
עקרון האקראיות וההתחשבות במשתנים מתערבים בניסוי – האם אנו עושים זאת נכון?

גל זיו, רוני לידור, יעל נץ

המכללה האקדמית בוינגייט

תקציר

אחד הגורמים המאיימים על יכולתו של החוקר במדעי הספורט והתנועה להצביע על קשר סיבתי בין המשתנה הבלתי-תלוי (המשתנה המסביר; הסיבה) למשתנה התלוי (המשתנה הנמדד; המוסבר) בניסוי (experiment) שהוא עורך, הוא אי-שוויון בין קבוצות המחקר למרות החלוקה האקראית. במילים אחרות, על החוקר לוודא שאת ההבדלים שנמצאו בקבוצות הניסוי במשתנה התלוי יהיה אפשר לייחס להבדלים במשתנה הבלתי-תלוי שהופעל עליהן ולא להבדלים שהיו קיימים ביניהן מראש במאפייני המשתתפים. מאמר זה מחזק את הטענה שחלוקה אקראית של משתתפים לקבוצות – אחד מ"כללי הברזל" בתכנון ניסוי – אין פירושה בהכרח שוויון בין הקבוצות. המאמר אף בוחן כמה דרכים שעל החוקר לנקוט כדי להתמודד עם האיום של חלוקה לא הולמת של המשתתפים בניסוי שהוא עורך. ביתר פירוט, ארבע מטרות למאמר זה: (א) להגדיר מהי חלוקה אקראית של משתתפים לקבוצות הניסוי ולדון בבעייתיות ביישום חלוקה זו; (ב) לבחון חלוקה אקראית חלופית של המשתתפים בניסוי – החלוקה האקראית הרבודה ואופן השימוש בה; (ג) לבחון כמה אסטרטגיות המסייעות לחוקר לטפל במשתנים מתערבים בניסוי; (ד) להמליץ על כמה שלבים לתכנון החלוקה לקבוצות ולטיפול במשתנים מתערבים. המחברים קוראים לחוקרים לשכלל את הליך הבחירה האקראית הנהוג במדעי הספורט והתנועה כדי לחזק את יכולתם להסביר את הקשר הסיבתי בין משתני הניסוי.

תאריכים: ניסוי, גורמים מאיימים, תוקף פנימי, חלוקה אקראית רבודה.

אחת משיטות המחקר השכיחות במדעי הספורט והתנועה (כמו ביומכניקה, למידה מוטורית, פיזיולוגיה של המאמץ, פסיכולוגיה של הספורט) היא הניסוי (experiment) (ראו: Gratton & Jones, 2016; Amonette, English, & Kraemer, 2016). הניסוי הוא עריכה מכוונת של חומרים או נבדקים בתנאים ובמצבים מסוימים, במטרה לאשר או לשלול הנחה מעשית או מדעית משוערת. ביתר פירוט: ניסוי הוא שיטת מחקר, שבעזרתה בודק החוקר השערה בדבר קשר סיבה ומסובב בין הגורמים הנחקרים, וזאת באמצעות השוואה בין שני מצבים (או יותר) (בייט-מרום וספורטה, 2005). במצב אחד החוקר מחולל שינוי משמעותי ויזום במערכת הנחקרת (קרי, מניפולציה) ובמצב השני החוקר אינו מחולל שינוי כזה. העובדה שהשינוי המשמעותי במערכת נעשה ביוזמת החוקר מאפשרת לדעת מה הם קשרי הסיבה והמסובב במערכת הנחקרת.

מטרתו העיקרית של החוקר בעריכת ניסוי היא להצביע על קשר סיבתי בין שני משתנים – משתנה בלתי-תלוי ומשתנה תלוי. המשתנה הבלתי-תלוי הוא זה שבוחר החוקר כדי לבחון את השפעתו על המשתנה התלוי, כלומר, הוא אותו גורם (הסיבה) המסביר ומנבא את המשתנה התלוי. המשתנה התלוי הוא זה שאותו החוקר מודד ומעוניין להסביר. המשתנה התלוי הוא גם המדד (או אחד מכמה מדדים אפשריים), שנבחר כדי לשקף תכונה/יכולת. השינויים החלים במשתנה התלוי מיוחסים למשתנה הבלתי-תלוי (Thomas, Nelson, & Silverman, 2015). כך לדוגמה, בניסוי אחד חוקר בלמידה מוטורית מעוניין לבחון את ההשפעה של הדגמה (modeling) על אחוזי הקליעה לסל בכדורסל בקרב שחקני נוער. המשתנה הבלתי-תלוי הוא ההדגמה בניסוי זה (קבוצה אחת תלמד לקלוע לסל עם הדגמה, וקבוצה שנייה תעשה זאת ללא הדגמה). המשתנה התלוי (המשתנה הנמדד) הוא אחוזי הקליעה לסל. הנחת החוקר בניסוי זה היא שההדגמה תסייע בשיפור אחוזי הקליעה לסל בקרב השחקנים הצעירים. כלומר, הוא מניח שההדגמה היא הסיבה לשיפור שחל באחוזי הקליעה של המשתתפים.

בניסוי אחר, חוקרת בפיזיולוגיה של המאמץ מעוניינת לבחון את ההשפעה של אימון אירובי עם דלגית (משתנה בלתי-תלוי) על יכולת הקשב (משתנה תלוי) של תלמידים עם הפרעות קשב. החוקרת מניחה שתלמידים שישתתפו באימונים אירוביים יפיקו מהם תועלת ויהיו קשובים יותר (יקבלו ציון גבוה יותר במבחן קשב). כלומר, לטענת החוקרת, הפעילות האירובית היא הסיבה לשיפור שיחול במדדי הקשב. בניסוי נוסף, חוקר מעוניין לבחון את ההשפעה של אימון כוח (משתנה בלתי-תלוי) על זריזות רגליים בקרב טניסאים צעירים ובוחר בריצת הלוך ושוב – שני מקטעי ריצה, כל אחד במרחק 5 מ' – כמדד (משתנה תלוי), שנועד לשקף את יכולת הזריזות בקרב הטניסאים.

כדי לבסס קשר סיבתי בין המשתנה הבלתי-תלוי לבין המשתנה התלוי על החוקר לאמץ מערך ניסוי, שיאפשר לו שליטה מרבית במשתני הניסוי. זאת ועוד, על החוקר לתת את הדעת על גורמים מאיימים פוטנציאליים, שעלולים למנוע ממנו לדווח שהמשתנה הבלתי-תלוי הוא אכן הסיבה לשינוי שחל במשתנה התלוי. במילים

אחרות, על החוקר לדחות הסברים אפשריים אחרים העולים בדעתו בנוגע לקשר הסיבתי בין המשתנה הבלתי-תלוי למשתנה התלוי. על החוקר להביא בחשבון הסברים אלו כבר בשלב התכנון של הניסוי.

