

บทที่ 4

ขนาดกลุ่มตัวอย่างในการวิจัยทางคลินิก

จรณิต แก้วกังวาล และ ประตาป สิงหศิวานนท์

ขนาดของกลุ่มตัวอย่างเป็นเรื่องที่
สนใจเรื่องหนึ่งของนักวิจัยทางคลินิก
เพราะขนาดกลุ่มตัวอย่างที่น้อยเกินไป
อาจทำให้ไม่อาจสรุปผลการวิจัยได้ เนื่องจาก
ไม่มีข้อมูลมากพอที่จะทราบผลการศึกษา
หรือค่าสถิติที่ได้ไม่อาจให้ผลสรุปหรือคำตอบ
ที่สนใจจะศึกษา ขนาดกลุ่มตัวอย่างที่มาก
เกินไป อาจจะทำให้สิ้นเปลืองทั้งแรงงานและ
เพิ่มต้นทุนการวิจัยโดยเปล่าประโยชน์
นอกจากนั้นขนาดกลุ่มตัวอย่างที่ใหญ่มากๆ
อาจนำไปสู่การพบความแตกต่างอย่างมี

นัยสำคัญทางสถิติ ที่อาจไม่มีความสำคัญทาง
คลินิกที่แท้จริง ขนาดกลุ่มตัวอย่างที่น้อย
เกินไปหรือมากเกินไปยังเป็นปัญหาทางด้าน
จริยธรรมการวิจัย หากน้อยเกินไป ก็จะทำให้
ผู้ที่เข้าร่วมการวิจัยที่ไม่มีคำตอบ เสียเวลา
และความตั้งใจโดยเปล่าประโยชน์ หากมาก
เกินไปก็จะทำให้ผู้เข้าร่วมงานวิจัยส่วนหนึ่ง
เพิ่มความเสี่ยงโดยไม่มีความจำเป็น ทั้งที่
ผู้วิจัยเองก็ยังไม่ทราบว่า สิ่งที่ได้ทดลอง (ยา,
อุปกรณ์, วัคซีน ฯลฯ) จะได้ผลหรือไม่ หรือ
อาจก่อโทษแทนที่จะให้คุณหรือไม่

คณะกรรมการจริยธรรมการวิจัยทางคลินิกของสถาบันหลายแห่ง มักจะขอให้ผู้วิจัยแสดงวิธีการคำนวณ หรือแสดงสูตรคำนวณขนาดของกลุ่มตัวอย่างไว้ในโครงร่างการวิจัย ทั้งนี้เพื่อจะได้พิจารณาว่าผู้วิจัยจะมีเหตุผลที่มีน้ำหนักพอและถูกต้องหรือไม่ ในการนำผู้เข้าร่วมวิจัยตามจำนวนที่คำนวณได้ มาเข้าสู่กระบวนการวิจัยทางคลินิก เพื่อให้ได้คำตอบอันมีประโยชน์ที่สุด

เนื้อหาในบทนี้จะได้กล่าวถึง วิธีการคำนวณขนาดของกลุ่มตัวอย่างที่เหมาะสมกับการวิจัยทางคลินิก โดยแบ่งเป็น 5 หัวข้อใหญ่ๆ คือ

1. รูปแบบของงานวิจัยและการคำนวณขนาดของกลุ่มตัวอย่าง
2. ค่าสถิติที่เกี่ยวข้องกับการคำนวณกลุ่มตัวอย่าง
3. สูตรคำนวณขนาดกลุ่มตัวอย่างของแต่ละรูปแบบการวิจัย
4. การปรับขนาดของกลุ่มตัวอย่างภายใต้เงื่อนไขเฉพาะ
5. การเขียนอธิบายขนาดกลุ่มตัวอย่างในโครงร่างการวิจัย

รูปแบบของงานวิจัยและการคำนวณขนาดตัวอย่าง

การคำนวณขนาดกลุ่มตัวอย่างขึ้นอยู่กับคำถามวิจัยและรูปแบบของการวิจัย โดยอาจจำแนกได้เป็น 2 รูปแบบใหญ่ นั่นคือ

รูปแบบที่ 1 การศึกษาเพื่อประมาณค่าพารามิเตอร์ (parameter estimation) เพื่อตอบคำถามวิจัยแบบพรรณนา (descriptive study) ตัวอย่างเช่น ผู้วิจัยต้องการประมาณค่าความชุก (prevalence) ของโรคในประชากร การคำนวณขนาดกลุ่มตัวอย่างจึงเป็นการคำนวณขนาดที่มากพอที่จะช่วยให้ผู้วิจัยสามารถประมาณค่าทางสถิติ (ค่าความชุก) จากกลุ่มตัวอย่างว่าเป็นค่าพารามิเตอร์ (ค่าความชุก) จริงๆ ในประชากรได้อย่างไร เชื่อมันที่ระดับใด (confidence level) หรือค่าสถิติ นั้นเป็นค่าที่ใกล้เคียงหรือตรง (precise/accurate) กับค่าพารามิเตอร์มากน้อยเพียงใด วิธีการนี้จึงเป็นการคำนวณหาขนาดกลุ่มตัวอย่างเพื่อประมาณค่าความถูกต้อง (precision analysis) หากค่าสถิติที่ประมาณได้มีช่วงความคลาดเคลื่อนจากการประมาณ (error of estimation) น้อยก็แสดงว่าการประมาณค่าถูกต้อง ซึ่งตรวจสอบได้จากช่วงห่างของค่าที่ประมาณได้ภายใต้ระดับความมั่นใจที่กำหนด (Confidence Interval หรือ CI) เช่น ถ้าค่าความชุกของโรคร้อยละ 10 เป็นค่าที่ได้จากการศึกษาประมาณ ค่าจากกลุ่มตัวอย่าง 20 ราย ก็อาจจะได้ค่าประมาณพารามิเตอร์ความชุกจริงที่ความเชื่อมั่นร้อยละ 95CI อยู่ระหว่าง ร้อยละ 1 ถึง 31 ซึ่งถือได้ว่ามีความกว้างมาก จนทำให้คำตอบค่าความชุกที่ได้นี้เกือบจะไม่บอกอะไรเกี่ยวกับค่าในประชากรได้มากนักแต่หากค่าความชุกร้อยละ 10 นี้

คำนวณได้จากขนาดกลุ่มตัวอย่าง 400 ราย และที่ระดับความเชื่อมั่นร้อยละ 95 CI ค่าประมาณความชุกในประชากรอาจจะอยู่ระหว่าง ร้อยละ 7 ถึง 13 ซึ่งเป็นช่วงที่แคบลง และมีความหมายมากขึ้นในการสรุปค่าของความชุกของโรคในประชากร

อย่างไรก็ตาม งานวิจัยทางคลินิกส่วนใหญ่มักไม่ใช่เรื่องการประมาณค่าพารามิเตอร์ ดังนั้นในที่นี้จะกล่าวถึงสูตรต่างๆ ที่จะใช้คำนวณขนาดตัวอย่างในเรื่องนี้

รูปแบบที่ 2 การศึกษาเพื่อทดสอบสมมติฐาน (hypothesis testing) เพื่อตอบคำถามวิจัยแบบวิเคราะห์ (analytic study) กล่าวคือเป็นการศึกษาเพื่อเปรียบเทียบหรือหาความสัมพันธ์ระหว่างผลของปัจจัยที่ศึกษาในกลุ่มตัวอย่างต่างกลุ่มกัน โดยผ่านการทดสอบสมมติฐานที่ผู้วิจัยกำหนดขึ้น แล้วใช้วิธีวิเคราะห์ทางสถิติเพื่อประกอบการพิจารณาที่จะปฏิเสธ หรือไม่ปฏิเสธสมมติฐานที่ผู้วิจัยตั้งไว้ การคำนวณขนาดของกลุ่มตัวอย่าง จึงมีความจำเป็นที่จะทำให้สามารถตอบได้อย่างมั่นใจว่า ความแตกต่างหรือความสัมพันธ์ที่เกิดขึ้นและสังเกตได้จากกลุ่มตัวอย่างนั้น เป็นสิ่งที่มีนัยสำคัญจริงๆ และมีความหมายทางคลินิกจริง โดยที่ผลลัพธ์ที่เกิดขึ้นซึ่งทดสอบได้จากกลุ่มตัวอย่างนั้นๆ มิใช่แค่เหตุบังเอิญ แต่เป็นจริงเหมือนกับได้ศึกษาในประชากร หากขนาดของกลุ่มตัวอย่างเล็กเกินไป แม้ว่าเราจะพบความ

แตกต่างกันอย่างมากและมีนัยสำคัญทางสถิติ ก็ยากที่จะบอกได้ว่าความแตกต่างที่เกิดขึ้นนั้น เป็นโดยเหตุบังเอิญ หรือเป็นความแตกต่างจริงที่มีความหมายทางคลินิก

การตั้งสมมติฐานมีหลายรูปแบบ ขึ้นอยู่กับจุดมุ่งหมายที่ผู้วิจัยกำหนดขึ้น หากกำหนดให้

μ_c = ค่าเฉลี่ยของผลการได้รับการรักษาแบบมาตรฐาน

μ_T = ค่าเฉลี่ยของผลการได้รับการรักษาด้วยวิธีการแบบใหม่

δ = ผลต่างระหว่าง 2 กลุ่มที่มีความหมายทางคลินิก (ตามที่ผู้วิจัยตั้งค่าเอาไว้)

เราอาจตั้งสมมติฐานไว้ 4 รูปแบบใหญ่ดังนี้

1. การทดสอบเพื่อแสดงความเท่ากัน (test for equality) ซึ่งมีจุดมุ่งหมายจะแสดงค่าผลลัพธ์ความแตกต่างที่อาจพบระหว่างกลุ่มตัวอย่าง 2 กลุ่มเป็นผลต่างที่มีความหมายทางคลินิกจริงหรือไม่

2. การทดสอบเพื่อแสดงความไม่ด้อยกว่า (test for non-inferiority) เพื่อแสดงว่าผลลัพธ์ของวิธีการแบบใหม่มีประสิทธิภาพผลไม่ด้อยกว่าผลลัพธ์ที่ได้จากวิธีการมาตรฐาน (นั่นคือ ผลต่างระหว่างผลลัพธ์จากวิธีการแบบใหม่และแบบมาตรฐานมีค่าน้อยกว่าค่าผลต่างระหว่าง 2 กลุ่มที่มีความหมายทางคลินิกที่ผู้วิจัยกำหนด)

3. การทดสอบเพื่อแสดงความเหนือกว่า (test for superiority) เพื่อแสดงผลลัพธ์ของวิธีการแบบใหม่ว่ามีประสิทธิภาพเหนือกว่าผลลัพธ์ของวิธีการมาตรฐาน (นั่นคือ ผลต่างระหว่างผลลัพธ์ที่ได้จากวิธีการแบบใหม่ กับวิธีการมาตรฐานมีค่ามากกว่าผลต่างระหว่าง 2 กลุ่มที่มีความหมายทางคลินิกตามที่ผู้วิจัยคาดหวังไว้) การทดสอบเพื่อแสดงความเหนือกว่านั้น บางครั้งก็เรียกว่า การทดสอบเพื่อแสดงความเหนือกว่าทางด้านคลินิก (clinical superiority) หากตั้งค่า $\delta = 0$ เราจะเรียกว่าเป็นการทดสอบเพื่อแสดงความเหนือกว่าทางสถิติ (statistical superiority)

4. การทดสอบเพื่อแสดงความเสมอภาคกัน (test for equivalence) เพื่อแสดงผลต่างระหว่างผลลัพธ์ที่ได้จากวิธีการแบบใหม่กับวิธีการแบบมาตรฐานว่าไม่มีความสำคัญทางสถิติ หากเราปฏิเสธสมมติฐานที่ว่าผลลัพธ์ของ 2 วิธีการนี้ไม่เสมอภาคกันก็

แสดงว่า วิธีการทั้ง 2 นี้ให้ผลลัพธ์ที่เสมอภาคเท่าเทียมกัน

ตารางที่ 1 แสดงการตั้งสมมติฐานที่จะใช้ในการทดสอบทางสถิติ (null hypothesis หรือ H_0) และสมมติฐานที่จะรองรับหากปฏิเสธสมมติฐานที่ทดสอบ (alternative hypothesis หรือ H_a) จะสังเกตได้ว่า H_0 และ H_a ของแต่ละรูปแบบการตั้งสมมติฐานจะแตกต่างกันออกไป ซึ่งจะมีผลต่อการคำนวณขนาดของกลุ่มตัวอย่าง ดังจะได้กล่าวในตอนถัดไป

ค่าสถิติที่เกี่ยวข้องกับการคำนวณขนาดตัวอย่าง

ก่อนการคำนวณหาขนาดกลุ่มตัวอย่าง ผู้วิจัยจะต้องมีความเข้าใจในโครงสร้างการวิจัยของตนเองอย่างถ่องแท้ และต้องสามารถตอบคำถามต่อไปนี้ได้ เพราะคำตอบของคำถามนี้จะมีผลต่อขนาดกลุ่มตัวอย่าง และสูตรที่จะใช้คำนวณหาขนาดกลุ่มตัวอย่าง ก็

ตารางที่ 1 การตั้งสมมติฐานที่จะใช้ในการทดสอบทางสถิติและสมมติฐานที่จะรองรับหากปฏิเสธสมมติฐานที่ทดสอบ

การทดสอบเพื่อแสดง	Null Hypothesis	Alternative Hypothesis
ความเท่ากัน (equality)	$H_0: \mu_T - \mu_c = 0$	$H_a: \mu_T - \mu_c \neq 0$
ความไม่ด้อยกว่า (non-inferiority)	$H_0: \mu_T - \mu_c \geq \delta$	$H_a: \mu_T - \mu_c < \delta$
ความเหนือกว่า (superiority)	$H_0: \mu_T - \mu_c \leq \delta$	$H_a: \mu_T - \mu_c > \delta$
ความเสมอภาคกัน (equivalence)	$H_0: \mu_T - \mu_c \geq \delta$	$H_a: \mu_T - \mu_c < \delta$

แตกต่างกันไป ในแต่ละเงื่อนไข

1. วัตถุประสงค์หลักของการวิจัยนี้ เพื่อประมาณค่าพารามิเตอร์ในประชากร หรือเพื่อการทดสอบสมมติฐาน? การทดสอบสมมติฐานเป็นการทดสอบเปรียบเทียบความแตกต่างระหว่างกลุ่ม หรือเป็นการหาค่าความสัมพันธ์ระหว่างตัวแปร? การวิจัยทางคลินิกส่วนใหญ่จะเป็นการเปรียบเทียบค่าผลลัพธ์ของปัจจัยที่ศึกษา (ยา, วัคซีน, เครื่องมือ วิธีการรักษา) ในระหว่างกลุ่มที่ได้รับหรือไม่ได้รับปัจจัยดังกล่าว

2. โครงร่างการวิจัยมีคำอธิบายปัจจัยที่ศึกษาอย่างละเอียดและชัดเจนหรือไม่?

3. ผลลัพธ์ที่ผู้วิจัยสนใจจะศึกษาคืออะไร?

- ค่าผลลัพธ์ที่จะเป็นผลสรุปของการศึกษาเป็นค่าแบบต่อเนื่อง (continuous) เช่น ระดับน้ำตาล ระดับค่า CD4 หรือเป็นแบบจำแนกแจกแจงค่า (categorical) เช่น หาย/ไม่หาย รอด/ไม่รอด ติดเชื้อ/ไม่ติดเชื้อ?
- ในการวิจัยนี้มีมากกว่า 1 ผลลัพธ์หลักหรือไม่?
- อะไรคือผลลัพธ์ที่เป็นวัตถุประสงค์หลัก (อาจมีมากกว่า 1 วัตถุประสงค์ก็ได้) เราจะคำนวณขนาดของกลุ่มตัวอย่างเพื่อตอบวัตถุประสงค์หลักของการวิจัยนี้

- อะไรคือผลลัพธ์ที่เป็นวัตถุประสงค์รองของการวิจัย?
- ผลลัพธ์ที่จะสรุปเป็นหลักนี้มี ความหมายทางคลินิกหรือไม่?
- เราสามารถวัดผลลัพธ์หลักนี้ได้จากผู้เข้าร่วมวิจัยทุกคนจริงหรือไม่?
- เราจะวัดค่าผลลัพธ์หลักกี่ครั้ง บ่อยแค่ไหน?
- ผลลัพธ์หลักนี้เป็นเพียงตัวแทนของผลลัพธ์ที่ไม่อาจวัดได้โดยตรง (surrogate outcomes) หรือไม่? ถ้าใช่มีเหตุผลใดที่เลือกใช้ผลลัพธ์ตัวนี้ ค่าของผลลัพธ์นี้เป็นตัวแทนที่เหมาะสมจริงหรือไม่? การแปรผลจากค่าตัวแทนค่าผลลัพธ์จริงจะก่อให้เกิดประโยชน์หรือโทษต่อค่าจริงที่ไม่อาจวัดได้โดยตรงในช่วงเวลาที่ศึกษานั้นๆ?

4. มีตัวแปรหรือปัจจัยอื่นๆ ที่ต้องมีการควบคุมในการดำเนินการวิจัยหรือไม่?

5. หน่วยของการสุ่มจำแนกกลุ่มตัวอย่าง เป็นกลุ่มย่อยเพื่อการทดสอบสมมติฐาน (randomization) คืออะไร? เป็นหน่วยที่เป็นตัวบุคคล ครอบครัว โรงพยาบาล ชุมชน หรือหน่วยอื่นๆ?

