

האם חישוב גודל מדגם מתחיל ונגמר בהצבה בנוסחה?

דקלה גבע, אינטגרטיסט

Diklah@IntegriStat.Com

נובמבר 2004

קבלתם החלטה לצאת למחקר, והגדרתם את מטרות המחקר. אז כמה זה עולה וכמה נבדקים דרושים? בשיחה הראשונית עם הסטטיסטיקאית נתבקשתם להביא כל מיני נתוני רקע, והיא שאלה שאלות רבות שבעצם לא ברור כיצד הן קשורות לקביעת גודל המדגם הדרוש. מדוע לא נתן להציב כמה מספרים בנוסחת חישוב ולתת תשובה?

מאמר זה סוקר את הגורמים המשפיעים על קביעת גודל המדגם. ישנם פרמטרים הסתברותיים הניתנים לכימות. גורמים אחרים אינם ניתנים לכימות ובכל זאת ישפיעו על גודל המדגם. הסטטיסטיקאית, בפגישה עם החוקר, מבררת את הגורמים המבניים אשר משפיעים על בחירת הנוסחה לחישוב גודל המדגם. לאחר מכן היא מנסה לדלות מכם אינפורמציה על הפרמטרים, בכדי להציב בנוסחה. ואז יחושבו המספרים ההתחלתיים, אשר בהמשך ידרשו לתיקונים שונים. כעת ניתן לקבוע את גודל המדגם המחושב והמתוקן.

הסטטיסטיקאית מוודאת את ההתאמה בין כל הפרטים ומוודאת ששאלת המחקר אכן נתנת לבדיקה בתוכנית הדגימה המוצעת. על החוקר להבין את השיקולים לקביעת גודל המדגם, וכיצד יוצגו התוצאות במידה והשערת המחקר תאוּש, ובמידה שלא תאוּש.

בסופו של תהליך זה, עליכם לקבל מהסטטיסטיקאית מסמך ובו קביעה מנומקת של גודל המדגם. בפרוטוקול לניסוי קליני נדרש לתעד את השיקולים לקביעת גודל המדגם, ולתמוך בה על ידי הדגמת היחס בין העוצמה הסטטיסטית, האפקט המינימלי וגודל המדגם, כל זאת תוך התחשבות בהקשר בו נערך המחקר — כפי שנפרט בהמשך.

בחירת הנוסחה לחישוב גודל המדגם

תכנון המחקר נועד לקבוע את הדרך היעילה ביותר לבחון את שאלת המחקר. מה הוא המשתנה הנמדד (האם רציף או בדיד, האם נורמלי?). איזה מבנה מחקרי יעיל ביותר, האם יש להשוות בין קבוצות? או להדגים קשר מתאמי כמנה-תגובה? האם יהיו במחקר חזרות, או מעקב לאורך זמן? האם יש משתנים נלווים (co-factors)? כל אלו ישפיעו על בחירת הכלים הסטטיסטיים לניתוח התוצאות, מכאן על הנוסחה לחישוב גודל מדגם.

לנתוחים סטטיסטיים פשוטים קיימות נוסחאות פשוטות. אבל כל חריגה מתכנון ניסוי פשוט, דורשת נוסחה מורכבת יותר ע"י ישום של סטטיסטיקה ברמה גבוהה ביותר. לרוב גם לא נמצאת נוסחה ידועה; למשל, אם ההתפלגות המצופה אינה נורמלית, הנוסחאות המקובלות לא מתאימות. במקרים כאלה נהוג להשתמש בסימולציות. במקרים אחרים, משתמשים בנוסחה מקובלת תוך כדי ציון הסטיה ושערוך כיוון הסטיה.

תוכנית הדגימה יכולה להיות קריטית. ללא ספק, זו הנקודה החשובה ביותר בייעוץ הסטטיסטי. האופן בו מגויסים המשתתפים, והאופן בו מוגדרות הקבוצות והחזרות יכול לקבוע האם בנתונים שנאספו ניתן לענות על שאלת המחקר. ייעוץ סטטיסטי נחוץ כדי להתאים תוכנית דגימה המאפשרת לענות על שאלת המחקר בצורה היעילה והברורה ביותר.

פניה מאוחרת לייעוץ סטטיסטי לא תתקן תכנון לקוי. מחקרים בעלי תוכנית דגימה עלובה מפיקים נתונים שאינם מאפשרים לענות על שאלת המחקר. הדבר העצוב ביותר בתוכנית דגימה שגויה הוא שגם המשך גיוס של משתתפים לא מתקן את המצב, ועיבוד סטטיסטי, מורכב ככל שיהיה, יתקשה להציל את המחקר.

