp} คือขนาดตัวอย่างที่ต้องการใช้เพื่อประมาณค่าความจำเพาะ

Se คือความไวที่คาดการณ์ไว้

และ Sp คือความจำเพาะที่คาดการณ์ไว้

ซึ่งขนาดตัวอย่างที่คำนวณได้จะเพียงพอสำหรับการประมาณค่าความไวหรือความจำเพาะเท่านั้นแต่ไม่ใช่ทั้งสอง ดังนั้น ในปี ค.ศ. 1996 Buderer⁷ ได้พิจารณาค่าความชุกของโรค (P)

สูตรการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าความไวคือ

$$n_{se} = \frac{Z_{1-\alpha/2}^2 \times Se \times (1 - Se)}{d^2 \times P} \quad (5)$$

และสูตรการคำนวณขนาดตัวอย่างเพื่อการประมาณค่าความจำเพาะคือ

$$n_{sp} = \frac{Z_{1-\alpha/2}^2 \times Sp \times (1 - Sp)}{d^2 \times (1 - P)} \quad (6)$$

เมื่อ P คือความชุกของโรค

การปรับขนาดตัวอย่าง (sample size: n)

1) กรณีประชากรอันตะ (finite population)²⁰

ถ้าหากประชากรที่มีจำนวนสมาชิกเป็นจำนวนจำกัด เมื่อคำนวณ n ตามสูตร (5)

$$n' = \frac{n}{1 + \frac{n}{N}} \quad (7)$$

จะสังเกตได้ว่า n' จะมีค่าน้อยกว่า n เสมอ แต่ถ้า N มีขนาดใหญ่มากๆ จะส่งผลให้ n' มีค่าใกล้เคียงกับ n

2) กรณีประชากรอนันต์ (infinite population)

ถ้าหากประชากรมีจำนวนสมาชิกเป็นจำนวนอนันต์ (infinity) ในการคำนวณขนาดตัวอย่างจะใช้เฉพาะสูตร (5) หรือสูตร (6)

สำหรับงานวิจัยทางการแพทย์โดยส่วนใหญ่จะไม่ทราบจำนวนผู้ป่วยทั้งหมด (N) จึงอาจจะถือได้ว่าประชากรมีจำนวนสมาชิกเป็นจำนวนอนันต์ ในการคำนวณขนาดตัวอย่างจึงใช้เฉพาะสูตร (5) หรือสูตร (6) เท่านั้น

เข้ามารวมไว้ในสูตรการคำนวณด้วยเมื่อผู้วิจัยไม่ทราบสถานะของโรคที่แท้จริง (true disease status) ณ เวลาที่ทำการสุ่มตัวอย่าง (sampling) ดังนี้

หรือสูตร (6) แล้ว จะต้องทำการปรับขนาดของ n ที่คำนวณได้ตามขนาดประชากร (N) ซึ่งจะได้ขนาดตัวอย่างใหม่ (n') โดยใช้สูตรการคำนวณดังนี้

ตัวอย่างที่ 1 ผู้วิจัยต้องการประเมินความแม่นยำของวิธี diaphragmatic height index ในการวินิจฉัยภาวะเส้นประสาทฟรีนิกได้รับบาดเจ็บในผู้ป่วยประสาทแขนได้รับบาดเจ็บจากอุบัติเหตุเทียบกับวิธีการตรวจวินิจฉัยมาตรฐานคือ intraoperative phrenic nerve stimulation เมื่อกำหนดค่าความไวและความจำเพาะที่คาดการณ์ไว้เท่ากับ 0.90 และ 0.85 ตามลำดับ ระดับนัยสำคัญเท่ากับ 0.05 ส่วนค่าความคลาดเคลื่อนที่ยอมรับได้ในการประมาณค่าความไว/ความจำเพาะเป็นค่าคงที่ 0.10 (ไม่เกิน 10%) ($d = 0.10$) และความชุกของภาวะเส้นประสาทฟรีนิกได้รับบาดเจ็บในผู้ป่วยชายประสาทแขนได้รับบาดเจ็บจากอุบัติเหตุเท่ากับ 24% ($P = 0.24$)