6. หน่วยที่จะใช้วิเคราะห์ข้อมูลในตอนสิ้นสุดโครงการเป็นหน่วยเดียวกับหน่วยที่ใช้

ในการจัดเก็บข้อมูลหรือไม่? หน่วยทั้งสองนี้อาจต่างกันก็ได้ เช่น หน่วยจัดเก็บข้อมูลอาจเป็นระดับครอบครัว แต่หน่วยวิเคราะห์ข้อมูลอาจเป็นระดับบุคคล และที่สำคัญคือจะใช้สถิติอะไรในการวิเคราะห์ผลการศึกษาค้างนี้

7. รูปแบบของการวิจัยมีลักษณะเป็นแบบใด? รูปแบบต่างกันจะมีสูตรคำนวณที่แตกต่างกันออกไป ดังนั้นต้องตีโจทย์ให้แตกว่างานวิจัยนั้นๆ มีรูปแบบเป็นแบบใด เช่น Randomized Controlled Trial หรือ RCT), cluster randomized trial, equivalence trial, non-randomized intervention study, repeated-measures design study ดังได้กล่าวแล้วในบทสรุปของการวิจัย

8. ผู้เข้าร่วมการวิจัยในโครงการนี้คือใคร? ประชากรเป้าหมายของการศึกษาค้างนี้เป็นใคร? มีเกณฑ์คัดเลือกเข้าสู่โครงการและคัดออกจากการเข้าร่วม (inclusion/exclusion criteria) อย่างไร? คาดว่าอัตราการคงอยู่ครบตลอดโครงการ (retention/follow-up rate) อัตราการถอนตัวจากโครงการ (drop-out rate) หรืออัตราการยอมทำตามกระบวนการวิจัยครบถ้วน (compliance rate) จะเป็นเท่าใด? ในการคำนวณขนาดกลุ่มตัวอย่างจะต้องมีการปรับค่านี้เข้าไปในภายหลัง ซึ่งจะได้กล่าวในตอนต่อไป

9. ขนาดของกลุ่มตัวอย่างที่อาจจะต้องคัดกรองเข้ามาสู่โครงการวิจัย (screening)

ต้องมีมากน้อยเท่าใด นั่นคือ จะมีจำนวนที่ไม่ผ่านเกณฑ์การคัดเลือกมากน้อยเพียงใด? จะต้องใช้วิธีการ/การทดสอบใดในการคัดกรองอาสาสมัคร/ผู้ป่วยเข้าโครงการหรือไม่อย่างไร?

10. มีโอกาสที่จะได้ค่าผลลัพธ์ที่ตั้งเป้าไว้จริงหรือไม่? ควรจะต้องมีการติดตามวัดผลซ้ำหรือไม่/หรือจะต้องวัดกี่ครั้ง? จึงจะสามารถบอกผลที่มีความหมายทางคลินิกได้?

หลังจากตอบคำถามข้างต้นได้แล้ว ผู้วิจัยต้องมีค่าสถิติซึ่งจะนำมาใส่ในสูตรคำนวณหาขนาดของกลุ่มตัวอย่างในการทดสอบสมมติฐานการวิจัยทางคลินิก 4 ค่าคือ

- ค่าข้อมูลพื้นฐานอันเป็นผลลัพธ์หลักของการวิจัยค้างนี้ (priori information)
- ค่าผลต่างที่คาดว่าจะพบและ/หรือที่มีความหมายทางคลินิก (effect size)
- ค่าระดับความเชื่อมั่น/นัยสำคัญทางสถิติ (significance level of test)
- ค่าอำนาจจำแนกความแตกต่างระหว่างกลุ่ม (power of test)

1. ค่าข้อมูลพื้นฐานของผลลัพธ์หลัก

ค่าสถิติที่สำคัญที่สุดที่ต้องใช้ในการคำนวณขนาดตัวอย่าง คือ ค่าข้อมูลพื้นฐานเกี่ยวกับผลลัพธ์หลักของการศึกษาค้างนี้ ตัวอย่างเช่น ในการวิจัยทางคลินิกเพื่อเปรียบเทียบ

ประเมินผลของยาตัวใหม่กับยามาตรฐานเดิมที่ใช้อยู่ ผู้วิจัยจะต้องมีคำจำกัดความที่ชัดเจนว่าประสิทธิภาพในที่นี้หมายถึงอะไร ถ้าหมายถึงอัตราการรักษาหายขาดจากโรค ผู้วิจัยก็ต้องมีตัวเลขอัตราการหายขาดจากโรคเมื่อใช้ยามาตรฐานที่ใช้กันอยู่ว่ามีค่าเท่าใด (เช่น cure rate ของยา A = ร้อยละ 70) หรือในการวิจัยทางคลินิกเพื่อศึกษาประสิทธิภาพของวัคซีนป้องกันการติดเชื้อเอชไอวีในกลุ่มผู้ติดเชื้อเสฟติด ซึ่งจะเป็นการเปรียบเทียบอัตราการติดเชื้อในกลุ่มผู้ติดเชื้อเสฟติดที่สู่มั้ได้วัคซีนทดลองกับกลุ่มผู้ติดเชื้อเสฟติดที่สู่มั้ได้สารเลียนแบบวัคซีน (placebo) ผู้วิจัยก็ต้องมีคำข้อมูลอัตราการติดเชื้อในกลุ่มผู้ติดเชื้อเสฟติดเป็นพื้นฐาน โดยถือเสมือนกับว่าอัตรานี้คือ อัตราที่จะพบในกลุ่มที่สู่มั้ได้สารเลียนแบบวัคซีน ซึ่งหมายถึง กรณีที่ไม่มีวัคซีนอยู่ หรือวัคซีนไม่ได้ผลนั่นเอง เช่น อัตราการติดเชื้อในกลุ่มผู้ติดเชื้อเสฟติดเท่ากับร้อยละ 8

คำข้อมูลพื้นฐานนี้ ผู้วิจัยจะต้องระบุที่มาว่าได้จากแหล่งข้อมูลใด อาจได้มาจากการทบทวนวรรณกรรม รายงานของโรงพยาบาล การสำรวจ/ศึกษาขนาดเล็กเพื่อหาคำข้อมูลดังกล่าวล่วงหน้า (pilot study) หรือจากการสอบถามผู้เชี่ยวชาญในวงการ (expert opinion) อย่างไรก็ตาม ในโครงการวิจัยต้องบอกที่มาของแหล่งข้อมูลนี้ให้ชัดเจน และสามารถตรวจสอบได้ แม้ว่าอาจจะเป็นเอกสารที่มีได้มีการตีพิมพ์เผยแพร่ทั่วไปก็ตาม

โดยทฤษฎีอาจดูเหมือนง่าย แต่ในทางปฏิบัติผู้วิจัยอาจเผชิญกับปัญหาที่ยุ่งยากในการเลือกค่าข้อมูลพื้นฐานตั้งต้นดังกล่าว ตัวอย่างเช่น ในการวางแผนการศึกษาผลของการให้โปรแกรมเพื่อลดอัตราการติดเชื้อ (Bacterial Vaginosis หรือ BV) คณะผู้วิจัยทำการทบทวนวรรณกรรมจากหลายแหล่งข้อมูล และพบว่าอัตราการติดเชื้อ BV แตกต่างกันมากมายในรายงานจากต่างที่กัน ซึ่งมีค่าตั้งแต่ร้อยละ 18 ไปจนถึงกว่าร้อยละ 30 คณะผู้วิจัยได้ศึกษาสำรวจขนาดเล็กในกลุ่มประชากรเป้าหมายของตนเอง พบว่ามีอัตราการติดเชื้ออยู่ที่ร้อยละ 10 เนื่องจากผลการศึกษานี้มาจากการสำรวจขนาดเล็ก ผู้วิจัยจึงไม่ใช้ร้อยละ 10 เป็นข้อมูลพื้นฐาน และได้ตัดสินใจเลือกเอาร้อยละ 15 เป็นค่าข้อมูลพื้นฐานสำหรับอัตราการติดเชื้อ BV กล่าวคือตัดสินใจเลือกเองระหว่างค่าร้อยละ 10 ที่สำรวจเองกับร้อยละ 18 ซึ่งเป็นค่าต่ำสุดจากรายงานอื่นๆ ทั้งนี้ในการตัดสินใจเลือกค่าตั้งต้นนี้ควรเป็นการเลือกที่มีการแสดงที่มาของแหล่งข้อมูลต่างๆ และแสดงเหตุผลของการเลือกกว่าเป็นการเลือกที่พยายามลดอคติ (bias) ที่จะใช้กลุ่มตัวอย่างขนาดเล็กเกินไปจนทำให้ไม่ได้คำตอบ แต่ไม่ต้องการใช้กลุ่มตัวอย่างที่ใหญ่เกินไปซึ่งเสี่ยงเกินไปสำหรับการเข้าสู่การวิจัยที่ยังไม่ทราบว่าจะมีผลดี/ร้ายอย่างไร ดังนั้น จึงจำเป็นต้องเลือกขนาดประชากรที่พอเหมาะเพื่อให้ได้คำตอบที่ต้องการ (conservative approach) ในกรณีนี้พบว่า

ยิ่งค่าตั้งต้นต่ำลง ขนาดกลุ่มตัวอย่างก็จะเพิ่มมากขึ้น หากใช้ค่าตั้งต้นของอัตราการติดเชื้อที่ร้อยละ 30 ขนาดกลุ่มตัวอย่างก็จะเล็กลงกว่าที่จะคำนวณได้เมื่อใช้ค่าตั้งต้นที่ร้อยละ 15 ดังกล่าว

2. ค่าผลต่างที่คาดว่าจะพบ/ต้องการจากค่าข้อมูลพื้นฐานอันเป็นค่าตั้งต้นเดิม
ผู้วิจัยจะต้องกำหนดค่าสถิติอีกค่าหนึ่งเพื่อคำนวณหาขนาดตัวอย่าง ค่าสถิตินี้จะเป็นตัวบ่งชี้ขนาดของผลต่างที่จะเกิดขึ้นจากการเปรียบเทียบระหว่างการได้หรือหรือไม่ได้รับปัจจัยที่จะศึกษา เรียกว่าเป็นค่าของขนาดผลต่างที่มีความหมาย ตัวอย่างเช่น กรณีของการเปรียบเทียบประสิทธิผลของยาใหม่เมื่อเทียบกับยามาตรฐาน หากอัตราการรักษาหายของยาเดิมคือร้อยละ 70 ผู้วิจัยอาจตั้งเป้าหมายว่ายาใหม่จะดีกว่ายาเก่าเมื่ออัตราการรักษาหายเพิ่มเป็นร้อยละ 80 นั่นคือขนาดผลต่างที่มีความหมายทางคลินิก ในที่นี้คือร้อยละ 10 หรือในกรณีของการทดสอบประสิทธิผลของวัคซีนป้องกันการติดเชื้อเอชไอวี (HIV) หากอัตราการติดเชื้อในกลุ่มที่ได้สารเลียนแบบ (เท่ากับอัตราการติดเชื้อในกลุ่มผู้ติดยาเสพติดทั่วไป) คือร้อยละ 8 และผู้วิจัยตั้งเป้าหมายไว้ว่าวัคซีนจะลดอัตราการติดเชื้อลงได้ร้อยละ 50 นั่นคือ คาดว่าอัตราการติดเชื้อในกลุ่มผู้ได้รับวัคซีนทดลองจะมีค่าเท่ากับร้อยละ 4 ผลต่างที่มีความหมายในที่นี้คือการลดลงจากร้อยละ 8 เป็นร้อยละ 4

ค่าขนาดผลต่างที่มีความหมายนี้ ไม่ใช่เรื่องการมีนัยสำคัญทางสถิติ เมื่อวิเคราะห์ข้อมูลเปรียบเทียบความแตกต่างด้วยวิธีการทดสอบทางสถิติ หากแต่หมายถึงผลต่างที่พบหรือคาดว่าจะพบนั้นมีความหมายทางคลินิกหรือสาธารณสุขหรือไม่ กล่าวคือถ้ายาตัวใหม่ดีกว่ายาเดิมร้อยละ 10 จึงจะถือว่ามีความหมายในแงุ่มทางคลินิกที่แพทย์จะพิจารณานำยาใหม่มาใช้แทนยาเก่า หรือวัคซีนลดอัตราการติดเชื้อได้ร้อยละ 50 จึงจะสามารถขึ้นทะเบียนยาได้ เป็นต้น ดังนั้นค่าว่าผลต่างที่มีความหมายจึงอาจหมายถึงค่าผลต่างที่มีความสำคัญทางคลินิก ค่าผลต่างที่ผู้วิจัยคิดว่ามีคุณค่ามากพอที่จะพยายามทำให้ได้ หรือค่าผลต่างที่ผู้วิจัยคาดว่าจะพบ

ขนาดกลุ่มตัวอย่างที่คำนวณได้นี้ จะแปรผกผันกับขนาดผลต่างที่มีความหมาย เพราะในทุกสูตรคำนวณค่านี้จะเป็นตัวหาร ดังนั้นถ้าขนาดผลต่างน้อย กลุ่มตัวอย่างก็จะมีขนาดใหญ่ แต่ถ้าขนาดผลต่างมากก็จะเป็นตัวคูณได้ขนาดกลุ่มตัวอย่างเล็กลง จึงต้องตั้งขนาดของผลต่างที่เล็กที่สุดเท่าที่จะทำให้ได้กลุ่มตัวอย่างขนาดพอเหมาะ ที่จะทำให้คำตอบจากการทดสอบสมมติฐานมีความน่าเชื่อถือว่าผลการวิเคราะห์ความแตกต่างนั้น เกิดจากความแตกต่างจากปัจจัยที่ใส่ลงไปจริง ไม่ใช่ด้วยเหตุบังเอิญ

การกำหนดค่าต่ำสุดของขนาดผลต่างที่มีความหมายนั้น อาจเขียนได้หลายรูปแบบ

ดังแสดงในตัวอย่างข้างต้นเป็นการเสนอขนาดผลต่างของ 2 สัดส่วน (เช่น อัตราการรักษาหายร้อยละ 70 ต่อร้อยละ 80) หรืออาจเป็นการระบุการลดอัตราลงก็เท่า (เช่น ลดอัตราการติดเชื้อลงร้อยละ 50 จากร้อยละ 8 เหลือร้อยละ 4) นอกจากนั้นอาจจะมีการระบุขนาดผลต่างด้วยค่าสถิติอื่นๆ เช่น ค่าอัตราสัดส่วนต่อการเกิด/ไม่เกิดผลลัพธ์ระหว่าง 2 กลุ่ม (Odd Ratio หรือ OR) อัตราความเสี่ยงสัมพัทธ์ (Relative Risk หรือ RR) อัตราการผลต่างความเสี่ยง (Risk Difference หรือ RD) ตัวอย่าง การระบุขนาดผลต่างที่มีความหมายเป็น RR ของโครงการทดสอบประสิทธิผลของวัคซีน อาจเขียนว่า ผู้วิจัยมีความต้องการประสิทธิผลของวัคซีน (Vaccine Efficacy หรือ VE) เท่ากับร้อยละ 50 ($VE = 1 - RR$; $RR = 0.5$) ดังนั้นหากอัตราการติดเชื้อในกลุ่มที่ได้สารเลียนแบบเท่ากับร้อยละ 8 ($risk = infection\ rate = 0.8$) อัตราการติดเชื้อของกลุ่มที่ได้วัคซีนจึงคาดว่าจะเท่ากับร้อยละ 4 ($RR = Risk\ กลุ่ม\ 2 / Risk\ กลุ่ม\ 1$; $Risk\ กลุ่ม\ 2 = 0.8 \times 0.5 = 0.4$)

หากผู้วิจัยต้องการทดสอบว่ายามีใหม่สามารถเพิ่มอัตราการอยู่รอดได้เท่ากับ 1.5 เท่าเมื่อเทียบกับอัตราการอยู่รอดของยาที่ใช้อยู่ปัจจุบัน ($RR = 1.5$) ดังนั้น ถ้าหากอัตราการอยู่รอดของยาที่ใช้กันอยู่เท่ากับร้อยละ 30 หมายความว่าผู้วิจัยคาดว่าอัตราการอยู่รอดของยาตัวใหม่เป็นร้อยละ 45 ($RR =$

$Risk\ กลุ่ม\ 2 / Risk\ กลุ่ม\ 1$; $Risk\ กลุ่ม\ 2 = 0.3 \times 1.5 = 0.45$)

อนึ่ง การกำหนดขนาดผลต่างที่มีความหมายทางคลินิกนี้ ควรเป็นการกำหนดอย่างมีหลักการไม่ใช่ตั้งขึ้นมาลอยๆ คณะผู้วิจัยต้องพิจารณาว่าขนาดผลต่างตรงไหนที่มีความหมายแท้จริง ในการวิจัยหลายๆ โครงการคณะผู้วิจัยใช้วิธีการประชุมระดมความคิดจากผู้เชี่ยวชาญในวงการ และกำหนดระดับขนาดที่คาดหวังและมีประโยชน์ในวงการ อย่างไรก็ตามในหลายกรณีก็จะมีกำหนดมาตรฐานขั้นต่ำเอาไว้ โดยบางองค์กรที่มีอำนาจหน้าที่ควบคุมดูแล เช่น การทดสอบประสิทธิผลของยาหรือวัคซีน องค์กรกำกับดูแลอาหารและยา (เช่น FDA ของสหรัฐอเมริกา) จะกำหนดขนาดผลต่างขั้นต่ำเอาไว้ว่าต้องอยู่ที่ระดับใดจึงจะใช้ขึ้นยาได้