הפרמטרים בנוסחת הישוב גודל מדגם

הפרמטרים שנלקחים בחשבון בנוסחה הנם הסתברותיים: מובהקות, עוצמה ופיזור, וניתן לאמוד את גודלם רק באופן סטטיסטי. לגבי כל אחד מהפרמטרים, נהוג לבדוק את טווח הפרמטר ולהתרשם מהרגישות של גודל המדגם המחושב לשינוי בפרמטר. בדיקה כזו מגלה את הטווח בו גודל המדגם רגיש מאד להגדרת הפרמטרים הללו, ובדרך כלל מוצגת גרפית בעת קביעת גודל המדגם. ההמלצה היא להימנע ממצב שבו שינוי זעיר בפרמטר מביא לשינוי גדול בגודל המדגם. יש להציג בדיקות רגישות אלו במסמך המתעד את קביעת גודל המדגם.

מובהקות סטטיסטית (α) – הינה ההסתברות לאשש את שאלת המחקר, למרות "שבאמת" לא היה צריך לאשש. לדוגמה, מצאנו הבדל מובהק בין קבוצת טיפול לקבוצת בקורת, אבל כשחזרנו על הניסוי עוד פעמים רבות גילינו שהתוצאה לא חוזרת על עצמה כלל והתוצאה הראשונית הנה חריגה ולא מייצגת. מקובל לאפשר טעות מסוג זה (נקראת גם טעות מסוג ראשון) בהסתברות של 5%, או $\alpha=0.05$. ז"א שסטטיסטית, מכל 100 ניסויים שבצענו, 5 יראו תוצאה חיובית למרות שבאמת לא קיים הבדל.

עוצמה סטטיסטית ($1-\beta$) – הינה ההסתברות לאשש את שאלת המחקר, כאשר אכן יש לאשש את שאלת המחקר. בעצם, β (טעות מסוג שני) הנה ההסתברות לאשש השערת המחקר כאשר "באמת" היה צריך

לאששה. טעות זו משמעותה לא למצוא הבדל בין קבוצת טיפול לביקורת, כאשר ההבדל קיים. מקובל לדרוש עוצמה סטטיסטית של 80% - 95%, (ז"א טעות מסוג שני, β של 20%-5%). באופן אינטואיטיבי, מדגם קטן מדי יהיה בעל עוצמה נמוכה, ז"א בעל יכולת נמוכה לעשש השערת מחקר, כאשר היא "האמת הנכונה". אותה השערה תאושש אם תבדק במחקר עם מספר רב יותר של נבדקים, ז"א בעל עוצמה גבוהה יותר.

אפקט קליני (δ) – הינו ההבדל בין הקבוצות או מידת המתאם שנצפה לראות במחקר, ויהיה בעל משמעות קלינית. גודל המדגם תלוי בגודל האפקט הקליני: זיהוי של אפקט קליני גס דורש מדגם קטן, בעוד זיהוי של אפקט זעיר דורש מדגם גדול יותר המאפשר רגישות רבה יותר. לדוגמה כאשר נרצה להדגים שתוכנית דיאטה מסייעת להורדה במשקל, נברר: מה ייחשב תוצאה חיובית? האם ירידה ממוצעת של 3 ק"ג במשך 3 חודשים, או האם ירידה ממוצעת של 30 גרם באותה תקופה? מקובל לדון עם החוקר והקלינאי ולהגדיר מראש איזה אפקט הוא מצפה לראות, והאם לאפקט זה יש משמעות קלינית.

מידת הפיזור של המשתנה הנמדד (σ^2) – הינה מדד להומוגניות של הנבדקים ולחזרתיות של המשתנה הנמדד בתוך כל קבוצה. אם קיים פיזור גבוה קשה לזהות הבדל משמעותי בין הקבוצות, ולכן ידרש מדגם גדול יותר. הפיזור יכול לנבוע מגורמים רבים: דיוק כלי המדידה, גורמים נלווים המשפיעים על המדידה, ומידת ההטרוגניות של קבוצת הנבדקים.