חוקרי החינוך קמפבל וסטנלי (Campbell & Stanley, 1963; ראה גם: בייט-מרום וספורטה, 2005; Thomas et al., 2015) הצביעו על שמונה גורמים פוטנציאליים המאיימים על הקשר הסיבתי המוצע בין המשתנה הבלתי-תלוי והמשתנה התלוי. בעבודה מאוחרת יותר הצביעו קוק וקמפבל על חמישה גורמים מאיימים נוספים (Cook & Campbell, 1979). כך לדוגמה, בין הגורמים המרכזיים שהוצעו על ידי החוקרים ניתן למנות את הבאים:

- א. ההיסטוריה של המשתתפים בניסוי – כל המאורעות שאינם קשורים לניסוי הקורים בין מדידת המשתנה התלוי לפני הטיפול הניסויי, קרי, הפעלת המשתנה הבלתי-תלוי, לבין מדידת המשתנה התלוי אחרי הטיפול.
- ב. בשלות – התהליך שעוברים משתתפי המחקר במהלך הניסוי כפונקציה של מעבר הזמן כמו התפתחות וצמא.
- ג. מכשור – השינויים שחלים במכשירי המדידה בין שתי מדידות. לדוגמה: לפני הטיפול ואחריו.

ד. נשירה – עזיבה (לא שווה) של משתתפים מקבוצות הניסוי השונות. אם גורמים אלו אינם נשלטים על-ידי החוקר הם עלולים להשפיע על מערכת היחסים בין המשתנה הבלתי-תלוי והמשתנה התלוי ולהחליש את הקשר הסיבתי המוצע בין שני המשתנים. במילים אחרות, קמפבל וסטנלי טוענים שהעדר מתן מענה לגורמים מאיימים אלו כבר בעת תכנון הניסוי עלול להחליש את התוקף הפנימי (internal validity) של הניסוי, כלומר, את יכולתו של החוקר להצביע על המשתנה הבלתי-תלוי כסיבה לשינוי שחל במשתנה התלוי.

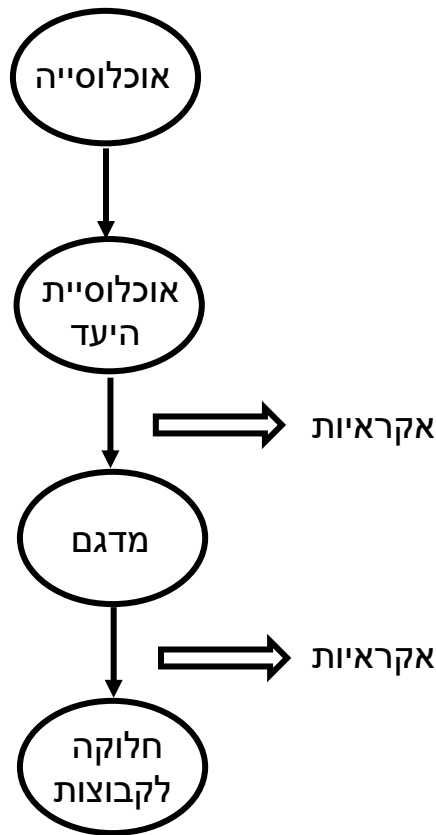
גורם נוסף שעלול לאיים על הקשר הסיבתי בין המשתנה הבלתי-תלוי למשתנה התלוי המוצע על-ידי החוקר בניסוי, ובו נתמקד במאמר זה, הוא גורם חלוקת המשתתפים לקבוצות הניסוי. אם החוקר אינו מקיים הליך נכון של חלוקת המשתתפים לקבוצות בניסוי שהוא עורך, הרי את ההבדלים בין קבוצות הניסוי יהיה אפשר לייחס להבדלים שהיו קיימים מראש במאפייני המשתתפים ולא להבדלים במשתנה הבלתי-תלוי שהופעל עליהם במהלך הניסוי עצמו. אי לכך, על החוקר לנקוט הליך חלוקה כזה שאמור לסייע לו לספק תשובה חיובית לשאלה – האם המשתנה הבלתי-תלוי (אימון/טיפול/תכנית התערבות) הוא אכן הסיבה לשינוי שחל במשתנה התלוי.

ארבע מטרות למאמר זה: (א) להגדיר מהי חלוקה אקראית של משתתפים לקבוצות הניסוי ולדון בבעייתיות ביישום חלוקה זו. (ב) לבחון חלוקה אקראית חלופית של המשתתפים בניסוי – החלוקה האקראית הרבודה, ואת אופן השימוש בה. (ג) לבחון כמה אסטרטגיות המסייעות לחוקר לטפל במשתנים מתערבים בניסוי. (ד) להמליץ על שלבי התכנון של החלוקה לקבוצות וטיפול במשתנים מתערבים.

בהקשר זה, משתנה מתערב עשוי להיות גם משתנה תלוי אחד, או כמה, שבהם נמצא הבדל התחלתי בין קבוצות הניסוי.

עקרון האקראיות

אחד העקרונות המרכזיים בתכנון של ניסוי הוא עקרון האקראיות (randomization): על החוקר לבחור אקראית את המדגם המשתתף בניסוי מאוכלוסיית היעד שהגדיר. לדוגמה: ספורטאי עילית בכדורסל, נשים בהיריון העוסקות בפעילות גופנית אירובית או ילדות השוחות שלוש שנים במסגרת תחרותית (Thomas et al., 2015). משום שמטרתו של החוקר בעריכת ניסוי היא להצביע על קשר סיבתי בין המשתנה הבלתי-תלוי למשתנה התלוי, עליו לאפשר לכל אחד/אחת מאוכלוסיית היעד סיכוי שווה להשתתף בניסוי שהוא עורך. כך יוכל החוקר לבחון את התרומה של המשתנה הבלתי-תלוי לזה התלוי, ללא הטיה בתהליך בחירת המשתתפים למחקר. תהליך זה הוא גם חיוני, אם החוקר מעוניין לייצג במדגם את כלל אוכלוסיית היעד כדי שיוכל להכליל את ממצאיו מעבר למדגם זה. נוסף לכך, על החוקר לחלק את משתתפי המדגם לקבוצות באופן אקראי [בחרנו להשתמש במונח **קבוצות**, משום שזהו המונח המקובל בספרות המבטא הקצאה טכנית של משתתפי הניסוי לקבוצות – ניסוי וביקורת (ראו: Thomas et al., 2015)]. עם זאת, ספרות המחקר העוסקת במתודולוגיה משתמשת אף במונחים נוספים כמו **תנאי הניסוי ו-סוגי ההתערבות** כדי לתת ביטוי למשתנים הבלתי-תלויים שאת תרומתם בוחן החוקר בניסוי (ראו איור 1 – תהליך הבחירה וההקצאה לקבוצות). החלוקה האקראית לקבוצות מאפשרת לכל משתתף סיכוי שווה להיכלל באחת מקבוצות הניסוי. החלוקה האקראית לקבוצות אמורה ליצור מצב שבו המשתנים המאפיינים את המשתתפים בקבוצות המחקר דומים, וכך ניתן יהיה לשייך הבדלים שיימצאו בין הקבוצות בתום הניסוי לטיפול או למניפולציה שעורך החוקר ולא לגורמים/משתנים אחרים. שיוך ההבדלים למשתנה הבלתי-תלוי יסייע לחוקר לספק ראיות מוצקות (evidence-based) לאיכות האימון/טיפול/תכנית ההתערבות שבחן בניסויו. ללא יישום של חלוקה אקראית לקבוצות יתקשה החוקר להצביע על המשתנה הבלתי-תלוי כסיבה לשינוי שחל במשתנה התלוי.