3. ค่าระดับความเชื่อมั่นและนัยสำคัญทางสถิติ ผู้วิจัยจะต้องกำหนดค่าระดับความเชื่อมั่นในการยอมรับสมมติฐานที่จะทดสอบว่าเป็นสมมติฐานที่แท้จริง หากสามารถทำการวิจัยได้ในประชากร ค่าระดับความเชื่อมั่นนี้จะมีความสัมพันธ์กับการตั้งระดับนัยสำคัญทางสถิติ ซึ่งหมายถึงโอกาสที่จะปฏิเสธสมมติฐานที่จะทดสอบ (H_0) และในทางกลับกัน ค่าระดับนัยสำคัญที่ตั้งขึ้นเพื่อจะปฏิเสธสมมติฐานนี้ ยังหมายถึงโอกาสที่อาจจะปฏิเสธอย่างผิดๆ ก็ได้ ซึ่งมักเรียกกันว่าเป็น “โอกาสผิดแบบที่ 1” (Type I error หรือ

α -possibility) โดยปกติแล้ว นักวิจัยมักจะตั้งระดับนัยสำคัญทางสถิติที่จะใช้ในการปฏิเสธสมมติฐานที่ระดับร้อยละ 5 ($\alpha = 0.05$) นั่นคือ ระดับความเชื่อมั่นจะอยู่ที่ร้อยละ 95 อย่างไรก็ตาม ไม่จำเป็นต้องตั้งที่ระดับร้อยละ 5 เสมอไป ขึ้นอยู่กับผู้วิจัยว่ามีเป้าหมายจะตรวจสอบเรื่องที่มีความละเอียดมากน้อยแค่ไหน หากเป็นการเปรียบเทียบเพื่อสรุปผลในเรื่องที่มีผลกระทบสูง (sensitive) อาจตั้งค่าระดับนัยสำคัญที่จะปฏิเสธสมมติฐานที่ระดับร้อยละ 1 ($\alpha = 0.01$) หรือมีความเชื่อมั่นที่ระดับร้อยละ 99 ก็ได้ ทั้งนี้หากกำหนดค่าผิดพลาดคลาดเคลื่อนต่ำลง และค่าความเชื่อมั่นสูงขึ้น ก็จะทำให้มีขนาดของกลุ่มตัวอย่างที่จะคำนวณ ได้มีขนาดใหญ่ขึ้น

การกำหนดค่าระดับความเชื่อมั่นและนัยสำคัญทางสถิตินี้ ยังมีส่วนเกี่ยวข้องกับ การกำหนดวิธีการที่จะตั้งสมมติฐานและทดสอบสมมติฐานของผู้วิจัยอีกด้วย กล่าวคือการตั้งสมมติฐานที่จะทดสอบอาจมีได้ 2 แบบ คือ การตั้งสมมติฐานเพื่อทดสอบแบบทิศทางเดียว (one-sided test of significance) และการตั้งสมมติฐานเพื่อจะทดสอบแบบ 2 ทิศทาง (two-sided test of significance) สำหรับการตั้งสมมติฐานแบบ 2 ทิศทาง ผู้วิจัยจะตั้งสมมติฐานที่จะทดสอบว่าปัจจัยที่จะศึกษาไม่ส่งผลก่อให้เกิดความแตกต่าง ดังนั้นสมมติฐานรอบรับ (H_a) จึงมีลักษณะว่าปัจจัยที่ศึกษาอาจส่งผลทางใด

ก็ได้ นั่นคือผลลัพธ์ที่ได้จากกลุ่มทดลองที่ได้ปัจจัยที่ศึกษาอาจจะดีกว่าหรือแย่กว่ากลุ่มควบคุมที่ไม่ได้ปัจจัยศึกษา ส่วนการตั้งสมมติฐานเพื่อทดสอบแบบทิศทางเดียว เป็นการตั้งโดยกำหนดทิศทางชัดเจนในทางใดทางหนึ่ง เช่น ผลลัพธ์ที่ได้จากกลุ่มที่ได้ปัจจัยศึกษาจะดีกว่ากลุ่มที่ไม่ได้รับปัจจัยศึกษา การตั้งสมมติฐานแบบทิศทางเดียวนี้นี้ จะทำให้ผู้วิจัยมีโอกาสปฏิเสธสมมติฐานที่จะทดสอบได้ง่ายกว่า เพราะค่าของจุดตัดที่ระดับนัยสำคัญที่กำหนดจะต่ำกว่าค่าจุดตัดที่ระดับนัยสำคัญที่ระดับเดียวกันแบบ 2 ทิศทาง ดังนั้น งานวิจัยส่วนใหญ่จึงมักจะตั้งและทดสอบสมมติฐานแบบ 2 ทิศทาง ยกเว้นว่าจะมีเหตุผลและหลักฐานมากพอที่จะตั้งแบบมาตรฐานเป็นแบบทิศทางเดียว การตั้งสมมติฐานแบบทิศทางเดียว อาจจะเหมาะสมเมื่อมีข้อเสนอนที่น่าเชื่อถือได้ว่าผลลัพธ์จะเป็นไปในแบบ 2 ทิศทางได้ยาก ซึ่งอาจจะเป็นไปได้ในกรณีศึกษาด้านพิษวิทยา การประเมินความปลอดภัย การประเมินผลข้างเคียงอันเป็นปฏิกิริยาจากยา หรือการศึกษาในเรื่องภาวะเสี่ยง เป็นต้น

4. ค่าอำนาจจำแนกความแตกต่างระหว่างกลุ่มในการทดสอบสมมติฐาน ผู้วิจัยต้องตัดสินใจจากข้อมูลที่รวบรวมได้จากกลุ่มตัวอย่างว่าจะปฏิเสธสมมติฐานที่จะทดสอบ หรือไม่ปฏิเสธสมมติฐานดังกล่าว ทั้งนี้สมมติฐาน H_0 อาจจะเป็นจริงหรือไม่จริงก็ได้ หากสามารถศึกษาวิจัยได้ในประชากรจริง

ตารางที่ 2 ค่าอำนาจจำแนกความแตกต่างระหว่างกลุ่มในการทดสอบสมมติฐาน

	ปฏิเสธ H_0	ไม่ปฏิเสธ H_0
H_0 เป็นจริง	ตัดสินใจผิดพลาดแบบที่ 1 (probability = α)	ตัดสินใจถูก (confidence level = $1 - \alpha$)
H_0 เป็นเท็จ	ตัดสินใจถูก (power = $1 - \beta$)	ตัดสินใจผิดพลาดแบบที่ 2 (probability = β)

(ดูตารางที่ 2) ผู้วิจัยอาจจะตัดสินใจ ถูกได้ 2 กรณี คือปฏิเสธ H_0 ในกรณีที่ H_0 เป็นเท็จ และไม่ปฏิเสธ H_0 ในกรณีที่ H_0 เป็นจริง ในทางตรงกันข้ามผู้วิจัยอาจจะตัดสินใจผิดได้ 2 กรณี คือ ปฏิเสธสมมติฐานที่เป็นจริง (เรียกว่าเป็นการตัดสินใจผิดพลาดแบบที่ 1 คือ Type I error หรือ α) และไม่ปฏิเสธสมมติฐานที่เป็นเท็จ (เรียกว่าเป็นการตัดสินใจผิดพลาดแบบที่ 2 คือ Type II error - β)

การตัดสินใจถูกในกรณีไม่ปฏิเสธสมมติฐาน H_0 ที่เป็นจริง หมายถึงการสรุปผลว่าปัจจัยที่ศึกษาไม่ก่อให้เกิดความแตกต่างที่มีความหมายทางคลินิกระหว่างผลลัพธ์ที่ได้จากกลุ่มตัวอย่างที่นำมาเปรียบเทียบกัน และจะเป็นจริงด้วยหากได้ศึกษาในประชากร ซึ่งก็จะเป็นมุมกลับของการตัดสินใจผิดพลาดแบบที่ 1 หากผู้วิจัยกำหนดค่าโอกาสที่จะยอมให้มีความผิดพลาดแบบที่ 1 ไว้ร้อยละ 5 ($\alpha = 0.05$) ผู้วิจัยก็มีโอกาสตัดสินใจถูกในกรณีนี้ร้อยละ 95 ($1 - \alpha = 0.95$) นั่นคือผู้วิจัยจะมีระดับความเชื่อมั่น ร้อยละ 95 ที่จะบอก

ว่าผลลัพธ์ซึ่งได้จากกลุ่มเปรียบเทียบไม่แตกต่างกัน

การตัดสินใจถูกในกรณีปฏิเสธสมมติฐาน H_0 ที่เป็นเท็จ หมายถึงการสรุปผลว่าปัจจัยที่ศึกษาก่อให้เกิดความแตกต่างจริง และมีความหมายทางคลินิกระหว่างผลลัพธ์ที่ได้จากกลุ่มตัวอย่างที่นำมาเปรียบเทียบกัน และจะเป็นจริงหากได้ศึกษาในประชากร ซึ่งการตัดสินใจถูกนี้จะเป็นมุมกลับกับการตัดสินใจผิดพลาดแบบที่ 2 หากผู้วิจัยกำหนดค่าโอกาสที่จะยอมให้ตัดสินใจผิดพลาดแบบที่ 2 เท่ากับร้อยละ 10 ($\alpha = 0.10$) ผู้วิจัยก็จะมีโอกาสตัดสินใจถูกในกรณีเช่นนี้ร้อยละ 90 ($\beta = 0.90$) นั่นคือผู้วิจัยมีอำนาจ (power) ร้อยละ 90 ที่จะบอกว่าผลลัพธ์ที่ได้จากกลุ่มตัวอย่างที่นำมาเปรียบเทียบกันนี้แตกต่างกัน หรืออาจกล่าวได้ว่า การทดสอบนี้มีอำนาจจำแนกความแตกต่างได้อย่างถูกต้องที่ระดับร้อยละ 90

โดยปกติแล้ว ในการคำนวณขนาดตัวอย่างที่จะนำมาทดสอบสมมติฐาน มักจะกำหนดค่าระดับนัยสำคัญทางสถิติเพื่อ

ตัดสินใจปฏิเสธหรือไม่ปฏิเสธสมมติฐานที่ระดับร้อยละ 5 ดังได้กล่าวแล้ว และมักจะกำหนดค่าอำนาจจำแนกความแตกต่างที่ระดับร้อยละ 80 หรือร้อยละ 90 ทั้งนี้หากกำหนดค่าอำนาจจำแนกสูงขึ้น ขนาดของกลุ่มตัวอย่างก็จะเพิ่มมากขึ้นด้วย

สูตรคำนวณขนาดกลุ่มตัวอย่างในการวิจัยทางคลินิก

ทุกวันนี้จะมีโปรแกรมสำเร็จรูปช่วยคำนวณขนาดตัวอย่างที่จะใช้ในการวิจัยมาก แต่การใช้โปรแกรมคำนวณขนาดกลุ่มตัวอย่างนั้น ผู้วิจัยต้องกำหนดค่าสถิติพื้นฐานต่างๆ ดังกล่าวมาแล้ว สามารถหาได้ในหนังสือตำราสถิติ และบทความทางสถิติทั่วไป ดังปรากฏในตัวอย่างของเอกสารอ้างอิงท้ายบท [1-16] ในที่นี้เป็นการรวบรวมมิได้สรุปโดย Thabane L. [17] สูตรคำนวณขนาดตัวอย่าง (ดูตารางที่ 3 และตารางที่ 4) เป็นสูตรมาตรฐานที่จะพบทั่วไปในการวิจัยทางคลินิก ซึ่งเป็นการเปรียบเทียบปัจจัยที่ศึกษาในกลุ่มตัวอย่างในการศึกษาแบบกลุ่มเดียว แบบ 2 กลุ่ม หรือแบบไขว้กลุ่ม การวิจัยแต่ละรูปแบบจะมีการตั้งสมมติฐานและสูตรคำนวณต่างกันออกไป

ในกรณีที่ผลลัพธ์หลักที่สนใจจะศึกษามีเป็นค่าข้อมูลแบบค่าต่อเนื่อง ค่าข้อมูลพื้นฐานของตัวแปรผลลัพธ์ที่จะต้องหามาจากการทบทวนวรรณกรรมหรือการสำรวจล่วงหน้า คือค่าส่วนเบี่ยงเบนมาตรฐาน หรือ

ค่าการกระจายตัวของผลลัพธ์ (σ) ในกลุ่มที่ไม่ได้รับปัจจัยที่ศึกษา ค่าผลต่างที่มีความหมายทางคลินิก (δ) หมายถึงถึงค่าผลต่างระหว่างค่าเฉลี่ยของ 2 กลุ่มที่คาดว่าจะได้จากการศึกษา และจะทดสอบว่ามีความหมายทางคลินิกหรือไม่ ($\mu_1 - \mu_2$) คำนัยสำคัญทางสถิติ และค่าอำนาจจำแนกภายในสูตรคำนวณ คือค่าจุดตัดภายใต้โค้งปกติ หากกำหนดให้คำนัยสำคัญทางสถิติแบบทดสอบ 2 ทิศทางเท่ากับร้อยละ 5 ค่าจุดตัดภายใต้โค้งปกติ $Z_{\alpha/2} = 1.96$ และถ้าค่าอำนาจจำแนกเท่ากับร้อยละ 80 ค่าจุดตัดภายใต้โค้งปกติ $Z_\beta = 0.84$

ในกรณีที่ผลลัพธ์หลักที่สนใจจะศึกษาเป็นค่าข้อมูลแบบจำแนกแจกแจงค่า (เช่น หาย/ไม่หาย ติดเชื้อ/ไม่ติดเชื้อ) ค่าข้อมูลพื้นฐานของผลลัพธ์หลักที่จะต้องหามาจากการทบทวนวรรณกรรม หรือจากการวิจัยอื่นๆ ก่อนหน้านี้ คือค่าสัดส่วน (proportion) ค่าความชุก หรือค่าอัตราอุบัติการณ์ในกลุ่มที่ไม่ได้รับปัจจัยที่จะศึกษา (Π_1) ส่วนค่าผลต่างที่มีความหมายทางสถิติ (δ) หมายถึงถึงผลต่างของค่าสัดส่วนใน 2 กลุ่มที่จะได้จากการศึกษาและจะทดสอบว่ามีความหมายทางคลินิกหรือไม่ ($\Pi_1 - \Pi_2$) การกำหนดคำนัยสำคัญทางสถิติ และค่าอำนาจจำแนกในการทดสอบความสามารถระหว่างกลุ่ม จะใช้วิธีการเดียวกับที่ได้กล่าวแล้วข้างต้น

ตารางที่ 3 สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง (แบบที่ 1)

รูปแบบการวิจัย	สมมติฐาน	สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง		
		H_0	H_a	สูตรพื้นฐาน
กลุ่มเดียว (One-sample)	Equality	$\mu - \mu_0 = 0$	$\mu - \mu_0 \neq 0$	$n = \frac{\left[\frac{z_\alpha}{2} + z_\beta \right]^2 \sigma^2}{(\mu - \mu_0)^2}$
	Superiority	$\mu - \mu_0 \leq \delta$	$\mu - \mu_0 > \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$
	Equivalence	$ \mu - \mu_0 \geq \delta$	$ \mu - \mu_0 < \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu - \mu_0 - \delta)^2}$
สองกลุ่มคู่ขนาน (Two-sample parallel)	Equality	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{2 \left[\frac{z_\alpha}{2} + z_\beta \right]^2 \sigma^2}{(\mu_1 - \mu_2)^2}$
	Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
	Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
	Equivalence	$ \mu_1 - \mu_2 \geq \delta$	$ \mu_1 - \mu_2 < \delta$	$n_i = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_2 - \delta)^2}$
สองกลุ่มไขว้ (Two-sample crossover)	Equality	$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$n_i = \frac{\left[\frac{z_\alpha}{2} + z_\beta \right]^2 \sigma^2}{2(\mu_1 - \mu_2)^2}$
	Non-inferiority	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\mu_1 - \mu_2 - \delta)^2}$
	Superiority	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\mu_1 - \mu_2 - \delta)^2}$
	Equivalence	$ \mu_1 - \mu_2 \geq \delta$	$ \mu_1 - \mu_2 < \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{2(\mu_1 - \mu_2 - \delta)^2}$

ตารางที่ 4 สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง (แบบที่ 2)

รูปแบบการวิจัย	สมมติฐาน	สมมติฐานและสูตรคำนวณขนาดกลุ่มตัวอย่าง	
		H_0	สูตรพื้นฐาน
กลุ่มเดียว (One-sample)	Equality	$\pi - \pi_0 = 0$	$n = \frac{\left(\frac{z_\alpha}{2} + z_\beta\right)^2 \pi(1-\pi)}{(\pi - \pi_0)^2}$
	Superiority	$\pi - \pi_0 \leq \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \pi(1-\pi)}{(\pi - \pi_0 - \delta)^2}$
	Equivalence	$ \pi - \pi_0 \geq \delta$	$n = \frac{(z_\alpha + z_\beta)^2 \pi(1-\pi)}{(\pi - \pi_0 - \delta)^2}$
สองกลุ่มคู่ขนาน (Two-sample parallel)	Equality	$\pi_1 - \pi_2 = 0$	$n_i = \frac{\left(\frac{z_\alpha}{2} + z_\beta\right)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2)^2}$
	Non-inferiority	$\pi_1 - \pi_2 \geq \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2 - \delta)^2}$
	Superiority	$\pi_1 - \pi_2 \leq \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2 + \delta)^2}$
	Equivalence	$ \pi_1 - \pi_2 \geq \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2))}{(\pi_1 - \pi_2 - \delta)^2}$
สองกลุ่มไขว้ (Two-sample crossover)	Equality	$\pi_1 - \pi_2 = 0$	$n_i = \frac{\left(\frac{z_\alpha}{2} + z_\beta\right)^2 \sigma_d^2}{2(\pi_1 - \pi_2)^2}$
	Non-inferiority	$\pi_1 - \pi_2 \geq \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$
	Superiority	$\pi_1 - \pi_2 \leq \delta$	$n_i = \frac{(z_\alpha + z_\beta)^2 \sigma_d^2}{2(\pi_1 - \pi_2 + \delta)^2}$
	Equivalence	$ \pi_1 - \pi_2 \geq \delta$	$n_i = \frac{(z_\alpha + z_{\beta/2})^2 \sigma_d^2}{2(\pi_1 - \pi_2 - \delta)^2}$