הנוסחאות הקלסיות לחישוב גודל מדגם מתיחסות לארבעת הפרמטרים הללו בלבד. בדרך כלל α ו- $(1-\beta)$ נקבעים לפי המקובל, $\alpha = 5\%$ ו- $(1-\beta) = 85\%$. בעוד שמידת הפיזור- σ וגודל האפקט- δ נקבעים על סמך מחקרים קודמים. יש הנוהגים לבטא את גודל האפקט ביחידות של σ ובכך לבטל את הצורך להגדיר בנפרד את הפיזור וגודל האפקט.

תיקונים בחישוב

כעת חושב גודל המדגם, על בסיס הנוסחה המתאימה והפרמטרים ההסתברותיים. אולם בשטח ישנן מספר בעיות אשר עלולות לאלץ אותנו לתקן את המספר ההתחלתי. האם נתן לגייס מספר שווה של נבדקים לשתי קבוצות המחקר? האם יש השוואות מרובות? מה מידת הנשירה? וכדומה.

השוואות מרובות: במידה ונעשות השוואות מרובות במחקר, ואנו מעוניינים לבחון יותר מהשערה אחת, עלינו לתקן את רמת המובהקות למספר ההשוואות. נניח שמחקר בודק השפעת טיפול בהשוואה לביקורת על פי ירידה במשקל ושיפור באיכות חיים. ירידה במשקל נמדדת ב- ק"ג, בהפחתת אינדקס מסת הגוף, ואחוז שומן. בנוסף איכות חיים נמדדת על פי מדד דכאון, והשתתפות בפעילות חברתית. בדוגמה זו מנינו כ- 5 מדדי תוצאה שלפיהם נבחן את השפעת הטיפול. במקרה זה, אם כל בדיקה תערך במובהקות של 5% אז בעצם, באופן אינטואיטיבי, רמת המובהקות הכללית עלתה ל- 25%, כי ההסתברות לאשש את המחקר בטעות כאשר חוזרים על המחקר 100 פעמים היא 5% עבור כל מדד, וסה"כ 25%.

במקרה של השוואות מרובות נפעיל תיקון לפי כלל אצבע האומר לחלק את סה"כ המובהקות - 5%, למספר ההשוואות. במידה ונערכו 5 השוואות, כל השוואה תעשה ברמת מובהקות של 1%, וסה"כ רמת המובהקות של המחקר כולו נשמרת על 5%. זה כמובן גורם לעליה משמעותית בגודל המדגם, ולכן לעיתים מקובל להקצות את סה"כ המובהקות למשתנה עיקרי אחד, ויתר המדדים במחקר נחשבים משניים. רשויות רגולטוריות מקפידות לבדוק שהמשתנה העיקרי (ה- endpoint) יצוין א-פריורי בפרוטוקול, ושהשיקולים הסטטיסטיים הולמים את תכנון המחקר.

נשירה- במחקרים רבים לא כל המשתתפים יסיימו את המחקר, מסיבות שונות. חשש במקרה כזה הוא שהמספר יקטן ונאבד מהעצמה הסטטיסטית הנדרשת. מענה לחשש זה ניתן על ידי תיקון גודל המדגם עבור נשירה. באופן כללי מקובל להניח 15%-25% נשירה, אך מידת הנשירה שונה מאד בין דיסציפלינות שונות, ויש להעריך את שיעור הנשירה המתאים. עצם ההתייחסות לנושא הנשירה מוסיף אמינות למחקר. גם הניתוח הסטטיסטי נדרש לנושא הנשירה: משווה בין הנושרים והמסיימים, ובודק האם הנשירה זהה בקבוצות הטיפול והביקורת. למשל מקרה בו הטיפול גורם לנשירה ורק קבוצה קטנה מצליחה לשרוד את הטיפול עד סופו יוביל לסטיה בהשוואה בין הקבוצות. המענה כאן הוא בעיקר ברמת הניתוח: יש להתייחס בניתוח גם לנושרים על בסיס כוונת הטיפול (Intent to Treat). מכאן שיש לתעד את הנושרים באותה מידה של קפדנות בה מתועדים הנבדקים שסיימו.

הקונטקסט המחקרי:

ההקשר בו נערך המחקר ומחקרים קודמים בתחום הספציפי ישפיעו במידה רבה על קביעת גודל המדגם.

מטרת המחקר: אנו מבחינים בין מחקר אישור למחקר גישוש (Confirmatory או Exploratory). האם המחקר נועד לאשרר תוצאות והנחות שכבר נבדקו במחקרים קודמים או שזה מחקר גישוש הבודק היתכנות. במחקר גישוש לרוב לא יודעים במדויק את מידת הפיזור של המשתנה הנמדד, ולא מתקנים עבור השוואות מרובות, מכין שאנו רוצים לקבל יותר תוצאות ולכן מתפשרים על המהימנות. במחקר אשרור, עלינו לוודא את כל הפרטים המוקדמים ולרכז את כל רמת המובהקות בשאלת המחקר העיקרית.