ขนาดตัวอย่างที่ต้องใช้เพื่อการประมาณค่าความไวสามารถหาได้ดังนี้

$$n_{se} = \frac{Z_{1-\alpha/2}^2 \times Se \times (1 - Se)}{d^2 \times P} = \frac{1.96^2 \times 0.90 \times (1 - 0.90)}{0.10^2 \times 0.24} = 144.06 \approx 145$$

และขนาดตัวอย่างที่ต้องใช้เพื่อการประมาณค่าความจำเพาะสามารถหาได้ดังนี้

$$n_{sp} = \frac{Z_{1-\alpha/2}^2 \times Sp \times (1 - Sp)}{d^2 \times (1 - P)} = \frac{1.96^2 \times 0.85 \times (1 - 0.85)}{0.10^2 \times (1 - 0.24)} = 64.44 \approx 65$$

เนื่องจาก n_{se} ที่คำนวณได้มีค่ามากกว่า n_{sp} ดังนั้น การศึกษานี้ต้องใช้ตัวอย่างผู้ป่วยที่มีภาวะประสาทแขนได้รับบาดเจ็บจากอุบัติเหตุจำนวน 145 ราย เพื่อให้ครอบคลุมทั้งการศึกษาในการประมาณค่าความไวและความจำเพาะ

นอกจากนี้ ผู้นิพนธ์ได้นำเสนอตารางขนาดตัวอย่างที่จำเป็นต้องใช้ในงานวิจัยที่มีวัตถุประสงค์เพื่อต้องการประมาณค่าความไวและความจำเพาะของการตรวจวินิจฉัยแบบใหม่ แสดงดังตารางที่ 3 และ 4 ดังนี้

ตารางที่ 3 ขนาดตัวอย่างที่ต้องการใช้เพื่อการประมาณค่าความไวและความจำเพาะ เมื่อกำหนดระดับนัยสำคัญ 0.05 และ $d = 0.05$

Prevalence of disease (P)	Anticipated sensitivity (Se)					Anticipated specificity (Sp)				
	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
0.10	3227	2882	2459	1960	1383	359	321	274	218	154
0.20	1614	1441	1230	980	692	404	361	308	245	173
0.30	1076	961	820	654	461	461	412	352	280	198
0.40	807	721	615	490	346	538	481	410	327	231
0.50	646	577	492	392	277	646	577	492	392	277
0.60	538	481	410	327	231	807	721	615	490	346
0.70	461	412	352	280	198	1076	961	820	654	461
0.80	404	361	308	245	173	1614	1441	1230	980	692
0.90	359	321	274	218	154	3227	2882	2459	1960	1383

ตารางที่ 4 ขนาดตัวอย่างที่ต้องการใช้เพื่อการประมาณค่าความไวและความจำเพาะ เมื่อกำหนดระดับนัยสำคัญ 0.05 และ $d = 0.10$

Prevalence of disease (P)	Anticipated sensitivity (Se)					Anticipated specificity (Sp)				
	0.70	0.75	0.80	0.85	0.90	0.70	0.75	0.80	0.85	0.90
0.10	807	721	615	490	346	90	81	69	55	39
0.20	404	361	308	245	173	101	91	77	62	44
0.30	269	241	205	164	116	116	103	88	70	50
0.40	202	181	154	123	87	135	121	103	82	58
0.50	162	145	123	98	70	162	145	123	98	70
0.60	135	121	103	82	58	202	181	154	123	87
0.70	116	103	88	70	50	269	241	205	164	116
0.80	101	91	77	62	44	404	361	308	245	173
0.90	90	81	69	55	39	807	721	615	490	346