איור 1. עקרון האקראיות בניסוי

אולם החלוקה האקראית הפשוטה (simple randomization) לקבוצות, שבה לכל משתתף סיכוי שווה להיכלל באחת מקבוצות הניסוי, אינה מביאה תמיד למצב שבו הקבוצות אינן שונות במשתנים שלהם פוטנציאל להטות את תוצאות המחקר גדולים ($n > 200$), החלוקה האקראית הפשוטה מבטיחה באופן מוחלט כמעט שהקבוצות אכן תהיינה דומות במגוון רב של משתנים, אולם במדגמים קטנים ($n < 200$), אין הדבר כך (Kernan et al., 1999).

בדרך כלל מספר המשתתפים בניסויים במדעי הספורט והתנועה, המשווים בין שתיים עד ארבע קבוצות מחקר, קטן משמעותית מ-200: בדרך כלל, בכל קבוצה ישנם כ-20-25 משתתפים. אם כך, חשוב להעריך מראש את הסיכוי שבמדגמים קטנים אלו תהיינה קבוצות המחקר שונות זו מזו במשתנים שעלולים

להשפיע על ממצאי המחקר או על הפרשנות שלהם. לכן, על-פי ההנחיות כיצד לדווח על מחקרים קליניים (CONSORT - Consolidated Standards of Reporting Trials, 2010) מומלץ להשתמש בשיטות חלוקה אקראיות אחרות כאשר גודל המדגם קטן. אחת משיטות אלו היא החלוקה האקראית הרבודה (stratified randomization) (Moher et al., 2010). שיטת חלוקה אקראית זו מאפשרת לחוקר, גם במדגמים קטנים, להגיע למצב שמשתנה מתערב אחד או שניים יאפיינו בצורה דומה את המשתתפים בקבוצות הניסוי השונות, דבר שהיה מתרחש מעצמו בחלוקה אקראית פשוטה במדגמים גדולים.

בעיית הדגימה האקראית הפשוטה במדגמים קטנים

כדי להתמודד עם האפשרות שדגימה אקראית פשוטה לא תמנע הבדלים בין קבוצות המחקר השונות באותם משתנים מתערבים הרלוונטיים לניסוי, נכתב על-ידי המחבר הראשון של מאמר זה קוד בשפת התכנות הסטטיסטית R (R Core Team, 2013). התוכנה היא חופשית לשימוש (קוד פתוח, open source) ופופולרית בקרב אנשי אקדמיה ותעשייה כאחד. באמצעות קוד זה ניתן להריץ מספר רב של סימולציות ניסוי הכולל שתי קבוצות, ובו התוצאות של מספר המשתתפים המוגדר על-ידי המשתמש הן אקראיות (בדומה למצופה מחלוקה אקראית לקבוצות). בדוגמאות שיובאו להלן, הסימולציה מייצרת תוצאות אקראיות של 1,000 ניסויים ומספקת את אחוז הניסויים שבהם ההבדל בין הקבוצות גדול מהאחוז שהחוקר מגדיר כמקובל. ההבדל בין הקבוצות בתחילת המחקר ייחשב למשמעותי בהתאם להערכת החוקר את הסיכוי שהבדל זה עלול למנוע ממנו להצביע על קשר סיבתי בין המשתנה הבלתי-תלוי לזה התלוי. לדוגמה, ייתכן שבמחקר אחד יחליט החוקר שהבדל של 10% במשתנה מסוים בין קבוצות המחקר הוא משמעותי, ואילו במחקר אחר עשוי חוקר אחר להחליט שהבדל של 15% בין הקבוצות בתחילת המחקר הוא משמעותי, וזאת בהתאם לספרות קודמת ולניסיון החוקר עצמו. קוד R לחישוב הסיכוי לקבלת אי-שיוויון בין שתי קבוצות במשתנה אחד והוראות השימוש בו מוצגים בנספח 1.

כך לדוגמה, חוקר בפסיכולוגיה של הספורט מעוניין לבחון את ההשפעה של הנחיות למיקוד קשב על יכולת הקליעה לסל מטווח קצר של שחקני עילית בכדורסל. לשם כך הוא מתכנן ניסוי הכולל שתי קבוצות: קבוצה המבצעת אימון קליעות רגיל (ללא הנחיות קשב) ושנייה המבצעת אימון קליעות עם הנחיות קשב. על-פי הספרות המקצועית והניסיון של החוקר, אחד המשתתפים המתערבים הפוטנציאליים שעלול להשפיע על תוצאות הניסוי הוא הוותק של המשתתפים במשחק (מספר שנות משחק). שחקן ששיחק שנים רבות יותר הוא בעל ניסיון רב יותר בקליעה לסל, ולכן הישגיו עשויים להיות טובים יותר בהשוואה לשחקן ששיחק שנים מעטות יותר. במדגם שהחוקר עתיד לבחון ישנם משתתפים בעלי ניסיון משחק שבין שנה אחת ל-10 שנים. החוקר מתכנן לגייס 20 משתתפים לכל קבוצה ומעוניין לדעת מה הסיכוי שבדגימה אקראית פשוטה יהיו עדיין הבדלים במשתנה שנות המשחק בקרב המשתתפים בשתי קבוצות המחקר. כמו כן, החוקר מגדיר שאי-שוויון בין הקבוצות הגדול מ-20% עלול להשפיע על התוצאות. כאשר מריצים 1,000 ניסויים עם הגדרות אלו מגלים כי יש

סיכוי של 17-18% שיהיה הבדל הגדול מ-20% בין הקבוצות בשנות הניסיון במשחק הכדורסל.