האג'נדה המחקרית: במסגרת של מחקרים קליניים נבדוק באיזה שלב מדובר (Phase I, II, III, IV). בשלבים המוקדמים ניתן להסתפק במדגם קטן יותר. כדאי לבדוק אם במקביל מתקיימים מחקרים דומים שניתן לשלב ביניהם. נברר אם המטרה היא פרסום אקדמי או דרישה של מחקר לאישור של רשויות בריאות ופיקוח. כל אלה גורמים שנלקחים בחשבון בתכנון המחקר, תוכנית הדגימה וקביעת גודל המדגם.

מה מקובל בתחום: בסופו של דבר, התחום הרפואי בו מתבצע המחקר יקבע את המסגרת וסדרי הגודל המקובלים. במחקרים העוסקים בנתוחים כירורגיים מורכבים לרוב נראה מדגמים קטנים - 10-15 בכל קבוצה. גם במחקרים של מחלות נדירות לא מתאפשרים מדגמים גדולים. במחקרים בהם קל יחסית לאסוף מידע רב בלי התערבות פולשנית נראה מדגמים גדולים יותר. תחומי מחקר עם הרבה משתנים מתערבים וקבוצות הטרוגניות דורשים את המדגמים הגדולים ביותר, לעיתים בסביבות 100-200 נבדקים בקבוצה. על החוקר להכיר את התחום בו הוא פועל, ולבדוק מה גדלי המדגמים במחקרים דומים המפורסמים בעיתונות המקצועית.

התקציב: ומעל הכל, מסגרת התקציב ומסגרת הזמן קובעים את תכנון המחקר וגודלו. בתקציב מצומצם יש לצמצם את היקף השאלה המחקרית. לדוגמה, ניתן לשקול התרכזות בתת קבוצה הומוגנית.

דיון

סקרנו כאן בקצרה את התהליך לקביעת גודל המדגם, והגורמים המשתתפים בו. הסטטיסטיקאית בודקת עם החוקר את תכנון המחקר האופטימלי כאשר כל הגורמים שצוינו נלקחים בחשבון. בדיקה זו אמורה להביא לתיאום ציפיות בין ההקשר המחקרי, התקציב, והתוצאה האפשרית. בראש ובראשונה, הסטטיסטיקאית מוודאת שניתן יהיה לענות על שאלת המחקר בכלים סטטיסטיים. במקרים מסוימים יתכן שהמסקנה תהיה לא להתחיל הניסוי כלל, וברוב המקרים החוקר יבין טוב יותר למה הוא מצפה בניסוי, איזה משקל מדעי יהיה לתוצאות מהניסוי שלו, ומה מגבלות המחקר. היעוץ הסטטיסטי אמור להניב מסמך המתאר את השיקולים בבחירת גודל המדגם.

אמנם ניתן למצא באינטרנט פרוצדורות רבות לחישוב גודל המדגם, אך חשוב לזכור שקביעת גודל מדגם אינה מסתכמת בהצבה בנוסחה ובתוצאה מספרית. זוהי תחנה חשובה בתהליך המחקרי שנועדה לאשר את תכנון המחקר. ההמלצה הישנה והמוכרת לכולנו, **תמיד להתייעץ עם הסטטיסטיקאית לפני תחילת הניסוי** תחסוך הרבה זמן, כסף ויכולה להיות ההבדל בין הצלחה לכשלון.

מקורות:

Zar J. H. Biostatistical Analysis 4th edition, Prentice hall, 1999

ICH E-9 [Statistical Principles for Clinical Trials](#)

דקלה גבע, בעלת תואר שני בביוסטטיסטיקה מאוניברסיטת פיטסבורג (פנסילבניה, ארה"ב), ותואר שני בחדשנות ויזמות של ISEMI (שלוחה בישראל של אוניברסיטת סווינבורן, אוסטרליה). לדקלה נסיון עשיר של יותר מ-15 שנה בביוסטטיסטיקה בתחום הניסויים הקליניים, בפרוייקטים מגוונים במרכזים רפואיים, בתעשייה, ובאקדמיה. (לפרטים נוספים: www.integrstat.com)