จากตารางที่ 3 และ 4 เมื่อกำหนด $Se = Sp$ ระดับนัยสำคัญ 0.05 และ $d = 0.05$ และ 0.10 ตามลำดับ จะเห็นได้ชัดเจนว่า ที่ค่าความชุกของโรคเท่ากับ 0.50 ขนาดตัวอย่างที่คำนวณได้จากสูตร (5) และ (6) จะเท่ากัน ซึ่งเป็นผลเนื่องมาจากเทอมที่เป็นตัวหารจากสูตร (5) และ (6) มีค่าเท่ากัน แต่ถ้าความชุกของโรคมีค่าต่ำกว่า 0.50 ขนาดตัวอย่างที่คำนวณจากสูตร (5) จะมีค่ามากกว่าขนาดตัวอย่างที่คำนวณได้จากสูตร (6) เพราะเทอมที่เป็นตัวหารจากสูตร (5) มีค่าน้อยกว่าสูตร (6) ในทางกลับกันถ้าความชุกของโรคมีค่าสูงกว่า 0.50 ขนาดตัวอย่างที่คำนวณได้จากสูตร (5) จะมีค่าน้อยกว่าขนาดตัวอย่างที่คำนวณได้จากสูตร (6) เพราะเทอมที่เป็นตัวหารจากสูตร (5) มีค่ามากกว่าสูตร (6) ดังนั้น ค่าความชุกของโรคจึงเป็นปัจจัยสำคัญที่มีผลต่อขนาดตัวอย่างที่คำนวณได้

ในทางปฏิบัติการพิจารณาเลือกใช้วิธีคำนวณขนาดตัวอย่างด้วยสูตร (5) หรือสูตร (6) นั้นขึ้นอยู่กับวัตถุประสงค์หลักที่สนใจศึกษา (primary interested in diagnostic purpose) แต่ถ้าผู้วิจัยเห็นว่า ค่าความไวและความจำเพาะมีความสำคัญเท่าๆ กัน ในการหาขนาดตัวอย่างจึงควรคำนวณโดยใช้ทั้งสูตร (5) และ

สูตร (6) แล้วเลือกค่าที่มากที่สุดของขนาดตัวอย่างที่คำนวณได้^{7,13}

2. วิธีการคำนวณขนาดตัวอย่างเพื่อการทดสอบความไวหรือความจำเพาะของการตรวจวินิจฉัย

สำหรับสูตรที่ใช้ในการคำนวณขนาดตัวอย่างเพื่อการทดสอบความไวหรือความจำเพาะของการตรวจวินิจฉัยขึ้นอยู่กับพื้นฐานของการแจกแจงใกล้เคียงการแจกแจงปกติ โดยมีรายละเอียดดังนี้¹⁸