ตัวอย่างที่ 1 การเปรียบเทียบอัตราการติดเชื้อของ 2 กลุ่ม

วัตถุประสงค์ ผู้วิจัยต้องการศึกษาประสิทธิผลของการให้ CESS (Colony Stimulating Factors) ในการลดอัตราการเกิดภาวะพิษเหตุติดเชื้อ (sepsis) ในทารกเกิดก่อนกำหนด มีการวิจัยก่อนหน้านี้รายงานว่าการให้ CESS ลดอัตราการเกิดภาวะพิษเหตุติดเชื้อในทารกกลุ่มนี้อยู่ที่ระดับร้อยละ 50 ผู้วิจัยตั้งเป้าหมายว่าหากลดอัตราการเกิดลงไปถึงระดับร้อยละ 34 จึงจะถือว่ามีความหมายทางคลินิก

การคำนวณ ในกรณีนี้เป็นการคำนวณเปรียบเทียบ 2 สัดส่วน สูตรที่ใช้ในการคำนวณค่ากลุ่มตัวอย่างสำหรับ 1 กลุ่ม คือ

$$n = \frac{[Z_{\alpha/2} + Z_{\beta}]^2 \times [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{(\pi_1 - \pi_2)^2}$$

โดยที่ n = ขนาดกลุ่มตัวอย่างของแต่ละกลุ่มที่ต้องการ (ต้องเพิ่มเป็น 2 เท่าสำหรับ 2 กลุ่ม รวมกัน)

π_1 = ค่าสัดส่วนที่ 1

π_2 = ค่าสัดส่วนที่ 2

$\pi_1 - \pi_2$ = ขนาดความแตกต่างที่มีนัยสำคัญทางคลินิก

$Z_{\alpha/2}$ = ค่าจุดตัดสินใจที่มีนัยสำคัญทางสถิติ ที่ความเชื่อมั่นร้อยละ 50 (อาจตัดสินใจผิดแบบที่ 1 = ร้อยละ 5)

Z_{β} = ค่าจุดตัดสินใจที่บ่งบอกความสามารถในการจำแนกความแตกต่างร้อยละ 80 (อาจตัดสินใจผิดแบบที่ 2 = ร้อยละ 20)

แทนค่าในสูตรจะได้คำตอบว่า :

$$n = \frac{[1.96 + 0.84]^2 \times [(0.50 \times 0.50) + (0.34 \times 0.66)]}{[0.16]^2} = 146$$

ดังนั้นจะต้องมีผู้เข้าร่วมการศึกษากลุ่มละ 146 คน รวม 2 กลุ่ม คือ 292 คน

ตัวอย่างที่ 2 การเปรียบเทียบค่าคะแนนของ 2 กลุ่ม

วัตถุประสงค์ ผู้วิจัยวางแผนจะประเมินผลการให้โปรแกรมทางจิตวิทยาแบบใหม่เมื่อเปรียบเทียบกับวิธีการรักษาแบบมาตรฐานทั่วไป เพื่อจะลดภาวะรู้สึกต้องการฆ่าตัวตายในกลุ่มผู้ป่วยที่กินยาพิษและเข้ารับการรักษาอยู่ภายในโรงพยาบาล การวัดภาวะรู้สึกต้องการฆ่าตัวตายวัดจากแบบสอบถามมาตรฐาน beck scale มีรายงานการวิจัยก่อนหน้านี้ว่าค่าคะแนนจากแบบสอบถาม beck scale มีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.7 ผู้วิจัยตั้งข้อกำหนดไว้ว่าหากผลต่างของค่าเฉลี่ยของคะแนนจาก 2 กลุ่มเท่ากับ 5 คะแนน จึงจะถือว่าโปรแกรมใหม่สามารถลดลงได้อย่างมีความหมายทางคลินิก ผู้วิจัยยังคาดหวังว่าอาจจะมีผู้อยู่ครบการศึกษาวิจัยเพียงแค่ 2 ใน 3 กล่าวคือ อาจมีผู้ถอนตัวจากโครงการออกไปในการศึกษาเสร็จสิ้นประมาณ 1 ใน 3

การคำนวณ ในกรณีนี้เป็นการคำนวณเปรียบเทียบค่าเฉลี่ย 2 ค่า สูตรที่ใช้คือ

$$n = \frac{\left[\frac{z_{\alpha}}{2} + z_{\beta} \right]^2 \times 2\sigma^2}{\delta^2}$$

โดยที่ n = ขนาดกลุ่มตัวอย่างของแต่ละกลุ่มที่ต้องการ

σ = ส่วนเบี่ยงเบนมาตรฐานของตัวแปรผลลัพธ์หลัก

δ = ขนาดความแตกต่างที่มีความสำคัญทางคลินิก

$$Z_{\frac{\alpha}{2}} = 1.96$$

$$Z_{\beta} = 0.84$$

แทนค่าในสูตรจะได้คำตอบว่า

$$n = \frac{[1.96 \times 0.84]^2 \times 2 \times 7.7^2}{5.0^2} = 38$$

ดังนั้น จะต้องต้องมีผู้เข้าร่วมโครงการวิจัย กลุ่มละ 38 คน รวม 2 กลุ่ม คือ 76 คน แต่เนื่องจากอาจมีผู้ถอนตัวออกไป 1 ใน 3 ดังนั้นจะต้องการผู้เข้าร่วมการวิจัยในตอนต้นประมาณ กลุ่มละ 60 คน รวม 2 กลุ่ม คือ 120 คน เพื่อให้ได้ยอดขนาดกลุ่มตัวอย่างในตอนจบโครงการเท่าจำนวนที่คาดหวังไว้ในตอนแรก (คือ 76 คน)

สูตรคำนวณขนาดกลุ่มตัวอย่างของรูปแบบการวิจัยแบบอื่นๆ

นอกจากสูตรคำนวณขนาดกลุ่มตัวอย่างมาตรฐานที่ได้กล่าวในตอนที่แล้ว ยังมีสูตรคำนวณที่แตกต่างออกไป ในกรณีที่ผู้วิจัยมีรูปแบบการวิจัยทางคลินิกแบบอื่นๆ อีกมากมาย ที่ผลลัพธ์ไม่ใช่แค่ค่าเฉลี่ยหรือค่าสัดส่วนอย่างทั่วไป ซึ่งอาจหาได้จากหนังสือสถิติหรือบทความต่างๆ ในเว็บ ในที่นี้จะนำเสนอเพียงบางสูตรที่พบบ่อยในการวิจัยทางคลินิก ทุกวันนี้เราคงไม่ต้องกังวลกับการคำนวณขนาดตัวอย่างด้วยตนเอง เพราะมีโปรแกรมสำเร็จรูปช่วยคำนวณอยู่มากมาย แต่ก็ควรรู้จักสูตรคำนวณเอาไว้บ้าง เพราะจะช่วยให้ผู้วิจัยรู้ว่าจะต้องมีข้อมูลพื้นฐานอะไรบ้างที่จะต้องหามาใส่เข้าไปในสูตรคำนวณ

1. การวิเคราะห์ค่าความเปลี่ยนแปลงของผลลัพธ์อันเป็นผลต่างจากค่าข้อมูลผลลัพธ์ที่จุดตั้งต้น (Baseline) ในการคำนวณค่าความเปลี่ยนแปลงของผลลัพธ์ที่เป็นผลต่างจากค่าข้อมูลผลลัพธ์ที่วัดครั้งสุดท้าย (หรือครั้งที่สนใจจะวิเคราะห์) กับค่าข้อมูลผลลัพธ์ที่วัดครั้งแรก หากกำหนดให้ค่าต่างๆ ที่จะใช้ในการคำนวณ คือ

Y = ตัวแปรผลลัพธ์ที่วัดติดตามผล (response variable)

Y_b = ค่าตัวแปรผลลัพธ์ที่จุดเริ่มต้น

μ = ค่าเฉลี่ยของการกระจายตัวของตัวแปรผลลัพธ์ที่วัดติดตามผล

μ_b = ค่าเฉลี่ยที่จุดเริ่มต้น

ρ = ค่าความสัมพันธ์ระหว่างค่าผลลัพธ์
ติดตามผลกับค่าที่วัดที่จุดเริ่มต้น

$d = Y - Y_b$ ผลต่างจากค่าที่วัดที่จุด
ตั้งต้น

$S = \mu - \mu_b$

σ = ส่วนเบี่ยงเบนมาตรฐานของค่า d

สูตรคำนวณขนาดตัวอย่างจากค่า d คือ

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 2\sigma^2 (1 - \rho)}{\delta^2}$$

ในการวิเคราะห์ความเปลี่ยนแปลงจากข้อมูลที่จุดตั้งต้น หากค่า $\rho > 0.5$ ก็จะทำให้ได้ขนาดกลุ่มตัวอย่างที่ไม่ใหญ่มากนัก ซึ่งเป็นข้อดีกว่าการคำนวณขนาดกลุ่มตัวอย่างโดยการเปรียบเทียบแบบ 2 กลุ่มคู่ขนานกันไป[17] สำหรับโครงการวิจัยทางคลินิกโดยส่วนใหญ่แล้วมักจะมีการเก็บรวมข้อมูลเพื่อประเมินผลลัพธ์สุดท้ายเป็นระยะๆ เป็นการคำนวณขนาดกลุ่มตัวอย่าง มักจะเป็นการคำนวณเพื่อการวิเคราะห์ข้อมูลครั้งสุดท้ายเท่านั้น โดยอาจจะเป็นการเปรียบเทียบกับค่าของตัวแปรผลลัพธ์ที่จุดตั้งต้นหรือไม่ก็ได้ แต่หากผู้วิจัยต้องการดำเนินการวิจัยและวิเคราะห์ข้อมูลแบบซ้ำๆ กัน (repeated measures) ในเชิงระบาดวิทยา สูตรคำนวณของกลุ่มตัวอย่างจะแตกต่างออกไป โดยอ่านรายละเอียดในเอกสารอ้างอิงท้ายบทของ Vickers AJ.[18]

2. การเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างที่มีค่าข้อมูลผลลัพธ์กระจายตัวแบบปัวซอง (Poisson populations) ในกรณีที่ต้องการคำนวณความเท่ากัน (equality) ของค่าเฉลี่ยของ 2 กลุ่มที่มาจากการประชากรที่มีลักษณะข้อมูลแบบนี้ หากกำหนดให้

θ_1 = ค่าเฉลี่ยในประชากรกลุ่มที่ 1

θ_2 = ค่าเฉลี่ยในประชากรกลุ่มที่ 2

สูตรคำนวณขนาดกลุ่มตัวอย่างของแต่ละกลุ่ม[17]จะมีค่าเท่ากับ

$$n = \frac{4}{(\sqrt{\theta_1} - \sqrt{\theta_2})^2}$$

3. การเปรียบเทียบความแตกต่างระหว่างกลุ่มที่ผลลัพธ์เป็นช่วงระยะเวลาการเกิดผลลัพธ์สุดท้าย (Survival time) ในปัจจุบันมีการวิจัยเป็นจำนวนมากที่สนใจผลลัพธ์เป็นเวลา เช่น ระยะเวลาตั้งแต่ให้ยาไปจนถึงแพ้หาย ระยะเวลาตั้งแต่ผ่าตัดไปจนผู้ป่วยฟื้นตัว ระยะเวลาการอยู่รอด โดยวิธีวิเคราะห์ข้อมูลที่ใช้เป็นการวิเคราะห์แบบการวิเคราะห์การอยู่รอด (survival analysis) ส่วนสูตรที่ใช้คำนวณขนาดกลุ่มตัวอย่างมีหลายสูตร ซึ่งรวบรวมโดย Southworth H. [19] ในสูตรต่างๆ สำหรับการวิเคราะห์ตัวแปรผลลัพธ์แบบช่วงเวลานี้จะใช้ค่าสถิติพื้นฐานต่อไปนี้

d เป็นจำนวนผู้ที่เกิดผลลัพธ์สุดท้าย (ตาย, หาย, อยู่รอด, ติดเชื้อ)

- s_1, s_2 เป็นโอกาสการอยู่รอด (survival probabilities) ในช่วงเวลาที่กำหนดของกลุ่มที่ 1 และกลุ่มที่ 2
- λ_1, λ_2 เป็นค่าการอยู่รอดในประชากรที่มีลักษณะเป็นไค้งเอกซ์โปเนนเชียล (exponential survival parameters) ของกลุ่มที่ 1 และกลุ่มที่ 2
- h ค่าอัตราส่วนการอยู่รอด (hazard ratio)
- θ เป็นค่าอัตราส่วนการอยู่รอดในหน่วยล็อก (log-hazard ratio)
- π_1, π_2 เป็นค่าสัดส่วนจำนวนกลุ่มตัวอย่างกลุ่ม 1 ต่อกลุ่ม 2 หากไม่กำหนดจะถือว่าเท่ากัน
- T เป็นระยะเวลาดำเนินการศึกษา (รวมทั้งเวลาการสรรหาอาสาสมัคร/ผู้ป่วย/ผู้เข้าร่วมวิจัยเข้าโครงการ (accrual period) ด้วย
- τ เป็นระยะเวลาการสรรหาอาสาสมัคร/ผู้ป่วยเข้าโครงการ
- δ เป็นค่าอัตราการถอนตัวจากโครงการที่มีลักษณะเป็นแบบเอกซ์โปเนนเชียล พารามิเตอร์ (exponential parameter)
- γ เป็นค่าอัตราการสรรหาผู้เข้าร่วมโครงการ (exponential parameter in accrual) โดยปกติถ้าไม่ระบุจะกำหนดให้เท่ากับ 0 คือมีค่าเป็นแบบสม่ำเสมอ

ตลอดกระบวนการ (uniform accrual)

สูตรที่ 1 การประเมินจำนวนการเกิดผลลัพธ์สุดท้ายนำเสนอโดย Collett D. [20], อ้างถึง [19])

$$d = \frac{(z_{\alpha/2} + z_\beta)^2}{\pi_1 \pi_2 \theta^2}$$

โดยที่ d = จำนวนการเกิดผลลัพธ์สุดท้าย (ตาย, หาย, ฯลฯ) ตลอดการศึกษา

ในกรณีที่ต้องการให้ขนาดกลุ่มตัวอย่าง 2 กลุ่มมีขนาดเท่ากัน สูตรคำนวณ คือ

$$d = \frac{4(z_{\alpha/2} + z_\beta)^2}{\theta^2}$$

สูตรที่ 2 การคำนวณขนาดกลุ่มตัวอย่างกรณีที่กำหนดว่าอัตราการอยู่รอดมีลักษณะเป็นแบบไค้งเอกซ์โปเนนเชียล และอัตราการถอนตัวจากโครงการเป็นแบบคงที่ (uniform) และกลุ่มตัวอย่างทั้ง 2 กลุ่มมีขนาดเท่ากันนำเสนอโดย Rubinstein LV. [21] และ Lakatos E. & Lan KKG. [22], (อ้างถึงใน [19]) สูตรคำนวณขนาดกลุ่มตัวอย่าง คือ

$$n = \left(\frac{z_{\alpha/2} + z_\beta}{\theta} \right)^2 \left(\frac{1}{E(P_1)} + \frac{1}{E(P_2)} \right)$$

โดยที่

$$E(P_i) = \frac{\lambda_i}{\lambda_i + \delta} \left(1 - \frac{e^{-(\lambda_i + \delta)(T - \tau)} - e^{-(\lambda_i + \delta)\tau}}{(\lambda_i + \delta)\tau} \right), i = 1, 2$$

สูตรที่ 3 การคำนวณขนาดกลุ่มตัวอย่าง ที่ผู้วิจัยกำหนดให้อัตราการอยู่รอด อัตราการถอนตัวจากโครงการ และอัตราการสรรหาผู้เข้าร่วมโครงการ เป็นแบบโค้งเอกซ์โปเนนเชียล โดยขนาดกลุ่มตัวอย่าง 2 กลุ่ม ไม่จำเป็นต้องเท่ากัน (π_1/π_2) และยินยอมให้มีการปรับเปลี่ยนปัจจัยที่ให้

(ยา/วิธีการให้ยา) ได้อีกด้วย และผู้เข้าร่วมโครงการอาจเลิกจากวิธีการรักษาแบบที่ 1 ในกลุ่มที่ 1 แล้วได้รับวิธีการรักษาในกลุ่มที่ 2 แทนที่ นั่นคือมี “non-compliance” ที่เรียกว่าเป็น “therapeutic crossover” (Ladin, 1986)

สูตรคำนวณขนาดตัวอย่างคือ

$$N = \left\{ \frac{1}{|\lambda_1 - \lambda_2|} \left(z_{\alpha/2} \sqrt{\phi(\pi_1 \lambda_1 + \pi_2 \lambda_2)(1/\pi_1 + 1/\pi_2)} + z_{\beta} \sqrt{\phi(\lambda_1)/\pi_1 + \phi(\lambda_2)/\pi_2} \right) \right\}^2$$

โดยที่

$$\phi(\bar{\lambda}) = \bar{\lambda}^2 \left(\frac{\bar{\lambda}}{\bar{\lambda} + \delta} + \frac{\bar{\lambda} \gamma e^{-(\bar{\lambda} + \delta)(T - \tau)} [1 - e^{(\bar{\lambda} + \delta - \gamma)\tau}]}{(1 - e^{-\gamma\tau})(\bar{\lambda} + \delta)(\bar{\lambda} + \delta - \gamma)} \right)$$

ค่า $\phi(\cdot)$ จะง่ายขึ้นหากกำหนดให้อัตราการสรรหา (accrual) เป็นแบบสม่ำเสมอตลอดโครงการ และไม่มีการถอนตัว (dropout) จากโครงการ

$$\phi(\bar{\lambda}) = \bar{\lambda}^2 \left[1 - \frac{e^{-\bar{\lambda}(T - \tau)} - e^{-\bar{\lambda}T}}{\lambda\tau} \right]^{-1}$$

อัตราการไม่ได้รับปัจจัยการศึกษาตามกำหนดมีการปรับเปลี่ยนวิธีการศึกษา (non-compliance) จะนำมาคิดเพื่อปรับยอดผู้เข้าร่วมโครงการทั้งหมด (N) ดังนี้

$$N \rightarrow \frac{N}{(1 - \omega_1 - \omega_2)^2}$$

โดย ω_1 เป็นสัดส่วนของผู้ที่ non-compliant ในกลุ่มที่ i ในที่นี้ non-compliance ก็คือ

การปรับเปลี่ยนจากวิธีการรักษาหนึ่งไปอีกรักษาหนึ่ง (drop-ins and drop-outs).