בניסוי אחר, חוקרת בלמידה מוטורית מעוניינת לבחון את התרומה של שתי שיטות אימון – חדגוני ומגוון – לשיפור דיוק בחבטה בגולף (putting) בקרב שחקניות גולף מתחילות. החוקרת יודעת שיכולת החבטה הבסיסית של השחקניות עלולה להשפיע על התוצאות, ומקווה שהדגימה האקראית הפשוטה תגרום לשוויון בין הקבוצות ביכולת זו. החוקרת מתכוונת לגייס שתי קבוצות של 20 משתתפות בכל קבוצה, ומניחה שהבדל של יותר מ-15% ביכולת הביצוע ההתחלתי בין שתי הקבוצות יקשה עליה להסביר את התרומה של אחת משיטות האימון לשיפור הדיוק, אם אכן שיטה זו תימצא כמשפרת דיוק. מהרצת 1,000 ניסויים עם הנתונים האלו עולה שיש סיכוי של כ-12% שההבדל בין הקבוצות ביכולת הבסיסית יהיה גדול מ-15%.

בניסוי נוסף, חוקרת בפיזיולוגיה של המאמץ מעוניינת לדעת אם שתי תכניות אימון משפרות צריכת חמצן מרבית. בכוונתה לגייס 20 משתתפים בגילי 20-29 שנים לכל אחת מקבוצות הניסוי. בהתבסס על הספרות המחקרית, היא מניחה שטווח צריכת החמצן היחסית בגילים אלה הוא 35-51 מ"ל חמצן * ק"ג⁻¹ * דקה⁻¹. נוסף לכך, החוקרת מעוניינת שההבדל בין הקבוצות לא יהיה גבוה מ-10%. כאשר מריצים 1,000 ניסויים, מתברר שהסיכוי שההבדל בצריכת החמצן היחסית המרבית בין שתי הקבוצות יהיה יותר מ-10% הוא בין 0.3 ל-0.5%. אם חשוב לחוקרת שההבדל בצריכת החמצן המרבי בין הקבוצות לא יעלה על 5%, הסיכוי להבדלים בין הקבוצות עולה לכ-15%.

בשלוש הדוגמאות הנסקרות לעיל, החוקרים נדרשים להחליט אם דגימה אקראית פשוטה מספקת אותם, או שמא עליהם לנקוט שיטה אחרת של דגימה כדי להקטין את הסיכון שההבדל בין הקבוצות במשתנה התלוי, או באחד מהמשתנים המתערבים הצפויים בתחילת הניסוי, יהיה כזה שימנע מהם לענות על השאלה – האם המניפולציה שערך החוקר, קרי האימון/טיפול/תכנית ההתערבות שניתן למשתתפים, הוא אכן הגורם שהביא לתוצאות שהתקבלו בניסוי. מהסימולציות הסטטיסטיות שבוצעו והודגמו קודם לכן ניתן לראות שלמעט מקרה אחד (הבדל של 10% בצריכת חמצן מרבית בקרב משתתפות בטווח גילים 20-29 שנים), הסיכון לאי-שוויון בין קבוצות הניסוי הוא בין 10 ל-15%. סיכון זה אינו מבוטל, וייתכן שכדאי לנקוט שיטת דגימה אחרת.

דוגמאות נוספות של מקרים בניסויים במדעי הספורט והתנועה, שבהם צפויים הבדלים בין קבוצות הניסוי בשלב ההתחלתי שלהם, מוצגות בלוח 1. בשני ניסויים המתוארים בלוח, שמטרתם היא שיפור יכולת של ניתור אנכי בקרב שחקני כדורגל בכיתה י"ב ושיפור בצריכת חמצן מרבית – דגימה אקראית פשוטה בהחלט עשויה להספיק לגבי המשתנה הרלוונטי הנמדד. עם זאת, בשאר המקרים המוצגים בלוח, בהתאם להפרש הרצוי בין קבוצות הניסוי בנקודת ההתחלה, על החוקר לשקול שימוש בשיטת דגימה אקראית אחרת. אחת מהשיטות החלופיות היא חלוקה אקראית רבודה (stratified randomization).

לוח 1.
דיוגמאות של ניסויים שבהם קיימים סיכויים שיהיו הבדלים התחלתיים בין שתי קבוצות
הניסוי לאחר תלוקה אקראית פשוטה לקבוצות

סיכון לאי-שוויון בין הקבוצות	טווח תוצאות צפוי	משתנה מתערב	מס' משתתפים	מטרת הניסוי
20% הובדל	15% הובדל	אפשרי	20	שיפור תבנית קצורה בגוף
10% הובדל	70-20 ס"מ מהגומה בממוצע 10 החבטות הראשונות	יכולת דיוק בסיסית	20	שיפור יכולת ניתור אנכי בקרב שחקני כדורגל בכיתה י"ב
3-4%	11-12% 30-35%	יכולת ניתור בסיסית	20	שיפור יכולת ניתור אנכי בקרב שחקני כדורגל בכיתה י"ב
0%	7-1.7%	יכולת ניתור בסיסית	20	שיפור יכולת ניתור אנכי בקרב שחקני כדורגל בכיתה י"ב
7-10%	20-22% 42-44%	שנות ניסיון במשחק	15	שיפור דיוק בבעיטת עונשין מ-11 מ'
1-3%	2-3% 14-15%	50-20 מספר תשובות נכללות במבחן ממוחשב	25	שיפור יכולת קוגניטיבית לאחר אמון גופני חד-פעמי
0%	0-1% 0-2%	1^{-1} *תומ' $kg^{-1} \cdot mlO_2$ צריכת חמצן מרבית בסיסית	15	שיפור בצריכת חמצן מרבית
2-4%	11-13% 29-31%	52-15 ק"ג *	20	לחיצה בשכיבה התחלתית
			20	לחיצה בשכיבה IRM בקרב ילדים גילאי 9-12

* הנתונים נלקחו מ-Hoffman (2006)