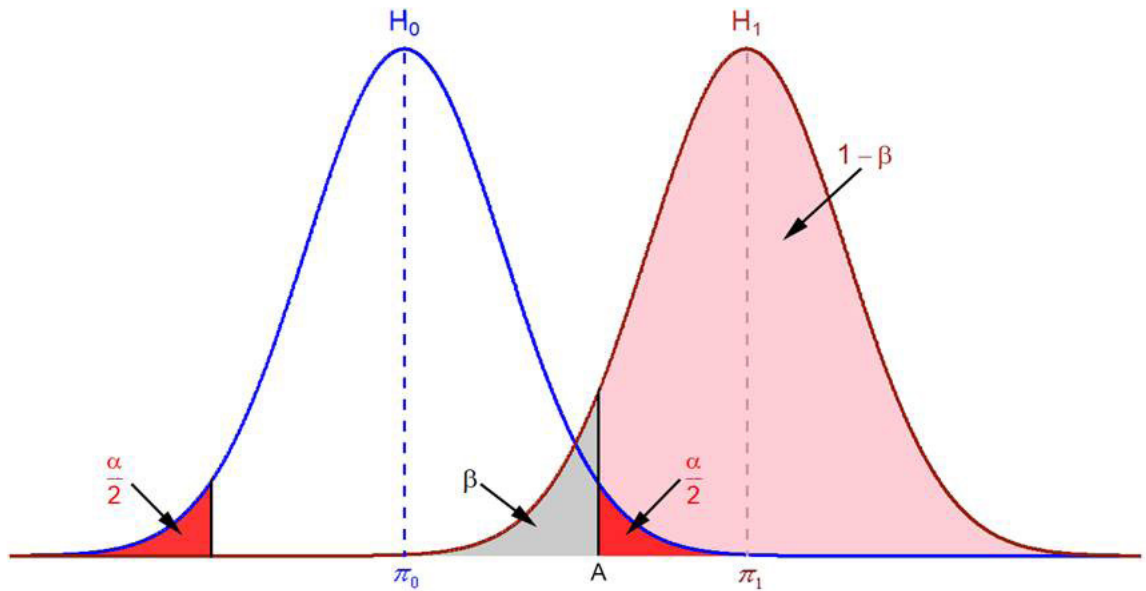
จากสมมติฐานของการทดสอบค่าสัดส่วนคือ

$$H_0 : \pi = \pi_0 \text{ และ}$$

$$H_1 : \pi \neq \pi_0 \text{ (or } \pi = \pi_1)$$

โดยที่ π_0 คือค่าสัดส่วนที่คาดการณ์ไว้และ π_1 คือค่าสัดส่วนภายใต้สมมติฐานทางเลือก (alternative hypothesis: H_1)

เมื่อกำหนดให้ type I error เท่ากับ α และ type II error เท่ากับ β นั่นคือเราต้องการอำนาจการทดสอบ (power of the test) เท่ากับ $1 - \beta$ โดยการแจกแจงค่าตัวอย่างแสดงดังรูปที่ 2



รูปที่ 2 แสดงการแจกแจงค่าตัวอย่างสำหรับการทดสอบสมมติฐานเกี่ยวกับค่าสัดส่วนของประชากรกลุ่มเดียว (one-sample test) แบบสองทาง (two-sided)

จากรูปที่ 2 จะเห็นได้ว่าการแจกแจงค่าตัวอย่างจะมีศูนย์กลางที่ π_0 เมื่อสมมติฐาน H_0 เป็นจริง ดังนั้น จุด A จึงแทน the upper $100(\alpha/2)$ percent point ของการแจกแจงของ p คือ

$$A = \pi_0 + Z_{1-\alpha/2} \sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}} \quad (8)$$

และสำหรับการแจกแจงค่าตัวอย่างจะมีศูนย์กลางที่ π_1 เมื่อสมมติฐาน H_1 เป็นจริง ฉะนั้นจุด A จึงแทน the lower $100(\beta)$ percent point ของการแจกแจงของ p คือ

$$A = \pi_1 - Z_{1-\beta} \sqrt{\frac{\pi_1 * (1 - \pi_1)}{n}} \quad (9)$$

เมื่อ $Z_{1-\beta}$ คือ Z-score for type II error เช่น กำหนดให้ $\beta = 0.20$ ดังนั้น $Z_{0.80} = 0.84$ หลังจากนั้นให้สมการ (8) เท่ากับสมการ (9) เพื่อแก้สมการหาค่า n ดังนี้

$$\pi_0 + Z_{1-\alpha/2} \sqrt{\frac{\pi_0 * (1 - \pi_0)}{n}} = \pi_1 - Z_{1-\beta} \sqrt{\frac{\pi_1 * (1 - \pi_1)}{n}}$$

จะได้ว่า

$$n = \left[\frac{Z_{1-\alpha/2} \sqrt{\pi_0 * (1 - \pi_0)} + Z_{1-\beta} \sqrt{\pi_1 * (1 - \pi_1)}}{\pi_1 - \pi_0} \right]^2 \quad (10)$$