สูตรที่ 4 การคำนวณขนาดกลุ่มตัวอย่าง อย่างง่าย โดยผู้วิจัยกำหนดให้อัตราการอยู่รอดมีลักษณะเป็นแบบเอกซ์โปเนนเชียล อัตราการสรรหาผู้เข้าร่วมโครงการเป็นแบบสม่ำเสมอและไม่มีการถอนตัวจากโครงการ นำเสนอโดย Collett D. [20], (อ้างถึง [19]) สูตรคำนวณ คือ

$$n = \frac{d}{P(\text{death})}$$

โดย d เป็นจำนวนการเกิดผลลัพธ์สุดท้าย (ตาย/หาย/ติดเชื้อ)

$$P(\text{death}) = 1 - \frac{1}{6}(\bar{s}(T - \tau) + 4\bar{s}(T - \tau/2) + \bar{s}(T))$$

ทั้งนี้โอกาสการอยู่รอด (survival possibility) ที่ระยะเวลา t เป็นค่าเฉลี่ยระหว่าง 2 กลุ่มตัวอย่างดังนี้

$$\bar{s}(t) = \frac{s_1(t) + s_2(t)}{2}$$

สูตรที่ 5 การคำนวณขนาดกลุ่มตัวอย่าง
ในกรณีที่ผู้วิจัยกำหนดจะใช้วิธีการเปรียบเทียบอัตราการอยู่รอดด้วยวิธีการเปรียบเทียบระหว่างกลุ่ม โดยใช้สถิติเป็น “log-rank test” (Freedman, 1982) โดยที่สูตรคำนวณคือ

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 (h+1)^2}{(2 - s_1 - s_2)(h-1)^2}$$

สูตรที่ 6 การคำนวณขนาดกลุ่มตัวอย่าง
อย่างง่ายที่ผู้วิจัยสามารถกำหนดค่าเฉลี่ยของอัตราการอยู่ระหว่าง 2 กลุ่มเปรียบเทียบ นำเสนอโดย Thabane L. [17]

M_t = ค่าเฉลี่ย survival time ในกลุ่ม
ทดลอง

M_c = ค่าเฉลี่ย survival time ในกลุ่ม
ควบคุม

หากกำหนดว่า survival times มีการกระจาย
ตัวเป็นแบบ exponential สูตรคำนวณขนาด
กลุ่มตัวอย่างคือ

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2}{(\ln(M_t / M_c))^2}$$

**4. การเปรียบเทียบความแตกต่าง
ระหว่าง 2 กลุ่มที่เป็นการสุ่มแบบยกกลุ่ม
(clusters)** รูปแบบการวิจัยแบบสุ่มยกกลุ่ม
(cluster randomized design) นิยมใช้กัน
มากขึ้นในการวิจัยด้านสาธารณสุขชุมชน

ซึ่งใช้วิธีการสุ่มยกทั้งหน่วย (ครอบครัว/บ้าน/
กลุ่มบ้าน/หมู่บ้าน/ชุมชน) โดยที่หน่วยหนึ่ง
จะได้รับปัจจัยที่ศึกษา เช่น โปรแกรมสุขศึกษา
ป้องกัน และอีกหน่วยหนึ่งไม่ได้รับ การวิจัย
ลักษณะนี้อาจมีส่วนดีส่วนเสียในแง่ของ
การควบคุมการให้ปัจจัยที่ศึกษา การเสริม
หรือลดการทำตามเงื่อนไขต่างๆ ของโครงการ
(compliance/adherence) ความสะดวกใน
การบริหารจัดการโครงการเป็นต้น สูตร
คำนวณขนาดตัวอย่างในกรณี
นี้ขึ้นอยู่กับจำนวนหน่วยที่ใช้สุ่มและขนาด
ของกลุ่ม ดังนี้

k = จำนวน cluster

m = ขนาดเฉลี่ยของ cluster

ρ = ค่าสหสัมพันธ์ระหว่าง clusters

$IF = 1 + (m - 1) \rho$ = ปัจจัยเพิ่มขนาด
(inflation factor)

สูตรที่ 1 การทดสอบความเท่ากัน
(equality) ของค่าเฉลี่ย $\mu_1 - \mu_2 = 0$

$$n_1 = n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2 2\sigma^2 \times IF}{(\mu_1 - \mu_2)^2}$$

$$k = \frac{(z_{\alpha/2} + z_{\beta})^2 2\sigma^2 \times IF}{m(\mu_1 - \mu_2)^2}$$

สูตรที่ 2 การทดสอบความเท่ากันของ
สัดส่วน : $\pi_1 - \pi_2 = 0$

$$n_1 = n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)] \times IF}{(\pi_1 - \pi_2)^2}$$

$$k = \frac{(z_{\alpha/2} + z_{\beta})^2 [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)] \times IF}{m(\pi_1 - \pi_2)^2}$$

สูตรที่ 3 การทดสอบความเท่ากันของอุบัติการณ์ (incidence rates): $\lambda_1 = \lambda_2$

$$n_1 = n_2 = \frac{(z_{\alpha/2} + z_{\beta})^2 [\lambda_1 + \lambda_2] \times IF_t}{t(\lambda_1 - \lambda_2)^2}$$

$$IF_t = 1 + \frac{CV^2(\lambda_1^2 + \lambda_2^2)t}{(\lambda_1 + \lambda_2)}$$

$$CV = \frac{\sigma_1}{\lambda_1} = \frac{\sigma_2}{\lambda_2}$$

5. การศึกษาที่ค่าผลลัพธ์เป็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร 2 ตัว (Correlation coefficient) เป็นกรณีที่ผู้วิจัยสนใจจะศึกษาความสัมพันธ์ระหว่างตัวแปร (ρ) ว่ามีความหมายทางสถิติหรือไม่ สูตรคำนวณขนาดตัวอย่างนี้ยังใช้ในการศึกษาที่มีจุดมุ่งหมายจะทดสอบสมมติฐานว่าค่า slope = 0 ในสมการทำนายค่าความสัมพันธ์ระหว่างตัวแปร (regression) ได้อีกด้วย[17] สมมติฐานที่ต้องการจะทดสอบและขนาดตัวอย่างในที่นี้ คือ

$$H_0 : \rho = 0$$

$$H_a : \rho \neq 0$$

หากใช้การแปลงข้อมูลเป็นแบบ Fisher's arctanh (Z) transformation

$$Z = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right)$$

และข้อมูลมีลักษณะเป็นโค้งปกติ สูตรคำนวณขนาดกลุ่มตัวอย่าง คือ

$$n = 3 + \frac{2(Z_{\alpha/2} + Z_{\beta})^2}{\left(\ln \left(\frac{1 + \rho}{1 - \rho} \right) \right)^2}$$

ในกรณีที่ผู้วิจัยต้องการเปรียบเทียบค่าสหสัมพันธ์ระหว่างตัวแปรดังกล่าวระหว่างกลุ่มตัวอย่าง 2 กลุ่ม ว่ามีค่าสหสัมพันธ์เท่ากันหรือไม่ สูตรคำนวณขนาดตัวอย่าง คือ

$$H_0 : \rho_1 = \rho_2$$

$$H_a : \rho_1 \neq \rho_2$$

สูตรคำนวณขนาดกลุ่มตัวอย่างคือ

$$n = 3 + \frac{2(Z_{\alpha/2} + Z_{\beta})^2}{\left(\ln \left(\frac{1 + \rho_1}{1 - \rho_1} \right) - \ln \left(\frac{1 + \rho_2}{1 - \rho_2} \right) \right)^2}$$

6. การศึกษาที่มีวัตถุประสงค์หลักเกี่ยวกับความสัมพันธ์ระหว่างตัวแปร ในการวิเคราะห์สมการความสัมพันธ์ระหว่างตัวแปรผลลัพธ์กับตัวแปรอื่นๆ ที่เป็นตัวแปรทำนายค่า ค่าสถิติที่มีความสำคัญคือค่า R^2 ซึ่งเป็นค่าที่ใช้ในการบ่งบอกความแปรปรวนที่อธิบายได้ของเส้นสมการ ที่มีตัวแปรหลายตัวร่วมกันทำนาย (หรือมีความสัมพันธ์ร่วมกัน) ในการอธิบายค่าตัวแปรที่เป็นผลลัพธ์หลัก มีคำแนะนำจากนักสถิติหลายคนที่กำหนดขนาดตัวอย่างที่เหมาะสมสำหรับงานวิจัยในลักษณะนี้ ดังนี้

คำแนะนำที่ 1 Hair JF. [25] (อ้างถึงใน [17]) et.al เสนอว่าในกรณีสมการทำนายค่าจากหลายตัวแปร (multiple regression) ที่ต้องการค่าอำนาจจำแนก $\text{power} = \text{ร้อยละ } 80$ ที่ระดับนัยสำคัญ $= 0.05$ และต้องการประเมินค่า R^2 ว่า

ค่า $R^2 \geq 0.23$ จะได้จาก $n = 50$

ค่า $R^2 \geq 0.12$ จะได้จาก $n = 100$

หลักการสำคัญคือสัดส่วนของจำนวนตัวอย่าง (n) ต่อตัวแปรต้น (independent variable) ในสมการต่ำสุดคือ 5:1 หากมีสัดส่วนต่ำกว่านี้อาจจะทำให้ผลการวิเคราะห์มีปัญหาเรื่อง overfitting แต่แนะนำว่าสัดส่วนที่ดีกว่า คือ 15-20 หน่วยข้อมูลต่อ 1 ตัวแปรต้น

คำแนะนำที่ 2 Green SB. [26] (อ้างถึงใน [17]) เสนอให้ใช้สูตรคำนวณขนาดตัวอย่างในสมการทำนายค่าจากหลายตัวแปร คือ

$$n > 50 + 8m$$

โดยที่ $m =$ จำนวนตัวแปรต้น

และถ้าต้องการทดสอบความสัมพันธ์ระหว่างตัวแปรผลลัพธ์กับตัวแปรต้น แต่ละตัวสูตรโดยรวม คือ

$$n > 104 + m$$

คำแนะนำที่ 3 Tabachnick BG. & Fidell LS. [27] (อ้างถึงใน [17]) มีคำแนะนำเพิ่มเติมว่าให้เพิ่มขนาดตัวอย่างมากขึ้น

ในกรณีที่ตัวแปรผลลัพธ์มีลักษณะเป็นโค้งเบ้, ขนาดของผลต่างที่สนใจจะศึกษามีขนาดเล็ก, มีค่าความคลาดเคลื่อนจากการวัดพอสมควร (เครื่องมือที่ใช้วัดค่าข้อมูลไม่ดีนัก) และใช้วิธีการวิเคราะห์สมการแบบขั้นตอน (stepwise)

คำแนะนำที่ 4 Cohem J. [28] (อ้างถึงใน [17]) ในกรณีที่เป็นการทดสอบหาความสัมพันธ์โดยใช้สถิติไค-สแควร์ (Chi-square) ขนาดกลุ่มตัวอย่างที่เหมาะสม คือจำนวนคาดหวังต่ำสุดในช่องตารางใดๆ (cell) ไม่ควรน้อยกว่า 5 หากมีขนาดต่ำกว่านี้ จะเพิ่มโอกาสการตัดสินใจผิดพลาดแบบที่ 1 (Type I error) ได้ง่าย และขนาดของกลุ่มตัวอย่างรวมทั้งหมดไม่ควรต่ำกว่า 20 อีกประเด็นที่ควรคำนึงถึงคือ จำนวนช่องตาราง ทั้งหมดมีความเกี่ยวเนื่องโดยตรงกับค่าอำนาจจำแนกความแตกต่างของการทดสอบสมมติฐาน

คำแนะนำที่ 5 ในกรณีของการศึกษาความสัมพันธ์ระหว่างชุดตัวแปรต่างๆ ด้วยวิธีวิเคราะห์ข้อมูลแบบการวิเคราะห์องค์ประกอบ (factor analysis) มีคำแนะนำหลากหลาย ได้แก่ Tabachnick & Fidell LS. [27] (อ้างถึงใน [17]) แนะนำว่า $n \geq 300$ ส่วน Pedhazur & Schmelkin LP. [29] (อ้างถึงใน [17]) เสนอว่าควรมีอย่างน้อย 50 หน่วยข้อมูล ต่อ 1 ตัวแปร นอกจากนี้ Comrey AL. & Lee HB. [30] (อ้างถึงใน [17]) เสนอว่า $n = 50$ ไม่ดีมาก,

$n = 100$ ไม่ค่อยดี, $n = 200$ ปานกลาง, $n = 300$ ดี, $n = 500$ ดีมาก นั่นคือยิ่งมีสัดส่วนขนาดตัวอย่างต่อจำนวนตัวแปรมากขึ้น ก็จะทำให้ลดโอกาสการเกิด “over-fitting” (คือสร้างตัวแปรองค์ประกอบที่ไม่สามารถแปรผลได้อย่างแท้จริง)

7. การคำนวณขนาดตัวอย่างในกรณี
ที่ตัวแปรผลลัพธ์ต้องแปลงค่า (Data trans-
formation) สูตรการคำนวณตัวอย่างต่างๆ ที่กล่าวมาข้างต้นทั้งหมด มีข้อตกลงพื้นฐาน (assumption) ว่า ค่าข้อมูลวิธีการกระจายตัวเป็นโค้งปกติ หรือไม่ก็มีการกระจายตัวที่มีรูปแบบกำหนดชัดเจน (เช่น บัวซอง, เอกซ์โปเนนเชียล) แต่อย่างไรก็ตาม อาจจะมีหลายกรณีที่การกระจายตัวของข้อมูลไม่สอดคล้องกับข้อตกลงพื้นฐานที่กำหนด จึงอาจทำให้ต้องมีการแปลงค่าข้อมูลให้อยู่ในรูปแบบอื่นๆ (เช่น แปลงเป็นค่าเลขฐานสิบ ค่าเลขยกกำลัง) หากเป็นกรณีดังกล่าว การคำนวณขนาดกลุ่มตัวอย่างก็ต้องคำนวณจากค่าผลลัพธ์ที่แปลงค่าแล้ว ไม่ใช่ค่าที่เก็บมาแต่เดิม ตัวแปรผลลัพธ์ที่อาจต้องมีการแปลงค่า เช่น ระยะเวลาการเกิดอาการข้างเคียง ค่าใช้จ่ายในการรักษา ปริมาณไวรัสในเลือด ระยะเวลาพักรักษาตัวในโรงพยาบาล เป็นต้น ค่าตัวแปรประเภทรูปร่าง โดยธรรมชาติของมันมักจะมีลักษณะเป็นโค้งเบ้ (skewed distribution) ซึ่งสามารถแก้ไขให้มีลักษณะเป็นโค้งปกติได้ โดยการแปลงค่าเป็นเลขฐานลอการิทึม (logarithm) ดังนั้น ในการคำนวณ

ขนาดตัวอย่างก็จะใช้สูตรคำนวณที่ได้กล่าวมาแล้ว แต่แทนที่จะเป็นการคำนวณจากค่าตัวแปรผลลัพธ์ในเลขปกติ ก็จะคำนวณจากค่าผลลัพธ์ที่แปลงค่าแล้ว เช่น เป็นเลขฐานลอการิทึม และค่าผลต่างที่มีความหมายทางคลินิกก็จะอยู่ในเลขฐานที่เลือกไปด้วย เช่น ผู้วิจัยอาจต้องการทดสอบว่ายาที่ใช้สามารถลดปริมาณไวรัสในเลือดลงได้ 0.5 ล็อก แทนที่จะพิจารณาว่าจะลดลงให้ได้ 100 หรือ 1000 หรือ 10000 ก๊อปปี้ (copies)