חלוקה אקראית רבודה

חלוקה אקראית רבודה היא אחת האפשרויות המומלצות ליישום כאשר גודל המדגם קטן, וכתוצאה מכך הסיכון לאי-שוויון בין הקבוצות במשתנה אחד או שניים גדול באופן יחסי (Kernan et al., 1999; Moher et al., 2010). לדוגמה, אם חוקר מעוניין לבחון את ההשפעה של שתי תכניות אימון גופני על לחץ דם במנוחה, יש להניח שהבדלי הבסיס של המשתתפים בקבוצות הניסוי בכל הקשור לערכי לחץ הדם במנוחה יקשו על פירוש התוצאות. לכן, החוקר יכול להגדיר מראש כמה רבדים (strata), בהתבסס על מדידת לחץ הדם הבסיסי של המשתתפים בניסוי. כך לדוגמה, החוקר עשוי להגדיר שלושה רבדים של לחץ דם: (א) מתחת ל-110/70 מ"מ כספית; (ב) בין 110/70 ל-150/100 מ"מ כספית; (ג) מעל 150 מ"מ כספית. לחץ הדם של המשתתפים בניסוי ייבדק בתחילתו, וכל משתתף יוקצה לרובד (stratum) הרלוונטי. לאחר מכן, בכל רובד תבוצע חלוקה אקראית לקבוצות. חלוקה זו מבוצעת, בדרך כלל, על-ידי סדרות אקראיות של תמורות (permutations) על בסיס קבוצות הניסוי. הסדרות עשויות לכלול 2, 4 או 6 תמורות, ולעתים גם תכולת הסדרה יכולה להשתנות אקראית. לדוגמה, כאשר ישנן שתי קבוצות ניסוי – א ו-ב, תהיינה שש אפשרויות לסידורן בסדרה של ארבע תמורות: "אאבב", "בבאא", "אבאב", "באבא", "אבבא" ו-"באאב". כעת, בכל רובד תיבחרנה אקראית סדרות התמורות הללו, עד שהחוקר יגיע למספר המשתתפים בסדרה. אז, לפי סדר הגעת המשתתפים למחקר עצמו, הם משויכים לקבוצה הרלוונטית.

במקרה אחר, חוקר מעוניין לבחון את ההשפעה של הנחיות למיקוד קשב – פנימי וחיצוני – על עקיבות (שחזור) הביצוע של חבטת גולף לטווחים קצרים. לאחר חישוב הסיכוי למציאת הבדלים ביכולת הדיוק הבסיסית של הנבדקים שישתתפו במחקר, החוקר בוחר לבצע דגימה אקראית רבודה. לשם כך, 48 המשתתפים במחקר הגיעו למפגש ראשון שבו הם חבטו 10 פעמים בכדור גולף למטרה ניחת (גומה). החוקר מודד את עקיבות הביצוע של המשתתפים במפגש זה, לאחר מכן הוא מדרג את המשתתפים על-פי הישגיהם – מהמשתתף בעל העקיבות הנמוכה ביותר ועד לבעל העקיבות הגבוהה ביותר, ומגדיר שלושה רבדים – עקיבות נמוכה (16 משתתפים), עקיבות בינונית (16) ועקיבות גבוהה (16). לאחר שלב הדירוג, בוחר החוקר אקראית ארבע סדרות עם ארבע תמורות בכל אחת לכל רובד. למשל, לרובד הראשון הוא בוחר את ארבע הסדרות הבאות: "אבבא", "באבא", "אאבב" ו-"בבאא". המשתתפים מחולקים לקבוצות הניסוי על-פי תמורות אלו, כך שהמשתתף הראשון יהיה בקבוצה א, השני בקבוצה ב, השלישי ב-ב, הרביעי ב-א, החמישי בקבוצה ב וכך הלאה.

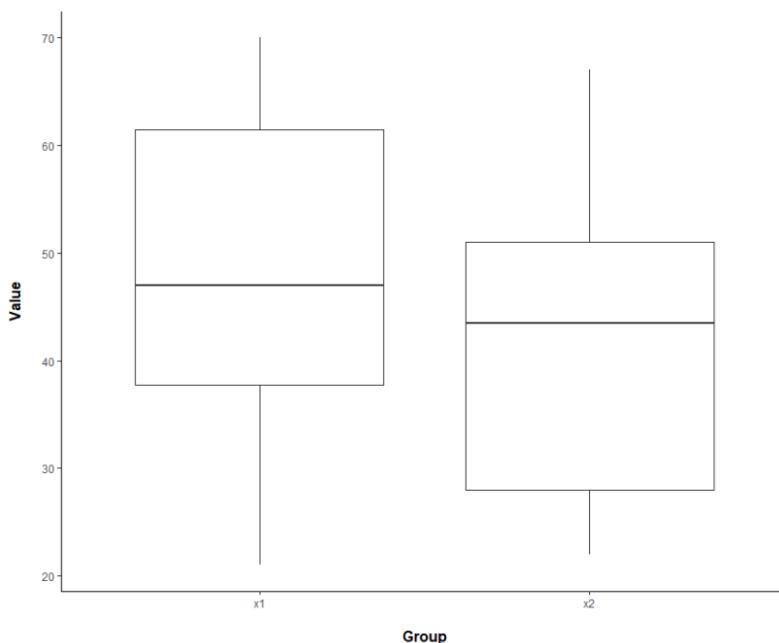
אחד היתרונות בחלוקה אקראית רבודה זו הוא שכל רובד מחולק אקראית לשתי קבוצות הניסוי, כך שאם יש די משתתפים בכל רובד ניתן למעשה לבחון את השפעת האימון/טיפול/תכנית התערבות בכל רובד בנפרד ואף לבחון את ההשפעה על כלל המשתתפים. יתרונות נוספים של החלוקה האקראית הרבודה הם הקטנת הסיכון לטעות מסוג I (Type I error); דחייה שגויה של השערת האפס. שגיאה כזו תגרום, בדרך כלל, למסקנה שאפקט מסוים קיים כשלמעשה אינו קיים). כמו כן,