เมื่อ n คือ ขนาดตัวอย่างที่ต้องการใช้เพื่อการทดสอบค่าสัดส่วน

ดังนั้น จะได้วิธีการคำนวณขนาดตัวอย่างที่ใช้เพื่อการทดสอบค่าสัดส่วนดังสูตร (10) ซึ่งในการคำนวณขนาดตัวอย่างที่ใช้เพื่อการทดสอบค่าความไวและความจำเพาะจึงประยุกต์มาจากสูตร (10) ดังนี้^{9,13}

สมมติให้ Se_0 และ Sp_0 คือค่าความไวและความจำเพาะที่คาดการณ์ไว้ของการตรวจวินิจฉัยแบบใหม่ ตามลำดับ

สมมติฐานของการทดสอบค่าความไวคือ

$$H_0 : Se = Se_0 \text{ และ } H_1 : Se \neq Se_0 (or Se = Se_1)$$

และสมมติฐานของการทดสอบค่าความจำเพาะคือ

$$H_0 : Sp = Sp_0 \text{ และ } H_1 : Sp \neq Sp_0 (or Sp = Sp_1)$$

เมื่อ Se_1 และ Sp_1 คือค่าความไวและความจำเพาะภายใต้สมมติฐานทางเลือก (alternative hypothesis: H_1) ตามลำดับ

สูตรการคำนวณขนาดตัวอย่างเพื่อการทดสอบความไวคือ

$$n_{disease} = \left[\frac{Z_{1-\alpha/2} \sqrt{Se_0 * (1 - Se_0)} + Z_{1-\beta} \sqrt{Se_1 * (1 - Se_1)}}{Se_1 - Se_0} \right]^2 \quad (11)$$

$$n_{total(Se)} = \frac{n_{disease}}{P} \quad (12)$$

และสูตรการคำนวณขนาดตัวอย่างเพื่อการทดสอบความจำเพาะคือ

$$n_{nondisease} = \left[\frac{Z_{1-\alpha/2} \sqrt{Sp_0 * (1 - Sp_0)} + Z_{1-\beta} \sqrt{Sp_1 * (1 - Sp_1)}}{Sp_1 - Sp_0} \right]^2 \quad (13)$$

$$n_{total(Sp)} = \frac{n_{nondisease}}{1 - P} \quad (14)$$

เมื่อ $n_{disease}$ คือขนาดตัวอย่างสำหรับคนที่เป็โรค
 $n_{total(Se)}$ คือขนาดตัวอย่างที่ต้องการใช้เพื่อการทดสอบความไว ซึ่งประกอบด้วยคนที่เป็โรคจำนวน $n_{disease}$ ราย และคนที่ไม่เป็โรคจำนวน $n_{total(Se)} - n_{disease}$ ราย
 $n_{nondisease}$ คือขนาดตัวอย่างสำหรับคนที่ไม่เป็โรค
 $n_{total(Sp)}$ คือขนาดตัวอย่างที่ต้องการใช้เพื่อการทดสอบความจำเพาะ ซึ่งประกอบด้วยคนที่ไม่เป็โรคจำนวน $n_{nondisease}$ ราย และคนที่เป็โรคจำนวน $n_{total(Sp)} - n_{nondisease}$ ราย

P คือความชุกของโรค

และ Power คืออำนาจการทดสอบสำหรับการตรวจพบความต่างของ $Se_1 - Se_0$ (หรือ $Sp_1 - Sp_0$)

ตัวอย่างที่ 2 ผู้วิจัยต้องการประเมินความแม่นยำของการทดสอบด้วย tuning fork และ stethoscope ในการวินิจฉัยภาวะกระดูกหักโดยเปรียบเทียบกับ การทดสอบอ้างอิงคือ radiographic testing เมื่อกำหนดสมมติฐานในการทดสอบความไวและความจำเพาะคือ $H_0 : Se = 0.80$ และ $H_1 : Se = 0.90$ และ $H_0 : Sp = 0.75$ และ $H_1 : Sp = 0.85$ ตามลำดับ ระดับนัยสำคัญเท่ากับ 0.05 อำนาจการทดสอบ (Power) เท่ากับ 80% และความชุกของภาวะกระดูกหักในผู้ใหญ่ในกลุ่มเสี่ยงเท่ากับ 48%