8. การคำนวณขนาดตัวอย่างของการ
ทดสอบสมมติฐานทางสถิติแบบวิธีการของ
นอน-พาราเมตริก (Non-parametric tests)
ในการศึกษาวิจัยทางคลินิกแบบ RCT ผู้วิจัยอาจต้องตัดสินใจใช้สถิติแบบนอน-พาราเมตริก ซึ่งเป็นการทดสอบที่ไม่ได้ระบุลักษณะการกระจายตัวของข้อมูล (distribution free) นั่นคือ ไม่มีข้อตกลงพื้นฐานหรือไม่กำหนดว่าการกระจายตัวของข้อมูลจะต้องมีลักษณะเป็นโค้งปกติหรือโค้งอื่นใด ข้อดีของการทดสอบแบบนี้คือ มีข้อจำกัดในเรื่องข้อตกลงพื้นฐานต่างๆ น้อย (เช่น ข้อตกลงเรื่องการกระจายตัวของข้อมูล ข้อตกลงเรื่องความเท่ากันของการกระจายตัวของข้อมูล) ค่าตัวแปรผลลัพธ์อาจจะเป็นค่าที่วัดในระดับนามบัญญัติเป็นการจำแนกแจกแจงค่า (nominal / categorical) หรือเป็นค่าข้อมูลแบบเรียงลำดับ (ordinal) ก็ได้ นั่นคือข้อมูลที่วัดไม่จำเป็นต้องเป็นค่าต่อเนื่อง (interval) เสมอไป ส่วนข้อด้อยของวิธีการ

แบบนี้คือ เป็นวิธีการทดสอบที่อาจจะไม่มีอำนาจจำแนก ได้ดีเท่าวิธีการแบบพารามेटริก (parametric) สำหรับผู้สนใจการคำนวณขนาดตัวอย่างในกรณีที่ต้องการทำการวิเคราะห์แบบนอน-พารามेटริก (non-parametric) ขอแนะนำให้อ่านบทที่ 11 ในหนังสือเรื่อง “Sample Size Calculations in Clinical Research” ของ Chow SC., Shao J., Wang H. [1]

9. การเปรียบเทียบผลลัพธ์ในการวิจัยที่มีมากกว่า 2 กลุ่ม สูตรคำนวณขนาดตัวอย่างทั้งหมดที่ได้กล่าวมาแล้ว เป็นสูตรของการเปรียบเทียบผลลัพธ์ของการศึกษา 2 กลุ่ม ซึ่งเป็นรูปแบบการวิจัยทางคลินิกส่วนใหญ่ แต่ในการดำเนินการวิจัยทางคลินิกเพื่อศึกษาผลของปัจจัยที่ศึกษาไม่ว่าจะเป็นยา, วัคซีน, วิธีการรักษา เราจะพบว่ามีย่อยครั้งที่มีการเปรียบเทียบมากกว่า 2 กลุ่ม ตัวอย่างเช่น ผู้วิจัยอาจต้องการเปรียบเทียบผลลัพธ์ของกลุ่มควบคุมที่ได้ยามาตรฐาน กับยาตัวใหม่ 2 ตัว ในกรณีเช่นนี้ผู้วิจัยอาจวางแผนกำหนดให้ขนาดตัวอย่างของกลุ่มควบคุมและกลุ่มทดลองอีก 2 กลุ่มมีขนาดเท่ากัน หรือต่างกันก็ได้ จำนวนรวมทั้งหมดของกลุ่มตัวอย่างในโครงการวิจัยมีที่ปัจจัยที่ต้องการศึกษาทดลอง r กลุ่มจะมีค่าเท่ากับ

$$N = m_t + n_c$$

โดยที่ $r =$ จำนวนกลุ่มที่จะได้รับปัจจัยที่จะศึกษาทดลอง

$n_t =$ จำนวนขนาดกลุ่มตัวอย่างของกลุ่มทดลอง

$n_c =$ จำนวนขนาดกลุ่มตัวอย่างของกลุ่มควบคุม

ในการคำนวณขนาดตัวอย่างของกลุ่มตัวอย่างและกลุ่มควบคุมอาจกำหนดให้มีขนาดเท่ากัน ($n_c = n_t$) ก็ได้ หรืออาจกำหนดขนาดไม่เท่ากัน โดยผู้วิจัยอาจกำหนดอัตราส่วนของกลุ่มทดลองต่อกลุ่มควบคุมดังนี้

$$\lambda = n_t / n_c$$

ในการคำนวณขนาดกลุ่มตัวอย่างผู้วิจัยอาจใช้วิธีการคำนวณตัวอย่างโดยใช้สูตรคำนวณปกติของ 2 กลุ่มดังได้กล่าวมาแล้ว แล้วค่อยคำนวณหาขนาดตัวอย่างของแต่ละกลุ่มและขนาดตัวอย่าง รวมทั้งหมดได้ตัวอย่างเช่น ถ้ากำหนดให้ $r = 3$ และ $\lambda = 0.5$ หากผลจากการคำนวณขนาดตัวอย่าง โดยใช้สูตรที่ได้กล่าวมาก่อนหน้าแล้วได้คำตอบว่า $n_t = 50$ ดังนั้น $n_c = 50/0.5 = 100$ ดังนั้นขนาดกลุ่มตัวอย่างที่ต้องใช้ทั้งหมดในการศึกษาครั้งนี้คือ 250

โดยปกติเรามักจะพบว่ามีย่อยครั้งที่กลุ่มตัวอย่างที่เป็นกลุ่มควบคุมเพียงกลุ่มเดียว แล้วนำไปเปรียบเทียบกับกลุ่มทดลอง r กลุ่ม แต่หากมีกรณีที่ผู้วิจัยมีกลุ่มควบคุมมากกว่า 1 กลุ่ม อาจปรับสูตรคำนวณขนาดกลุ่มตัวอย่างที่ต้องใช้ทั้งหมดในการศึกษาได้

$$N = m_t + sn_c$$

โดยที่ s = จำนวนกลุ่มควบคุม หรือ ผู้วิจัยอาจใช้สูตรคำนวณเพื่อเปรียบเทียบกลุ่มทดลองกับกลุ่มควบคุมที่ละคู่เปรียบเทียบแล้วคำนวณหาค่าขนาดตัวอย่างรวมทั้งหมด (N) หากการคำนวณชุดใดให้ค่าขนาดกลุ่มตัวอย่างทั้งหมดมากกว่า ให้ใช้ขนาดกลุ่มตัวอย่างดังกล่าวในการทำวิจัย (Meinert & Tanascia, 1986)

ตัวอย่าง การศึกษาการรักษาหลอดเลือดหัวใจ (Coronary Drug Project) ของ Meinert CL. et al [31])

วัตถุประสงค์ ศึกษาเปรียบเทียบสัดส่วนการเสียชีวิตใน 6 กลุ่ม (1 กลุ่มควบคุม และ 5 กลุ่มทดลอง) ที่มีการศึกษาติดตามผลลัพธ์จากการได้ยาภายใน 5 ปี ที่ผู้วิจัยได้กำหนดอัตราส่วนของขนาดกลุ่มตัวอย่างเป็น 1:1:1:1:1:2.5 ($\lambda = 1/2.5$ นั่นคือ ขนาดกลุ่มควบคุมต่างจากกลุ่มทดลอง 2.5 เท่า) ผู้วิจัยคาดว่าจะยาในกลุ่มทดลองโดยรวมจะสามารถลดสัดส่วนการตายลงได้ 1 ใน 4 ของสัดส่วนเดิม หากไม่ได้ยา (คือกลุ่มควบคุม) ซึ่งมีค่าเท่ากับร้อยละ 30 นั่นคือผู้วิจัยใช้วิธีการคิดผลต่าง 2 กลุ่ม (effect size) เป็นแบบค่าสัมพัทธ์: $(\pi_1 - \pi_2) = \pi_1 \times 0.25$

สูตรคำนวณมาตรฐานสำหรับ 2 กลุ่มคือ

$$n_1 = \frac{\left(z_{\alpha/2} + z_{\beta} \right)^2 \times [\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]}{(\pi_1 - \pi_2)^2}$$

แต่ผู้วิจัยใช้สูตรคำนวณที่ซับซ้อนขึ้นกว่านี้เล็กน้อย โดยปรับค่าอัตราการเกิดผลลัพธ์เฉลี่ย (π) จาก 2 กลุ่ม (π_1 และ π_2) และปรับสัดส่วนขนาดตัวอย่างที่ไม่เท่ากันระหว่างกลุ่ม (λ) ด้วย ซึ่งจะเป็นสูตรที่หลายคนนิยมใช้แทนสูตรข้างบน

$$n_1 = \frac{(z_{\alpha/2} \sqrt{\pi(1-\pi)(\lambda+1)\lambda} + z_{\beta} \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)/\lambda})^2}{(\pi_1 - \pi_2)^2}$$

โดยที่

$$\pi = (\pi_1 + \lambda\pi_2) / (1 + \lambda)$$

$$\lambda = n_2 / n_1$$

จากวัตถุประสงค์การวิจัยของการวิจัยข้างต้น จะได้ค่าสถิติต่อไปนี้

$$n_1 = \text{ขนาดตัวอย่างของกลุ่ม 1}$$

$$\pi_1 = \text{ค่าสัดส่วนการตายในช่วง 5 ปี ในกลุ่มควบคุม} = 0.30$$

$$\pi_2 = \text{ค่าสัดส่วนของการตายในช่วง 5 ปี กลุ่มทดลอง} = 0.225$$

$$\pi = \text{ค่าเฉลี่ยสัดส่วนการตายของ 2 กลุ่ม} = 0.279$$

$$z_{\alpha/2} = \text{จุดตัดที่มีนัยสำคัญทางสถิติที่ } \alpha = 0.01 \text{ และเป็นการทดสอบแบบทิศทางเดียว} = 2.326$$

$$z_{\beta} = \text{จุดตัดที่บอกความสามารถในการจำแนกความแตกต่างระหว่างกลุ่มที่ } \beta = 0.05 \text{ (Power = ร้อยละ 95)} = 1.645$$

$$\lambda = \text{สัดส่วนขนาดตัวอย่างกลุ่มทดลองต่อกลุ่มควบคุม} = 0.400 \text{ (ได้จาก } 1/2.5)$$

แทนค่าในสูตรข้างต้นจะได้ว่า

$$n_1 = \frac{\left(2.326[(0.279 \times 0.721) \times (0.4 + 1) \times 0.4]^{\frac{1}{2}} + 1.645[(0.3 \times 0.7) + (0.225 \times 0.775) / 0.4]^{\frac{1}{2}}\right)^2}{(0.075)^2}$$

ผลการคำนวณคือ

$$n_1 = 1906$$

$$n_2 = (1/2.5) \times 1906 = 762$$

ผู้วิจัยคาดว่าจะมีโอกาสสูญเสียผู้เข้าร่วมโครงการ กล่าวคืออาจมีผู้ขาดหายไปอยู่ไม่ครบโครงการ 5 ปี ไปจำนวนหนึ่งประมาณร้อยละ 3 ดังนั้นจึงปรับขนาดเพิ่มขึ้น

$$n_1 = n_c = (1/0.7) \times 1906 = 2723$$

$$n_2 = n_t = (1/0.7) \times 762 = 1089$$

$$N = 5n_t + n_c \\ = 5(1089) + 2723 = 8168$$

การปรับขนาดตัวอย่างภายใต้เงื่อนไขเฉพาะ

ในการคำนวณขนาดกลุ่มตัวอย่างนั้น ผู้วิจัยอาจเผชิญกับปัญหาเฉพาะหน้าหลายประเด็น ผู้วิจัยจึงอาจจำเป็นต้องปรับขนาดของกลุ่มตัวอย่างที่คำนวณได้จากสูตรต่างที่กล่าวมาแล้วหรือแม้แต่กำหนดเงื่อนไขเฉพาะบางอย่างในการคำนวณขนาดกลุ่มตัวอย่าง ทั้งนี้เพื่อให้สามารถดำเนินการวิจัยได้ โดยยังคงความถูกต้องทางวิทยาศาสตร์และจริยธรรมในการตอบคำถามวิจัยที่ต้องการ

1. จำนวนตัวแปรผลลัพธ์ที่ใช้ในการ

คำนวณขนาดตัวอย่าง ดังได้กล่าวในตอนก่อนหน้านี้แล้วว่า การคำนวณขนาดตัวอย่างนั้น เราจะคำนวณจากค่าตัวแปรที่เป็นผลลัพธ์หลัก (primary endpoint/outcome) ดังที่เขียนไว้ในวัตถุประสงค์หลักของการวิจัย (primary objectives) แต่อย่างไรก็ตามอาจมีบางโครงการวิจัยที่ผู้วิจัยมีวัตถุประสงค์หลักของการวิจัยที่มีตัวแปรผลลัพธ์ที่สำคัญหลายตัว เช่น ผู้วิจัยอาจสนใจวัดการเกิดโรคติดต่อทางเพศสัมพันธ์ ซึ่งมีหลายโรค โดยทุกโรคที่สนใจนี้ถือว่ามีความสำคัญเท่าๆ กัน ดังนั้นขนาดกลุ่มตัวอย่างที่ต้องการศึกษาควรจะมีขนาดใหญ่พอที่จะครอบคลุมโอกาสเกิดโรคนั้นๆ ได้ทั้งหมด วิธีการคำนวณขนาดกลุ่มตัวอย่างจึงอาจทำได้อย่างง่ายๆ คือ คำนวณขนาดตัวอย่างของแต่ละผลลัพธ์หลักดังกล่าวทุกตัว แล้วจึงเลือกเอาขนาดใหญ่ที่สุดเป็นขนาดที่จะใช้ในการวิจัย ตัวอย่างเช่น ในการคำนวณขนาดกลุ่มตัวอย่าง สำหรับศึกษาการป้องกันการเกิดโรคติดต่อทางเพศ 3 โรค ได้แก่ โกลโนเรีย, ซิฟิลิส, และคลาไมเดีย ผู้วิจัยคำนวณค่ากลุ่มตัวอย่างที่ต้องการทั้ง 3 โรค ได้คำตอบคือ 800, 1,200 และ 1,000 คน ผู้วิจัยจึงเลือกเอง 1,200 เป็นขนาดกลุ่มตัวอย่างของการวิจัยครั้งนั้น

แต่อย่างไรก็ตาม มีนักสถิติเสนอให้ปรับขนาดตัวอย่างให้เล็กลงได้ [17] หากตัวแปรนั้นๆ มีค่าความสัมพันธ์กันเท่ากับ ρ โดยผู้วิจัยสามารถปรับลดลงได้ โดยใช้สูตรดังนี้

$$n_{new} = n(1 - \rho^2)$$

2. อัตราส่วนขนาดของกลุ่มตัวอย่างในการวิจัย (Allocation ratio) โดยปรกติแล้วมักจะกำหนดให้อัตราส่วนของขนาดกลุ่มตัวอย่าง (กลุ่มทดลอง : กลุ่มควบคุม) เท่ากัน นั่นคือ $n/n_c = 1$ การกำหนดอัตราส่วนเท่ากันนี้ เรียกว่าเป็นการกำหนดแบบ “uniform allocation” ซึ่งถือว่าเป็นวิธีการที่เหมาะสมสำหรับการวิจัยทางคลินิกส่วนใหญ่ที่ต้องการเปรียบเทียบการเกิดผลลัพธ์หลักใน 2 กลุ่ม โดยทั้ง 2 กลุ่มมีโอกาสการเกิดผลลัพธ์หลักเท่าๆ กัน และผู้เข้าร่วมโครงการวิจัยมีโอกาสถูกสุ่มเลือกเข้ากลุ่มใดกลุ่มหนึ่งเท่าๆ กัน

ผู้วิจัยอาจพิจารณากำหนดขนาดกลุ่มตัวอย่างของ 2 กลุ่มไม่เท่ากัน เรียกว่า “non uniform allocation” ด้วยเหตุผลต่างๆ กันออกไป

กรณีที่ 1 วัตถุประสงค์ของการวิจัยให้ความสำคัญกับกลุ่มทดลองกลุ่มใดกลุ่มหนึ่งเป็นพิเศษ ดังนั้นจึงต้องการเพิ่มความชัดเจนในการสรุปผลเปรียบเทียบผลลัพธ์ระหว่างกลุ่มทดลองนั้นๆ โดยเฉพาะ ตัวอย่างเช่นในการศึกษาเปรียบเทียบยาแอสไพริน

(aspirin) กับยาแอสไพริน เพอเซนทีน (aspirin persantine) และยาหลอก ซึ่งเป็นกลุ่มควบคุม ผู้วิจัยกำหนดให้ขนาดของกลุ่มตัวอย่างของ 2 กลุ่มแรกมีขนาดใหญ่กว่ากลุ่มที่ได้รับยาหลอกเป็น 2 เท่า เพราะผู้วิจัยให้ความสำคัญกับการเปรียบเทียบ 2 กลุ่มแรกมากกว่าการเปรียบเทียบ 2 กลุ่มดังกล่าวกับกลุ่มควบคุม

กรณีที่ 2 กรณีการเกิดผลลัพธ์ในกลุ่มหรือกลุ่มใดเป็นที่รู้จักกันดีอยู่แล้ว หรือมีหลักฐานพยาน และผลงานวิจัยอื่นๆ รายงานผลลัพธ์มากอยู่แล้ว ตัวอย่างเช่นในการเปรียบเทียบประสิทธิผลการรักษามาลาเรียด้วยยาตัวใหม่เมื่อเทียบกับยาอาร์ทีซูเนต (artesunate) ซึ่งใช้กันมานานและมีรายงานผลการรักษาตลอดจนอาการข้างเคียงมากมายอยู่แล้ว ผู้วิจัยจึงกำหนดให้อัตราส่วนของขนาดตัวอย่างของกลุ่มที่จะได้รับยาตัวใหม่ เมื่อเทียบกับกลุ่มที่จะได้รับยาอาร์ทีซูเนต มีค่าเท่ากับ 3:1