לטעות מסוג II (Type II error); קבלה שגויה של השערת האפס. שגיאה כזו תגרום, בדרך כלל, לדחייה של אפקט מסוים, אף שהוא אכן קיים במציאות), ועל-ידי כך הגדלת העוצמה הסטטיסטית של הממצאים שעולים מהניסוי (Kernan et al., 1999). יש לציין שבכל רובד ניתן לחלק אקראית גם את סדרות התמורות וגם את המשתתפים עצמם. כך למשל, לאחר שהחוקר דירג את המשתתפים על-פי יכולת הביצוע ההתחלתית שלהם בגולף (עקיבות הביצוע), ניתן לחלק אותם לסדרות של התמורות האקראיות, או להקצות את המשתתפים באופן אקראי תחילה, ואז לחלקם לסדרות של התמורות האקראיות. במקרה הראשון, תהליך האקראיות מוגבל יותר, אולם השוויון במשתנה הנמדד יהיה חזק יותר. במקרה השני, האקראיות היא חזקה יותר, אולם השוויון במשתנה הנמדד עלול להיות חלש יותר. הדוגמאות הבאות תמחשנה את ההבדל בין שני המקרים: במקרה אחד ניתן לחלק 16 משתתפים לשני רבדים (שמונה משתתפים בכל רובד) ולדרגם על פי יכולת הביצוע ההתחלתית שלהם. כעת, נשתמש בסדרות של ארבע תמורות אקראיות לכל שמונה משתתפים. בדרך זו, למעשה, כל ארבעה משתתפים – הקרובים מאוד זה לזה ביכולת ההתחלתית שלהם – מחולקים אקראית לשתי קבוצות המחקר. באופן כזה, השוויון במשתנה הרלוונטי יהיה מוחלט כמעט. במקרה השני, אם לאחר דירוג המשתתפים והחלוקה לשני הרבדים, שמונת המשתתפים יעורבבו אקראית בכל רובד, ייתכן מצב שבו ארבעה משתתפים שעליהם תבוצע סדרת התמורות האקראית יהיו 1, 2, 7 ו-8. חלוקה זו עלולה להביא למצב שבו משתתפים המדורגים 1 ו-2 יהיו בקבוצה אחת, ואילו משתתפים 7 ו-8 יהיו בקבוצה שנייה. בדרך זו האקראיות אמנם גבוהה, אך ייתכנו הבדלים נוספים במשתנה שאותו אנו מעוניינים להשוות בין הקבוצות. עם זאת, בשני המקרים, ההבדלים בין הקבוצות במשתנה התלוי אינם אמורים להיות משמעותיים, כך שדרך החלוקה האקראית נתונה לשיקול דעתו של החוקר.

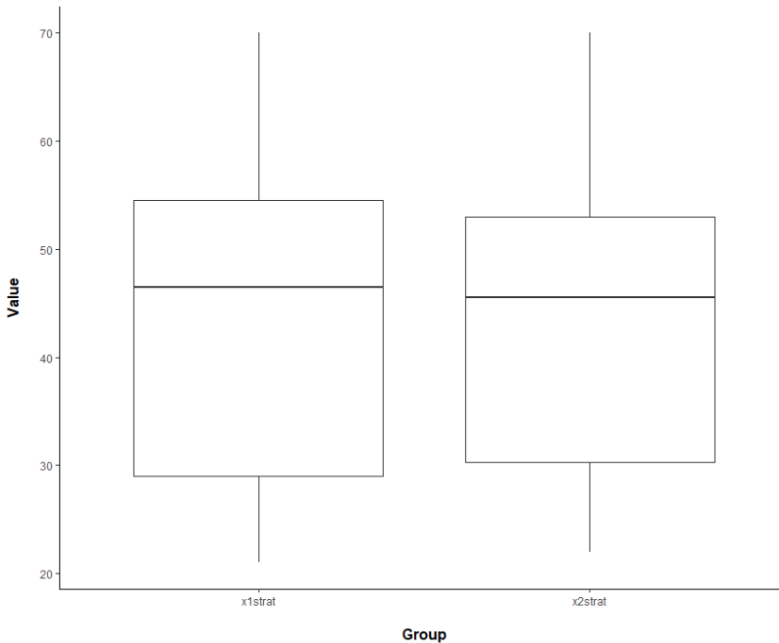
כאשר החלוקה לרבדים היא על-פי משתנים שמייים, מומלץ לחלק אקראית קודם את המשתתפים בכל רובד, משום שבכל מקרה, לסדר אין משמעות. קוד R לחישוב הסדרות של התמורות האקראיות לסדרות של ארבע תמורות, שש תמורות, או לבחירה אקראית של סדרות התמורות והוראות השימוש בו מוצגים בנספח 2.

דוגמה: חלוקה אקראית רבודה במחקר על חבטה בגולף. במחקר שמטרתו לבחון את תנועת העיניים (gaze behavior) בקרב שחקני גולף מתחילים, שלומדים כיצד לחבוט מטווח קצר, מעוניינים החוקרים שממוצע המרחק של 10 החבטות הראשונות יהיה דומה בקרב המשתתפים שיחולקו לשתי קבוצות למידה בנות 20 משתתפים כל אחת. בהרצת סימולציה של מדגם אקראי פשוט התקבל ממוצע של 48.3 ס"מ בקבוצה אחת ו-40.5 ס"מ בקבוצה שנייה. ההפרש של 7.8 ס"מ בין הקבוצות (כמעט 20%) הוא גדול מדי ויקשה על יכולתם של החוקרים להסביר את ממצאי הניסוי (איור 2 מציג את הממוצעים ופיזור התוצאות לאחר הרצת הסימולציה של מדגם אקראי פשוט). כדי לפתור את הבעיה השתמשו החוקרים במדגם אקראי רבוד עם סדרות של ארבע תמורות אקראיות. התקבל ממוצע של

44.25 ס"מ בקבוצה אחת ו-44.55 ס"מ בקבוצה שנייה. ההפרש הזניח של 0.3 ס"מ בין הקבוצות מעיד על רמת יכולת דומה של המשתתפים בתחילת הניסוי, עובדה שתסייע לחוקרים להסביר מדוע שיטת למידה מסוימת מטפחת למידה, אם אכן יתקבלו הממצאים המעידים על כך (הממוצעים ופיזור התוצאות לאחר הרצת הסימולציה של מדגם אקראי רבוד עם סדרות של ארבע תמורות אקראיות מוצגים באיור 3).



איור 2. ממוצעים ופיזור התוצאות לאחר הרצת סימולציה של מדגם אקראי פשוט



איור 3. ממוצעים ופיזור התוצאות לאחר הרצת סימולציה של מדגם אקראי רבד עם סדרות של ארבע תמורות אקראיות

השימוש בחלוקה אקראית רבודה – חסרונות. לשימוש בחלוקה אקראית רבודה במדעי הספורט והתנועה יש גם כמה חסרונות. כדי לבצע דגימה כזו עלינו למדוד, כבר לפני תחילת הניסוי, את המשתנה הרלוונטי לכל המשתתפים. כך נוכל לחלקם על-פי דירוגם במשתנה הרלוונטי הנמדד. ביצוע דגימה בשיטה זו מוסיף לניסוי המתוכנן מפגש אחד (לפחות), ובו יימדד המשתנה הרלוונטי. כמו כן, אם אחד המשתתפים פורש מהמחקר, איננו יודעים אם משתתף חדש שמחליף אותו מתאים לקבוצה או לא, משום שסביר להניח שהוא לא היה שותף לחלוקה האקראית הרבודה הראשונית. עם זאת, אם משתתפים בודדים פורשים במהלך הניסוי, יש סבירות נמוכה שהחלפתם במשתתפים חדשים אכן תוביל להבדלים משמעותיים בין הקבוצות. זאת מכיוון, שהחלוקה האקראית הרבודה הובילה לשוויון כמעט מוחלט בין הקבוצות. לכן, החלפת משתתף שפרש במשתתף חדש, שתוצאותיו ההתחלתיות יכולות אומנם להיות שונות מזה שאותו החליף, לא אמורה להוביל לאי-שוויון משמעותי בין הקבוצות.