ขนาดตัวอย่างที่ต้องใช้เพื่อการทดสอบค่าความไวสามารถหาได้ดังนี้

$$n_{disease} = \left[\frac{Z_{1-\alpha/2} \sqrt{Se_0 * (1 - Se_0)} + Z_{1-\beta} \sqrt{Se_1 * (1 - Se_1)}}{Se_1 - Se_0} \right]^2$$

$$= \left[\frac{\{1.96 \times \sqrt{0.80 \times (1 - 0.80)}\} + \{0.84 \times \sqrt{0.90 \times (1 - 0.90)}\}}{0.90 - 0.80} \right]^2 = 107.43$$

$$n_{total(Se)} = \frac{n_{disease}}{P} = \frac{107.43}{0.48} = 223.81 \approx 224$$

ขนาดตัวอย่างที่ต้องใช้เพื่อการทดสอบค่าความจำเพาะสามารถหาได้ดังนี้

$$n_{nondisease} = \left[\frac{Z_{1-\alpha/2} \sqrt{Sp_0 * (1 - Sp_0)} + Z_{1-\beta} \sqrt{Sp_1 * (1 - Sp_1)}}{Sp_1 - Sp_0} \right]^2$$

$$= \left[\frac{\{1.96 \times \sqrt{0.75 \times (1 - 0.75)}\} + \{0.84 \times \sqrt{0.85 \times (1 - 0.85)}\}}{0.85 - 0.75} \right]^2 = 132.07$$

$$n_{total(Sp)} = \frac{n_{nondisease}}{1 - P} = \frac{132.07}{1 - 0.48} = 253.98 \approx 254$$

เนื่องจาก $n_{total(Sp)}$ ที่คำนวณได้มีค่ามากกว่า $n_{total(Se)}$ ดังนั้น การศึกษานี้ต้องใช้ตัวอย่างผู้ป่วยที่สงสัยว่ามีกระดูกหัก จำนวน 254 ราย เพื่อให้ครอบคลุมทั้งการศึกษาในการทดสอบความไวและความจำเพาะ

นอกจากนี้ ผู้นิพนธ์ได้นำเสนอตารางขนาดตัวอย่างที่จำเป็นต้องใช้ในงานวิจัยที่มีวัตถุประสงค์เพื่อต้องการทดสอบความไวและความจำเพาะของการตรวจวินิจฉัยแบบใหม่ แสดงดังตารางที่ 5 ดังนี้

ตารางที่ 5 ขนาดตัวอย่างที่ต้องการใช้เพื่อการทดสอบความไวและความจำเพาะ เมื่อกำหนดระดับนัยสำคัญ 0.05 และ Power = 80%