กรณีที่ 3 โอกาสการเกิดผลลัพธ์ที่สนใจ (เช่น โอกาสการเกิดโรคหรือโอกาสการเกิดผลข้างเคียงจากการใช้ยา) ไม่สูงมากนักและผู้วิจัยสนใจจะศึกษาหาความชัดเจนการเกิดผลลัพธ์ในกลุ่มใดกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง ตัวอย่างเช่น ในการทดสอบประสิทธิผลของวัคซีนป้องกันโรคเอดส์เมื่อเปรียบเทียบกับสารเลียนแบบวัคซีน ในกลุ่มชายที่มีเพศสัมพันธ์กับชายในอเมริกา ผู้วิจัยต้องการพิสูจน์อัตราการติดเชื้อในกลุ่ม

ได้รับวัคซีนว่าได้ผลมากน้อยแค่ไหนเมื่อเทียบกับกลุ่มควบคุมและอัตราการติดเชื้อก็ไม่สูงมากนัก ดังนั้นผู้วิจัยจึงกำหนดให้อัตราส่วนขนาดกลุ่มตัวอย่างที่ได้รับวัคซีนต่อกลุ่มควบคุมเท่ากับ 2:1

กรณีที่ 4 การปรับอัตราส่วนของขนาดกลุ่มตัวอย่างต่างกันมีความสัมพันธ์กับค่าความแปรปรวน (variance) ของผลลัพธ์ กล่าวคือ การกำหนดขนาดต่างกันจะลดความแปรปรวน และความคลาดเคลื่อนจากการตัดสินใจผิดแบบที่ 1 (Type I error) ได้ด้วย โดยเฉพาะอย่างยิ่งในกรณีที่มีหลายกลุ่มเปรียบเทียบ และต้องทำการเปรียบเทียบรายคู่ทีละคู่ในภายหลัง (multiple comparisons) ตัวอย่างเช่น ในการศึกษาการรักษาโรคหลอดเลือดหัวใจ ที่มีกลุ่มทดลอง 5 กลุ่มเทียบกับกลุ่มควบคุม ผู้วิจัยกำหนดให้อัตราส่วน 1:1:1:1:1:2.5 เพื่อควบคุมความคลาดเคลื่อนจากการตัดสินใจผิดแบบที่ 1 สำหรับทุกคู่การเปรียบเทียบระหว่างกลุ่มทดลองกับกลุ่มควบคุม (5 คู่เปรียบเทียบ) ให้ค่าอยู่ภายใต้ระดับนัยสำคัญ (α) ที่กำหนด โดยวิธีการควบคุมค่า α ในการเปรียบเทียบรายคู่ของ Dunnett สามารถทำได้โดยการกำหนดให้ จำนวนผู้ป่วยที่จะถูกจัดเข้าอยู่ในกลุ่มควบคุมเท่ากับ $\sqrt{r}$ คนสำหรับทุก 1 คนที่ถูกจัดเข้ากลุ่มทดลอง หรือกล่าวง่ายๆ คือ อัตราส่วนของกลุ่มควบคุมต่อกลุ่มทดลองเท่ากับ $\sqrt{5}:1:1:1:1:1$ แต่

ผู้วิจัยปัดค่า $\sqrt{5}$ เป็น 2.5 เพื่อความสะดวกในการคำนวณอัตราส่วนขนาดตัวอย่างของการศึกษาครั้งนั้น

3. อัตราการเกิดผลลัพธ์ (Response rate) ไม่ครบถ้วนตามความคาดหวัง การคำนวณขนาดตัวอย่างที่ต้องการนั้น เป็นการคำนวณขนาดที่ต้องการนำไปวิเคราะห์เมื่อสิ้นสุดโครงการ แต่จะพบเสมอว่ามีโอกาสมากมายที่ผู้เข้าร่วมโครงการ มักหายไปจากโครงการวิจัยทางคลินิกซึ่งมักจะเป็นโครงการที่ต้องมีการติดตามวัดผลลัพธ์สุดท้ายภายในช่วงเวลาที่กำหนด ทำให้ได้ข้อมูลที่ไม่สมบูรณ์ หรือไม่ได้จำนวนมากพอที่จะตอบคำถามวิจัยได้ โอกาสการไม่ได้ค่าผลลัพธ์ครบถ้วนอาจมาจากหลายสาเหตุ ได้แก่ การตอบข้อมูลไม่ครบถ้วนสมบูรณ์ (non-response rate), การขาดหายไปจากโครงการวิจัย (follow-up rate), การถอนตัวออกจากโครงการ (drop-out rate) และการไม่ทำตามกระบวนการศึกษาภายในโครงการ (non-compliance/adherence rate)

การขาดหายไปเหล่านี้อาจทำให้เกิดอคติในการสรุปผลการวิจัย เพราะผู้ที่ขาดหายไปอาจมีลักษณะบางอย่างต่างไปจากผู้ที่ยังคงอยู่ อีกทั้งการขาดหายไปยังลดความเชื่อมั่นของค่าสถิติที่คำนวณได้จากกลุ่มตัวอย่าง การไม่ทำตามกระบวนการศึกษา เช่น ผู้ป่วยอาจไม่ยอมรับวิธีการรักษาหรือไม่ยอมรับขนาดยาที่แพทย์จัดให้ ผู้ป่วยอาจลด/

เพิ่มขนาดยาเอง หรือจัดหาจากแหล่งอื่น มาเสริมเพิ่มเติมเข้าไปเอง ทั้งหมดนี้ก่อให้เกิดความคลาดเคลื่อนในการสรุปผลได้

ผู้วิจัยต้องคำนึงถึงประเด็นนี้ล่วงหน้า ในบางงานวิจัยผู้วิจัยจะทำการสำรวจหาอัตราการขาด/ถอนตัว หรืออัตราการไม่ทำตามกระบวนการศึกษาล่วงหน้า เป็นการศึกษาสำรวจขนาดเล็ก ก่อนที่จะดำเนินการจริง แต่ถ้าไม่มีการดำเนินการดังกล่าวก็อาจใช้วิธีการคาดเดาจากการศึกษาอื่นๆ ซึ่งมักจะมีช่วงอัตราการไม่ครบถ้วนจากข้อมูลอยู่ระหว่างร้อยละ 10-30 ผู้วิจัยจึงต้องปรับขนาดตัวอย่างจากที่คำนวณได้จากสูตรคำนวณเพิ่มเติมเอาไว้ ดังนี้

$$n_{new} = \frac{n}{1-L}$$

โดยที่ L = อัตราการมีข้อมูลขาดหายไป

ตัวอย่างเช่น หากผู้วิจัยคำนวณค่า $n = 100$ และคาดว่าจะมีโอกาสข้อมูลสูญหาย และขอถอนตัวจากการวิจัยรวมร้อยละ 30 ดังนั้น ขนาดกลุ่มตัวอย่างที่จะต้องใช้ในการวิจัยครั้งนี้ จึงควรเท่ากับ $100/(1-.30)$ คน

4. การคำนวณค่าอำนาจบอกความแตกต่างระหว่างกลุ่ม ในหลายกรณีงานวิจัยไม่อาจทำได้เลยหากใช้ขนาดกลุ่มตัวอย่างที่คำนวณได้ เพราะขนาดกลุ่มตัวอย่างที่ต้องใช้มากเกินไป และเป็นไปไม่ได้ที่จะได้ผู้เข้าร่วมโครงการจากผู้ป่วยในจำนวนมากขนาดนั้น แม้ว่าจะเป็นการจัดเก็บข้อมูลจากหลาย

สถาบัน/แหล่งข้อมูล (multi-center/multi-countries) แล้วก็ตาม ดังนั้นผู้วิจัยอาจต้องใช้วิธีการคำนวณย้อนกลับ กล่าวคือใช้สูตรคำนวณที่ได้กล่าวมาแล้วข้างต้น แต่จะคำนวณจากขนาดกลุ่มตัวอย่างที่เป็นไปได้ที่รู้ว่าเราจะได้จริง (ค่า n) แล้วคำนวณย้อนกลับว่าที่ขนาดกลุ่มตัวอย่างดังกล่าว เราจะได้ค่าอำนาจบอกความแตกต่างระหว่างกลุ่มเท่าใด ถ้ายังอยู่ในขอบเขตที่พอรับได้ ไม่ต่ำกว่าร้อยละ 80 มากนัก งานวิจัยดังกล่าวก็อาจจะพอเป็นไปได้ที่จะได้รับการยอมรับ

วิธีการเช่นนี้อาจประยุกต์ใช้กับการณีของการศึกษาวิจัยที่เริ่มดำเนินการไปแล้ว โดยคาดว่าจะได้จำนวนขนาดตัวอย่างตามที่คำนวณได้ในตอนต้น แต่หลังจากเก็บข้อมูลมาจนครบระยะเวลาหนึ่งที่จะต้องปิดโครงการแล้ว แต่ยังไม่อาจได้ขนาดตัวอย่างดังกล่าวได้ ผู้วิจัยก็อาจใช้วิธีการคำนวณย้อนกลับจากข้อมูลที่รวบรวมได้ และสรุปสถิติของผลลัพธ์ที่ได้จากการวิเคราะห์เปรียบเทียบนั้นมีค่าอำนาจ $(1 - \beta)$ เท่าใด หากมีน้ำหนักมากพอ ก็อาจได้รับการพิจารณาตีพิมพ์

ในทำนองเดียวกันผู้วิจัยอาจคำนวณย้อนกลับ เพื่อหาค่าผลต่างที่อาจได้จากการเปรียบเทียบผลลัพธ์ระหว่างกลุ่ม (detectable difference) ซึ่งก็คือค่าผลต่างที่มีความหมายทางคลินิกนั่นเอง โดยผู้วิจัยอาจทดลองปรับค่าขนาดตัวอย่างที่ระดับต่างกันและ/หรือที่ค่าอำนาจจะต่างกัน เพื่อดูว่าจุดสมดุล

ลงตัวของขนาดผลต่างต่ำสุดที่ยังมีความหมายทางคลินิกอยู่ที่จุดใดและที่จุดนั้นมีค่าอำนาจเท่าใด ต้องใช้ขนาดกลุ่มตัวอย่างเท่าใด

โปรแกรมคำนวณขนาดกลุ่มตัวอย่างเกือบจะทุกโปรแกรมสามารถคำนวณย้อนกลับเช่นนี้ได้ นั่นคือแทนที่จะเลือกให้คำนวณขนาดตัวอย่างจากค่าสถิติพื้นฐานที่กำหนดที่ได้กล่าวมาในตอนต้นๆ ของบทนี้ ผู้วิจัยกลับกำหนดขนาดตัวอย่างและค่าสถิติอื่นๆ เข้าไปในสูตรเพื่อให้โปรแกรมคำนวณหาอำนาจ หรือค่าผลต่างที่จะได้รับออกมาแทน

การเขียนอธิบายขนาดตัวอย่างในโครงร่างวิจัย

ผู้เขียนโครงร่างวิจัยมักจะมีคำถามอยู่บ่อยครั้งว่าควรเขียนบรรยายวิธีการคำนวณขนาดกลุ่มตัวอย่างอย่างไร คำตอบก็คือควรบรรยายที่มาที่ไปของค่าสถิติพื้นฐานต่างๆ ที่ใช้ในการคำนวณขนาดตัวอย่าง

1. ตัวอย่างการเขียนบรรยายขนาดกลุ่มตัวอย่างที่สื่อความหมาย

ตัวอย่างที่ 1 ขนาดกลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นเด็กทารก 292 คน (146 ในกลุ่มทดลองและ 146 ในกลุ่มควบคุม) ซึ่งเป็นจำนวนซึ่งมากพอที่จะทดสอบความแตกต่างอันมีความหมายทางคลินิกของอัตราการเกิดภาวะพิษเหตุติดเชื้อ (sepsis) ภายใน 14 วัน ใน 2 กลุ่มตัวอย่างดังกล่าว ขนาด

ตัวอย่างนี้คำนวณจากสูตรเปรียบเทียบสัดส่วนของการเกิดภาวะพิษเหตุติดเชื้อในทารก โดยคาดว่าจะมีผลต่างที่มีความหมายทางคลินิกร้อยละ 16 ระหว่าง 2 กลุ่ม นั่นคือจากรายงานกลุ่มควบคุมมีสัดส่วนการเกิดภาวะพิษเหตุติดเชื้อ เท่ากับร้อยละ 50 และกลุ่มที่ได้วิธีการแบบใหม่จะมีสัดส่วนการเกิดภาวะพิษเหตุติดเชื้อ เท่ากับร้อยละ 34 การเปรียบเทียบค่าผลลัพธ์ระหว่าง 2 กลุ่มนี้เป็นการทดสอบสมมติฐานแบบ 2 ทิศทางที่ค่านัยสำคัญทางสถิติ 0.05 และค่าอำนาจจำแนกความแตกต่างร้อยละ 80

ตัวอย่างที่ 2 ขนาดกลุ่มตัวอย่างที่คำนวณได้คือ 38 คนต่อกลุ่ม ซึ่งเป็นขนาดที่มากพอที่จะบอกความแตกต่างที่มีความหมายทางคลินิกได้ในระดับ 5 คะแนนระหว่างคะแนนแนวโน้มการฆ่าตัวตายของกลุ่มที่ได้วิธีการรักษาแบบใหม่เมื่อเทียบกับกลุ่มที่ได้รับการรักษาแบบมาตรฐานเดิม ขนาดตัวอย่างนี้ได้จากการคำนวณค่าการกระจายตัว (standard deviation) ของคะแนนแนวโน้มฆ่าตัวตายเท่ากับ 7.7 (เอกสารอ้างอิงที่มาของเลข 7.7) โดยกำหนดค่านัยสำคัญทางสถิติ 0.05 และค่าอำนาจจำแนกความแตกต่างร้อยละ 80 การคำนวณครั้งนี้มีสมมติฐานการกระจายตัวของคะแนนแนวโน้มการฆ่าตัวตายที่ได้จากแบบทดสอบ beck มีลักษณะเป็นโค้งปกติ แต่อย่างไรก็ตาม ผู้วิจัยคาดว่าจะมีอัตราการขอลถอนตัวจากโครงการเท่ากับ 1/3 ของจำนวนทั้งหมด ดังนั้นจึงเพิ่มกลุ่ม

ตัวอย่างเป็นกลุ่มละ 60 คน รวมขนาดตัวอย่างของโครงการนี้คือ 120 คน

ตัวอย่างที่ 3 ขนาดกลุ่มตัวอย่างที่ต้องสรรหาเข้าโครงการคือ 1,200 คน โดยที่ผู้เข้าร่วมโครงการจะถูกสุ่มเข้ากลุ่มยาใหม่ต่อยามาตรฐานที่ใช้อยู่เดิม ในอัตราส่วน 2:1 การวัดผลลัพธ์ครั้งสุดท้ายจะวัดที่เดือนที่ 3 หลังจากรับเข้าโครงการแล้ว ในการคำนวณขนาดกลุ่มตัวอย่างที่ต้องการในการวิเคราะห์นี้ ได้ทำการตรวจสอบอัตราการรักษาหายของยามาตรฐานพบว่าเป็นยาที่มียอมรับมานานและมีอัตราการรักษาหายขาดตั้งแต่ร้อยละ 85-95 (รายการเอกสารอ้างอิงที่มาของเลข 85 - 95) ซึ่งในการวิจัยครั้งนี้จะถือว่าอัตราการรักษาหายด้วย ยามาตรฐานอยู่ที่ระดับร้อยละ 90 ดังนั้นการวิจัยครั้งนี้จึงมีจุดมุ่งหมายจะทดสอบสมมติฐานว่ายาดูใหม่ไม่ด้อยกว่ายามาตรฐานเดิม (non-inferiority test) โดยผลต่างของยาใหม่กับยาเก่าไม่ควรต่างกันเกินร้อยละ -5 จึงจะถือว่าไม่ด้อยกว่ากัน ในการทดสอบสมมติฐานนี้กำหนดค่านัยสำคัญทางสถิติที่ 0.05 ค่าอำนาจจำแนกความแตกต่างที่ร้อยละ 80 และไม่มีการทดสอบทางสถิติ

อื่นใดระหว่างการศึกษา (interim-analysis) กล่าวคือจะมีแต่การวิเคราะห์ข้อมูลเมื่อเสร็จสิ้นสุดโครงการ (final analysis) (ดูตารางที่ 5)

ในโครงการนี้ต้องใช้ขนาดกลุ่มตัวอย่างรวมทั้งสิ้น 1,092 คน เพื่อการทดสอบสมมติฐานแบบไม่ด้อยกว่ากัน ในการวิเคราะห์ข้อมูลตอนสิ้นสุดโครงการ ในอัตราส่วน 2:1 ขนาดกลุ่มตัวอย่างของยาใหม่ต่อยามาตรฐานเท่ากับ 728:364 การสรรหาผู้เข้าร่วมโครงการอาจใช้เวลาประมาณ 1 ปี และติดตามผลแต่ละราย 5 ครั้งใน 3 เดือน หลังจากเข้าร่วมโครงการ แต่อย่างไรก็ตามเพื่อปรับค่าโอกาสที่อาจมีผู้ขอลถอนตัวหรือขาดหายไปจากการติดตามผล 3 เดือน คาดว่าประมาณร้อยละ 10 ของผู้เข้าร่วมโครงการในตอนแรก ดังนั้น จึงต้องใช้ขนาดกลุ่มตัวอย่างทั้งหมด 1,200 คน ซึ่งจะถูกรวมเข้ากลุ่มยาใหม่ 800 รายและยามาตรฐาน 400 ราย

ตัวอย่างที่ 4 ขนาดกลุ่มตัวอย่างที่จะใช้ในโครงการนี้ ได้จากการคำนวณภายใต้

ตารางที่ 5 ขนาดกลุ่มตัวอย่างที่จะใช้ในการวิจัย

อัตราส่วน	ขนาดตัวอย่าง		ขนาดตัวอย่างปรับร้อยละ 10	
	ยามาตรฐาน	ยาใหม่	ยามาตรฐาน	ยาใหม่
1:2	364	728	400	800
1:3	323	969	355	1066