אתגר מתודולוגי נוסף מתרחש כאשר חוקר במדעי הספורט והתנועה מוסיף מפגש מדידה לניסוי בו נבחנת יכולת הלמידה של המשתתפים. במקרה כזה הוא עלול להתמודד עם גורם הלמידה (learning effect) (Thomas et al., 2015). אם החוקר מעוניין לדרג את המשתתפים על-פי יכולתם ההתחלתית, עליו לבחון אותם בביצוע

מוטורי מסוים כמו זריקת חצים למטרה נייחת, וזאת כדי לבחון יכולת בסיסית של זריקה ביד הדומיננטית. אפשרות אחרת היא בעיטה בכף-הרגל הפנימית לשער קטן-ממדים כדי לבחון יכולת בסיסית של בעיטה ברגל הלא-דומיננטית. וכך, המשתתפים מתנסים (קרי, לומדים) במטלה עוד לפני תחילת הניסוי. ביצועים אלו של המטלה המוטורית עשויים להשפיע על תוצאות הניסוי, שכן, המשתתף מגיע לניסוי לאחר התנסות במטלה שהוא עתיד ללמוד בניסוי עצמו. יש לבחור, אפוא, את המטלות הנבנות כך שהוספת ביצוע מקדים לא תסייע למשתתפים להגיע לאפקט תקרה (ceiling effect; המשתתפים במחקר מגיעים להישגים גבוהים, ולא ניתן להצביע על שיפור נוסף ביכולתם במהלך הניסוי עצמו). מצד שני, במחקרים רבים, המשתתפים מקבלים ממילא הזדמנות לבצע היכרות עם המטלה עוד לפני תחילת הניסוי, כך שמפגש מקדים יכול גם לשמש מפגש היכרות עם מטלת הניסוי הנלמדת.

שליטה במשתנים מתערבים

ללא קשר לבחירה של סוג החלוקה האקראית לקבוצות, על החוקר לתת את הדעת, בשלב תכנון הניסוי, על אותם משתנים מתערבים שעלולים להשפיע על התוצאות שתתקבלנה בניסוי. לדוגמה, אם הניסוי עוסק בהשפעתה של פעילות גופנית (משתנה בלתי-תלוי) על רמת לחץ הדם (משתנה תלוי), יש להניח שרמת לחץ הדם של המשתתפים בתחילת הניסוי עלולה להשפיע על רמת לחץ הדם שלהם בתום הניסוי. בניסוי אחר שבו נבחנת השפעתה של פעילות גופנית על בריאות האדם, משתנים כמו גיל, סוג תזונה ועישון עלולים להשפיע על תוצאות הניסוי. החוקר חייב אפוא להביא בחשבון משתנים אלו כמשתנים מתערבים.

אחת האסטרטגיות שבהן יכול החוקר להשתמש כדי לשלוט במשתנים המתערבים היא להתייחס למשתנים אלו כ-covariates (משתנים שנכללים בניתוח הסטטיסטי עקב הקשר ביניהם לבין המשתנה התלוי. סטטיסטי, ניתן לשלוט במשתנים אלו כדי לאפשר בחינה של השפעת משתנים אחרים, רלוונטיים יותר, על המשתנה התלוי) בניתוח הסטטיסטי (Warner, 2008). טעות נפוצה היא לחשוב שיש להשתמש במשתנה כ-covariate רק כאשר נמצא הבדל מובהק לגביו בין קבוצות הניסוי. ייתכן שהבדלים במשתנה מסוים בין קבוצות הניסוי אינם מובהקים סטטיסטית, אולם השפעתו של המשתנה עשויה, בכל זאת, להיות משמעותית באשר לתוצאות הניסוי. לכן, יש להקפיד לבחור את המשתנים המתערבים כבר בשלב תכנון הניסוי (de Boer, Waterlander, Kuijper, Steenhuis, & Twisk, 2015; Enas, 1990; Enas, Spradlin, Wilson, & Wiltse, 1990). הבחירה במשתנים מתערבים נדרשת להתבסס על הספרות הקיימת ועל ניסיונו של החוקר בתחום הדעת הנחקר.

יש לציין שבחלק הממצאים בדו"ח המחקר או במאמר אין צורך לדווח על ניתוחי מובהקות סטטיסטיים של ההבדלים בין קבוצות הניסוי בשלב ההתחלתי שלו (Moher et al., 2010). מטרת מבחני המובהקות היא לבחון אם ההבדלים בין קבוצות הניסוי אינם תוצר של אקראיות. מכיוון שהחוקר מחלק את הקבוצות באופן אקראי, הוא יודע כבר בתחילת הניסוי שאם יהיו הבדלים בין הקבוצות, הרי הם תוצר של

אקראיות, ולכן מבחני המובהקות לגילוי הבדלים בשלב ההתחלתי של הניסוי הם חסרי משמעות ואינם סבירים (Altman, 1985; Vickers, 2009).