Prevalence of disease (P)	The required sample size for testing sensitivity				The required sample size for testing specificity			
	H_0	H_1	$n_{disease}$	$n_{total}(Se)$	H_0	H_1	$n_{disease}$	$n_{total}(Sp)$
0.10	0.70	0.80	153	1525	0.70	0.80	17	170
0.10	0.80	0.90	108	1075	0.80	0.90	12	120
0.10	0.90	0.95	239	2381	0.90	0.95	27	265
0.20	0.70	0.80	153	763	0.70	0.80	39	191
0.20	0.80	0.90	108	538	0.80	0.90	27	135
0.20	0.90	0.95	239	1191	0.90	0.95	60	298
0.30	0.70	0.80	153	509	0.70	0.80	66	218
0.30	0.80	0.90	108	359	0.80	0.90	47	154
0.30	0.90	0.95	239	794	0.90	0.95	103	341
0.40	0.70	0.80	153	382	0.70	0.80	102	255
0.40	0.80	0.90	108	269	0.80	0.90	72	180
0.40	0.90	0.95	239	596	0.90	0.95	159	397
0.50	0.70	0.80	153	305	0.70	0.80	153	305
0.50	0.80	0.90	108	215	0.80	0.90	108	215
0.50	0.90	0.95	239	477	0.90	0.95	239	477
0.60	0.70	0.80	153	255	0.70	0.80	229	382
0.60	0.80	0.90	108	180	0.80	0.90	162	269
0.60	0.90	0.95	239	397	0.90	0.95	358	596
0.70	0.70	0.80	153	218	0.70	0.80	356	509
0.70	0.80	0.90	108	154	0.80	0.90	251	359
0.70	0.90	0.95	239	341	0.90	0.95	556	794
0.80	0.70	0.80	153	191	0.70	0.80	610	763
0.80	0.80	0.90	108	135	0.80	0.90	430	538
0.80	0.90	0.95	239	298	0.90	0.95	953	1191
0.90	0.70	0.80	153	170	0.70	0.80	1373	1525
0.90	0.80	0.90	108	120	0.80	0.90	967	1075
0.90	0.90	0.95	239	265	0.90	0.95	2143	2381

จากตารางที่ 5 จะสังเกตได้ว่า ที่ค่าความชุกของโรคเท่ากับ 0.50 ขนาดตัวอย่างที่ต้องการใช้เพื่อทดสอบความไวกับความจำเพาะจะมีค่าเท่ากัน ซึ่งเป็นผลเนื่องมาจากเทอมที่เป็นตัวหารจากสูตร (12) และ (14) มีค่าเท่ากัน แต่ถ้าความชุกของโรคมีค่าต่ำกว่า 0.50 ขนาดตัวอย่างที่ต้องการใช้เพื่อทดสอบความไวจะมีค่ามากกว่าขนาดตัวอย่างที่ต้องการใช้เพื่อทดสอบความจำเพาะเพราะเทอมที่เป็นตัวหารจากสูตร (12) มีค่าน้อยกว่าสูตร (14) ในทางกลับกันถ้าความชุกของโรคมีค่าสูงกว่า 0.50 ขนาดตัวอย่างที่ต้องการใช้เพื่อทดสอบความไวจะมีค่าน้อยกว่าขนาดตัวอย่างที่ต้องการใช้เพื่อทดสอบความจำเพาะเพราะเทอมที่เป็นตัวหารจากสูตร (12) มีค่ามากกว่าสูตร (14) ดังนั้นค่าความชุกของโรคจึงเป็นปัจจัยสำคัญที่มีผลต่อขนาดตัวอย่างที่คำนวณได้

ในทางปฏิบัติการพิจารณาเลือกใช้วิธีคำนวณขนาดตัวอย่างเพื่อทดสอบความไวหรือความจำเพาะนั้นขึ้นอยู่กับวัตถุประสงค์หลักที่สนใจศึกษา แต่ถ้าผู้วิจัยเห็นว่า ค่าความไวและความจำเพาะมีความสำคัญเท่าๆ กันในการหาขนาดตัวอย่างจึงควรคำนวณโดยใช้ทั้งสูตรสำหรับการทดสอบความไวและความจำเพาะแล้วเลือกค่าที่มากที่สุดของขนาดตัวอย่างที่คำนวณได้¹³

สรุปผล

การคำนวณขนาดตัวอย่างสำหรับการศึกษเกี่ยวกับ การตรวจวินิจฉัยที่ผลลัพธ์มีค่าได้เพียง 2 ค่า คือผลบวกและผลลบมีหลายวิธีขึ้นอยู่กับวัตถุประสงค์หลักของงานวิจัย โดยสูตรขนาดตัวอย่างที่กล่าวมาทั้งหมดข้างต้น จะเห็นได้ชัดว่าความชุกของโรคเป็นค่าที่สำคัญที่ส่งผลต่อขนาดตัวอย่าง กรณีที่ค่าความชุกของโรคเท่ากับ 0.50 ขนาดตัวอย่างที่