เงื่อนไขที่กำหนดต่อไปนี้เป็นคือ วัตถุประสงค์ของการวิจัยเป็นการเปรียบเทียบอัตราการรักษาหายของยามาตรฐานกับยาตัวใหม่ โดยกำหนดว่ายาตัวใหม่ต้องมีคุณสมบัติไม่ด้อยกว่ายามาตรฐานเดิม ทั้งนี้เพราะยามาตรฐานเดิมมีอัตราการรักษาหายในระดับสูง ตามรายงานต่างๆ พบว่ามีอัตราการรักษาหายอยู่ระหว่างร้อยละ 85-95 ซึ่งในการวิจัยครั้งนี้ถือว่าอัตราการรักษาหายของยามาตรฐานอยู่ที่ร้อยละ 90 ทั้งนี้ในการคำนวณกลุ่มตัวอย่าง ผู้วิจัยได้กำหนดอัตราส่วนของขนาดตัวอย่างของยามาตรฐานต่อยาใหม่ เป็น 2 ขนาดคือ 1:2 และ 1:3 และการทดสอบสมมติฐานจะสรุปผลที่ระดับนัยสำคัญ 0.05

ตารางที่ 6 แสดงผลการคำนวณว่า ในกรณีที่กำหนดให้ผลต่างระหว่างสองกลุ่มไม่ด้อยกว่ากันเป็นค่าคงที่อยู่ที่ระดับไม่เกินร้อยละ 5 ถ้าจำนวนขนาดตัวอย่างที่ได้ในกลุ่มยามาตรฐานและยาใหม่ต่างกันออกไปในอัตราส่วนที่ 1:2 หรือ 1:3 จะได้ค่าอำนาจจำแนกจากการทดสอบต่างๆ กันออกไป ตัวอย่างเช่น ในกรณีอัตราส่วนที่ 1:2 ถ้าเลือกขนาดตัวอย่าง 2 กลุ่มที่ 300 คน (100:200) ค่าอำนาจเท่ากับร้อยละ 31.3 ซึ่งถือว่าต่ำมากเกินไป หากเลือกขนาดตัวอย่าง 2 กลุ่มที่ 1,092 คน (364:728) ก็จะได้ค่าอำนาจที่เหมาะสมคือร้อยละ 80

ในทางตรงกันข้ามหากเลือกกำหนดค่าอำนาจเท่ากับร้อยละ 80 เป็นค่าคงที่ โดยต้องการใช้ขนาดกลุ่มตัวอย่าง n ที่ 300 คน

(100:200) ผลการทดสอบสมมติฐานแบบไม่ด้อยกว่ากันที่ระดับนัยสำคัญ 0.05 จะสามารถบอกความแตกต่างระหว่าง 2 กลุ่ม (detectable difference/effect size) ได้ที่ขนาดร้อยละ -10 แต่ถ้าเลือกขนาดตัวอย่าง 1,092 (364:728) ก็จะสามารถบอกความแตกต่างระหว่างกลุ่มได้ที่ระดับร้อยละ -5

โดยสรุปแล้ว ในการศึกษาครั้งนี้ผู้วิจัยกำหนดให้อัตราส่วนขนาดตัวอย่างของกลุ่มยามาตรฐานต่อยาใหม่เท่ากับ 1:2 และเป็นการทดสอบสมมติฐานแบบไม่ด้อยกว่าที่ค่าอำนาจ = ร้อยละ 80 และระดับนัยสำคัญ 0.05 โดยที่ถือว่าระดับผลต่างระหว่าง 2 กลุ่มที่สามารถบอกได้จากการทดสอบสมมติฐาน และมีความหมายทางคลินิกว่า ไม่ด้อยกว่ากันอยู่ที่ระดับร้อยละ -10 ดังนั้นขนาดกลุ่มตัวอย่างที่ต้องใช้ในการวิเคราะห์เมื่อปิดโครงการคือ 300 คน (100 คนในกลุ่มยามาตรฐาน และ 200 คนในกลุ่มยาใหม่) เมื่อปรับโอกาสร้อยละ 10 ของการขอถอนตัวหรือการขาดหายไปจากโครงการซึ่งจะเป็นการติดตามผล 3 เดือนแล้ว ขนาดกลุ่มตัวอย่างที่ต้องสรรหาเข้าโครงการในตอนเริ่มต้นคือ 333 คน (111:222) เป็นอย่างต่ำ

2. ตัวอย่างการเขียนบรรยายการคำนวณขนาดกลุ่มตัวอย่างที่ไม่เหมาะสม

วิธีการเขียนบรรยายขนาดตัวอย่างในโครงร่างวิจัยที่ควรหลีกเลี่ยง คือการบรรยายที่ไม่สื่อความหมาย และ/หรือมีความคลุมเครือไม่ชัดเจน ว่ามีที่มาที่ไปของค่าสถิติที่นำมา

ตารางที่ 6 ขนาดกลุ่มตัวอย่างภายใต้เงื่อนไขที่กำหนด

ขนาดตัวอย่าง		Overall Statistical Power	Detectable Difference at Power = 80%
ยามาตรฐาน	ยาใหม่		
Ratio 1:2			
100	200	31.3%	-0.10
150	300	4.2%	-0.08
200	400	55.4%	-0.07
364	728	80.1%	-0.05
Ration 1:3			
50	150	18.5%	-0.13
75	225	26.8%	-0.11
100	300	34.7%	-0.09
150	450	48.6%	-0.07
200	600	60.3%	-0.06
323	969	80.0%	-0.05

คำนวณอย่างไร การเขียนที่ไม่ชัดเจน หรือไม่มีเหตุผลสนับสนุนเพียงพออาจทำให้ต้องเขียนและเสนอกลับไปมาหลายครั้ง และอาจไม่ผ่านการอนุมัติจากคณะกรรมการพิจารณาโครงร่างการวิจัย ด้วยคณะกรรมการไม่เชื่อว่าผู้วิจัยจะได้คำตอบที่มีความหมายพอ ตัวอย่างการเขียนบรรยายที่ไม่เหมาะสม เช่น

ตัวอย่างที่ 1 “จากผลการวิจัยก่อนหน้านี้พบว่ากลุ่มตัวอย่างขนาด 150 คน ทำให้สามารถพบความแตกต่างที่มีนัยสำคัญทางสถิติ ($p = 0.014$) ดังนั้น การวิจัยครั้งนี้จึงจะใช้กลุ่มตัวอย่างขนาดเท่ากัน”

(ผู้วิจัยควรคำนวณขนาดกลุ่มตัวอย่างสำหรับการวิจัยครั้งนี้ โดยอาจใช้ข้อมูลจากการศึกษาเดิมเป็นฐานในการคิดคำนวณครั้งใหม่นี้)

ตัวอย่างที่ 2 “ในการวิจัยครั้งนี้ไม่ได้คำนวณขนาดกลุ่มตัวอย่าง เพราะไม่มีการวิจัยอื่นๆ ที่ได้ทำในเรื่องทำนองเดียวกันนี้มาก่อนหน้าการวิจัยครั้งนี้เลย”

(ผู้วิจัยควรศึกษาพื้นฐาน (pilot study) ก่อน เพื่อหาค่าสถิติพื้นฐานก่อนที่จะนำมาคำนวณหาขนาดกลุ่มตัวอย่างที่จะใช้ในการศึกษาเต็มรูปแบบ (full scale study)

หรือไม่ก็หาข้อมูลจากรายงานภายในหน่วยงานที่เกี่ยวข้องในเรื่องนั้นๆ มาเป็นพื้นฐานในการคำนวณขนาดตัวอย่าง)

ตัวอย่างที่ 3 “จากสถิติทางคลินิกเราพบว่าผู้ป่วยด้วยโรคนี้ประมาณปีละ 50 ราย ซึ่งคาดว่าจะมีประมาณร้อยละ 10 ที่ไม่ยอมเข้าร่วมโครงการวิจัย ดังนั้น ขนาดกลุ่มตัวอย่างของการวิจัยครั้งนี้จึงคาดว่าเท่ากับ 90 ราย ในการศึกษา 2 ปี”

(แม้ว่าการคำนวณขนาดตัวอย่างจะเป็นการชั่งน้ำหนักระหว่างความเป็นไปได้ของโครงการ (*feasibility*) กับค่าอำนาจจำแนกความแตกต่างก็จริง แต่ผู้วิจัยไม่สามารถจะตัดสินใจดำเนินการได้จากความเป็นไปได้เพียงอย่างเดียว ในกรณีนี้ผู้วิจัยควรจะคำนวณภายใต้ข้อจำกัดในเรื่องจำนวนผู้ป่วยที่มีอยู่นั้น เราจะได้ค่าอำนาจ หรือผลต่างที่จะพบ (*detectable difference*) เท่ากับเท่าใด และเป็นค่าที่มีความหมายทางคลินิกหรือไม่ ถ้าไม่ได้ผู้วิจัยก็จำเป็นต้องพิจารณาขยายระยะเวลาการศึกษา หรือเพิ่มหน่วยจัดเก็บข้อมูลไปยังคลินิกอื่นๆ)

บทสรุป

ขนาดกลุ่มตัวอย่างมีบทบาทสำคัญในการวิจัย เพราะเป็นหนึ่งในตัวกำหนดว่าจะตอบคำถามวิจัยได้อย่างแท้จริงหรือไม่ และเป็นประเด็นที่นักวิจัยส่วนใหญ่รู้สึกลำบากใจ เพราะจะต้องพยายามชั่งน้ำหนักระหว่างความเป็นไปได้ ทั้งด้านการบริหาร

จัดการโครงการ การเงิน ตลอดจนจริยธรรม และกระบวนการทางวิทยาศาสตร์/สถิติ

ทีมงานวิจัยจะต้องพยายามสรรหาผู้เข้าร่วมโครงการให้ได้ตามเป้าหมายที่กำหนดภายในระยะเวลาที่กำหนด ซึ่งอาจทำได้โดยการให้ค่าตอบแทนค่าเสียเวลาและการเดินทางของผู้เข้าร่วมโครงการ แต่ก็ต้องพยายามไม่ให้ละเมิดหลักการทางจริยธรรมที่ว่าด้วยการว่าจ้างล่อลวงผู้เข้าร่วมวิจัย

ทีมงานวิจัยจะต้องพยายามรักษาอัตราการคงอยู่ (*retention*) ของผู้เข้าร่วมวิจัยมิให้สูญหายหรือขอถอนตัวออกไปจากโครงการ การรักษาอัตราการคงอยู่นี้บางครั้งยากกว่าการสรรหาผู้เข้าร่วมโครงการเสียอีก โอกาสที่จะทำให้เกิดการสูญหาย ขอถอนตัว หรือไม่ทำตามวิธีการศึกษาเป็นไปได้จากหลายสาเหตุ ซึ่งผู้วิจัยต้องพยายามแก้ไขให้ทันกาล เช่น สถานการณ์แปลกปลอมเกิดขึ้นแทรกในระหว่างการวิจัย มีปัญหาด้านข่าวลือ การเมือง เจ้าหน้าที่กับผู้เข้าร่วมโครงการมีความขัดแย้ง เจ้าหน้าที่โครงการไม่เข้าใจโครงการอย่างแท้จริงและไม่ตั้งใจทำงานที่มอบหมาย มีการสรรหากลุ่มผู้เข้าร่วมโครงการที่ไม่เหมาะสม

วิธีการแก้ไขที่ควรจะต้องวางแผนล่วงหน้า เพื่อให้การสรรหาและการรักษาอัตราการคงอยู่ในโครงการครบถ้วน เพื่อให้ได้ขนาดตัวอย่างตามเป้าหมายที่กำหนดสุดท้ายได้แก่ (1) การทำการศึกษาวิจัยขนาดเล็ก (*pilot study*) ก่อน เพราะจะทำให้เห็นแนวโน้มของสิ่งที่อาจเกิดขึ้นในโครงการหลัก (2) การวางแผนงาน

ที่ดี ทั้งในแง่การบริหารจัดการทีมงานวิจัย และการบริหารจัดการผู้เข้าร่วมโครงการวิจัย, (3) การขอขยายเวลาการศึกษา และการบริหารจัดการหรือขอเพิ่มทุนวิจัย และ (4) การตั้งเป้าหมายเป็นระยะๆ ที่จะต้องบรรลุทั้งในแง่ของทีมงานวิจัยและผู้เข้าร่วมโครงการวิจัย

หมายเหตุ ในระหว่างเขียนบทนี้ ผู้เขียนได้พบเอกสารจาก [http:// www.lehanathabane.com](http://www.lehanathabane.com) ซึ่งเกือบจะเหมือนกับเนื้อหาที่ผู้เขียนวางแผนการเขียนไว้ ดังนั้นรายละเอียดของบทนี้ราวร้อยละ 60 รวมทั้งสูตรที่ใช้คำนวณขนาดตัวอย่างต่างๆ ส่วนใหญ่เป็นการอ้างอิงต่อมาจากการเรียบเรียงจากเอกสารเรื่อง “Sample Size Determination in Clinical Trials” ของ Dr. Lehana Thabane (2005) [17] ซึ่งเป็นอาจารย์สังกัดภาควิชา Clinical Epidemiology & Biostatistics, McMaster University ทั้งนี้ผู้เขียนได้รับอนุญาตจาก Dr. Thabane แล้ว

เอกสารอ้างอิง

1. Chow S-C, Shao J, Wang H. Sample Size calculations in clinical Research. New York: Chapman & Hall / CRC; 2008.
2. Lwanger SK & Lemeshow S. Sample Size Determination in Health Studies. Geneva:WHO; 1991.
3. Zar JH. Biostatistical Analysis. New York: Prentice-Hall, Inc; 1996.
4. Levy PS & Lemeshow S. Sample of Populations: Methods and Applications (3rd Ed) New York: John Wiley & Son, Inc.; 1999
5. Elashoff JD. nQuery Advisor (Version 6.0). Massachusetts A: Statistical Solutions, Ltd.; 2005
6. Fleiss JL, Levin BA, Paik MC. Statistical Methods for rates and Proportions. New Jersey: Wiley. Hoboken; 2003.
7. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating Sample sizes for comparing independent proportions. Biometrics. 1980; 36: 343-346.
8. Bland JM. Statistics Guide for Research Grant Applicants: Sample Size Calculation. Available from: <http://www.sgul.ac.uk>
9. Chan YK. Randomised Controlled Trials (RCTs) - Sample Size : The Magic Number? Singapore Med. J. 2003; 44(4):172-174.
10. Dell RB, Holleran S & Ramakrishnan R. Sample Size Determination. ILAR Journal. 2002; 43(4):207-213.

11. DeMets DL. Introduction to Clinical Trials: Sample Size Issues. Available from: <http://www.biostat.wisc.edu>
12. Fayers PM. et al. Sample Size Calculation for Clinical Trials: The Impact of Clinician Belief. British Journal of Cancer. 2000; 82(1):213-219.
13. Jones SR, Carley S & Harrison M. An Introduction to Power and Sample Size Estimation. Emerg Med. J. 2003; 20:453-458.
14. Kerry L & Durham NC. Sample Size and Interim Analysis Issues for Dose Selection. American Heart Journal. 2000; 139(4):S161-S165.
15. Lenth RV. Some Practical Guidelines for Effective Sample Size Determination. Available from: <http://www.stat.viowa.edu>
16. Xu C. Clinical Trial Writing II : Sample Size Calculation and Randomization. Available from: <http://www.cct.cuhk.edu.hk>
17. Thabane L. Sample Size Determination in Clinical Trials. Available from: <http://www.lebanathabane.com,2005>
18. Vickers AJ. How Many Reported Measures in Repeated Measures Designs? Statistical Issues for Comparative Trials. BMC Medical Research Methodology. 2003;[9pp.]. Available from: <http://www.biomedcentral.com/1471-2288/3/22>.
19. Southworth H. Sample Size Calculation for Survival Studies. Available from: <http://www.splus.com>
20. Collett D. Modelling Survival Data in Medical Research. London: Chapman & Hall; 1994.
21. Rubinstein LV, Gail MK & Santner TJ. Planning the duration of a comparative trial with loss to follow-up and a period of continued observation. Journal of Chronic Disease. 1991; 34: 469-479.
22. Laleatos E & Lan KKG. A Comparison of sample size methods for the logrank statistics in medicine. 1992; 11: 179-191.
23. Lachin JM & Foulkes MA Evaluation of sample size and power for analysis of survival with allowance for nonuniform patient entry, loss of follow-up, non-compliance and

- stratification. Biometrics. 1986; 42: 507-519.
24. Freedman LS. Tables of the number of patients required in clinical trials using logrank test. Statistics in Medicine. 1982; 1: 121-129.
25. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate Data Analysis (5th Edition) New Jersey: Prentice Hall; 1998.
26. Green SB. How many subject does it take to do a regression analysis? Multivariate Behavioral Research. 1991; 26; 499-510.
27. Tabachnick BG, Fidell LS. Using Multivariable Statistics. New York: Harper Collins; 2007.
28. Cohen J. Statistical power analysis for the behavioral sciences (2nd Edition). New Jersey: Erlbaum; 1988.
29. Pedhazur EJ & Schmelkin LP. Measurement, design and analysis: An integrated approach. New Jersey: Erlbaum; 1991.
30. Comrey AL & Lee HB. A first course in factor analysis (2nd Edition) New Jersey: Erlbaum; 1992.
31. Meinert CL. & Tonascia S. Clinical Trials Design Conduct and Analysis. New York: Oxford University Press; 1996.