ต้องการใช้เพื่อประมาณค่า/การทดสอบความไวและความจำเพาะจะมีค่าเหมือนกัน แต่ถ้าหากความชุกของโรคไม่เท่ากับ 0.50 ผู้นิพนธ์เห็นว่าการคำนวณขนาดตัวอย่างควรใช้ทั้งสูตรขนาดตัวอย่างเพื่อประมาณค่า/การทดสอบความไวและความจำเพาะแล้วเลือกขนาดตัวอย่างที่มีค่ามากที่สุดที่คำนวณได้เพื่อให้ครอบคลุมทั้งการศึกษาในการประมาณค่า/การทดสอบความไวและความจำเพาะ

References

1. Petrie A, Sabin C. Medical statistics at a glance. 3rd ed. West Sussex: Wiley-Blackwell; 2009. p. 115-7.
2. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. New Jersey: Julie Levin Alexander; 2009. p. 619-58.
3. Antonisamy B, Premkumar PS, Christopher S. Principles and practice of biostatistics. 1st ed. New Delhi: Elsevier; 2017. p. 213-26.
4. Jones S, Carley S, Harrison M. An introduction to power and sample size estimation. Emerg Med J 2003;20(5): 453-8.
5. Peat J, Barton B, Elliott E. Statistics workbook for evidence-based health care. 1st ed. West Sussex: Wiley-Blackwell; 2008. p. 147-53.
6. Lwanga SK, Lemeshow S. Sample size determination in health studies: a practical manual. 1st ed. Geneva: World Health Organization; 1991. p. VII-VIII.

7. Buderer NM. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;3(9):895-900.
8. Daly LE, Bourke GJ. Interpretation and uses of medical statistics. 5th ed. Oxford: Blackwell Science; 2000. p. 269-95.
9. Li J, Fine J. On sample size for sensitivity and specificity in prospective diagnostic accuracy studies. *Stat Med* 2004; 23(16):2537-50.
10. Chow SC, Shao J, Wang H. Sample size calculations in clinical research. 2nd ed. New York: Chapman & Hall/CRC; 2008. p. 1-47.
11. Machin D, Campbell MJ, Tan SB, et al. Sample size tables for clinical studies. 3rd ed. West Sussex: Wiley-Blackwell; 2009. p. 1-13.
12. Monette DR, Sullivan TJ, DeJong CR. Applied social research: a tool for the human services. 8th ed. Belmont, CA: Brooks/Cole, Cengage Learning; 2011. p. 134-62.
13. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193-204.
14. Connell FA, Koepsell TD. Measures of gain in certainty from a diagnostic test. *Am J Epidemiol* 1985;121(5):744-53.
15. Arkin CF, Wachtel MS. How many patients are necessary to assess test performance?. *JAMA* 1990;263(2):275-8.
16. Kent S, Wanzek J, Yun J. Screening in the upper elementary grades: Identifying fourth-grade students at-risk for failing the state reading assessment. *Assess Eff Interv* 2018;1:1-13.
17. Monette DR, Sullivan TJ, DeJong CR. Applied social research: a tool for the human services. 8th ed. Belmont, CA: Brooks/Cole, Cengage Learning; 2011. p. 425.
18. Lemeshow S, Hosmer Jr DW, Klar J, et al. Adequacy of sample size in health studies. 1st ed. Chichester: John Wiley & Sons Ltd.; 1990. p. 1-8.
19. Asadoorian OM, Kantarelis D. Essentials of inferential statistics. 4th ed. Oxford: University Press of America, Inc.; 2005. p. 94.
20. Machin D, Campbell MJ, Tan SB, et al. Sample size tables for clinical studies. 3rd ed. West Sussex: Wiley-Blackwell; 2009. p. 122-5.