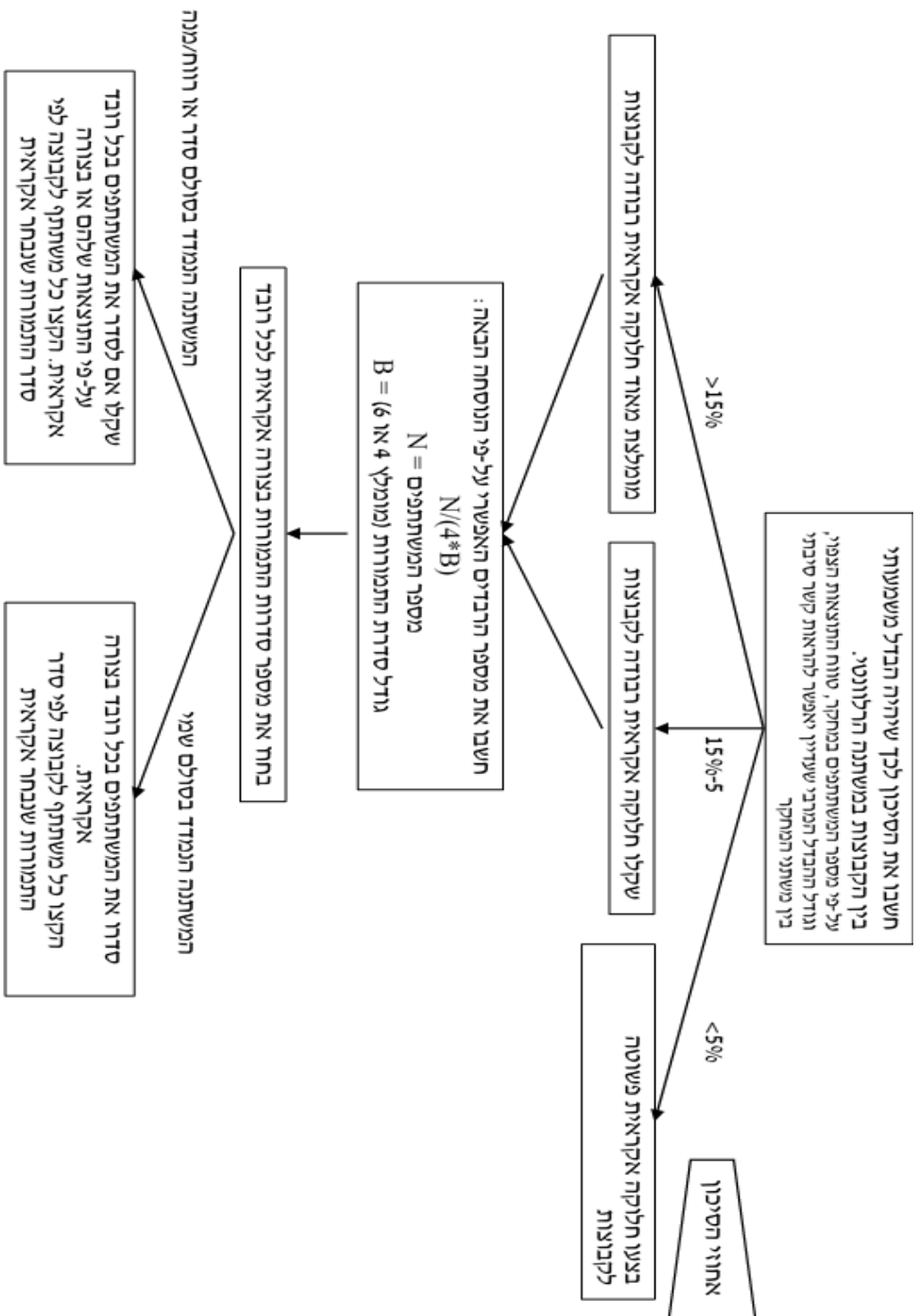
המלצות לחוקרים וסיכום

מומלץ להגדיר את שיטת החלוקה האקראית לקבוצות כבר בעת תכנון הניסוי, לאחר שהחוקר איתר את המשתנים המתערבים הרלוונטיים. מומלץ לחוקר לאמץ כמה צעדים בתהליך הגדרת החלוקה האקראית. ביניהם: (א) חישוב הסיכון לכך שיהיה הבדל משמעותי בין קבוצות הניסוי במשתנה המתערב הרלוונטי; (ב) חישוב מספר הרבדים האפשרי; (ג) בחירה אקראית של מספר סדרות התמורות לכל רובד (ראו צעדים נדרשים בתהליך קבלת ההחלטות לגבי שיטת החלוקה האקראית לקבוצות הניסוי באיור 4).

אם, לדעת החוקר, ישנם יותר משני משתנים מתערבים פוטנציאליים שעשויים להשפיע על תוצאות הניסוי, החלוקה האקראית הרבודה אינה יעילה, ולכן על החוקר לבחור שיטה אחרת (למשל: Covariate adaptive randomization/minimization) (Lin, Zhu, & Su, 2015; Taves, 1974). במאמר זה התמקדנו בטיפולו של החוקר בחלוקה אקראית בניסוי, תוך הבאה בחשבון של עד שני משתנים מתערבים פוטנציאליים.

כדי להצביע על קשר סיבתי בין משתנים בניסוי יש להקפיד, בין השאר, על חלוקה אקראית לקבוצות. אלא שלמרות נוחות השימוש בה, החלוקה האקראית הפשוטה, מאבדת מעילותה ככול שהמדגמים הולכים וקטנים. במאמר זה בחנו שיטה חלופית לחלוקה אקראית לקבוצות – החלוקה האקראית הרבודה. אמנם שיטה זו אינה נטולת חסרונות, אך מומלץ להשתמש בה כדי לוודא שקבוצות הניסוי "מתחילות מאותה נקודה" במשתנה שעשוי להשפיע על תוצאות הניסוי.

השימוש בקוד R (ראו נספחים 1 ו-2) יכול לסייע לחוקרים לחשב את הסיכון בהגעה למצב שבו הקבוצות אינן שוות ביכולתן בתחילת הניסוי. חישוב זה עשוי לעזור בקבלת ההחלטה לגבי סוג החלוקה האקראית. לבסוף, ללא קשר לסוג החלוקה האקראית, על החוקר להגדיר, בשלב תכנון הניסוי, את המשתנים המתערבים האפשריים (covariates) ולהתייחס אליהם בשלב ניתוח הנתונים גם כאשר אלו אינם שונים בצורה מובהקת בין הקבוצות בשלב ההתחלתי של הניסוי.



איור 4. צעדים נדרשים בתהליך קבלת החלטות לגבי שיטת החלוקה האקראית לקבוצות הניסוי

רשימת המקורות

- בייט-מרום, ר., וספורטה, ק. (2005). שיטות מחקר במדעי החברה – עקרונות המחקר וסגנונותיו (הניסוי–יחידה 6) (מהדורה שנייה). רעננה: האוניברסיטה הפתוחה.
- Altman, D. G. (1985). Comparability of randomized groups. *The Statistician*, 34, 125-136.
- Amonette, W. E., Kirk, E. L., & Kraemer, W. J. (2016). *Evidence-based practice in exercise science*. Champaign, IL: Human Kinetics.
- de Boer, M. R., Waterlander, W. E., Kuijper, L. D., Steenhuis, I. H., & Twisk, J. W. (2015). Testing for baseline differences in randomized controlled trials: An unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12, 4.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Dallas, TX: Houghton Mifflin Company.
- Cook, T. D., & Campbell, D. T. (1979). The design and conduct of quasi-experiments and true-experiments in field settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Skokie, IL: Rand-McNally.
- Enas, G. G., Enas, N. H., Spradlin, C. T., Wilson, M. G., & Wiltse, C. G. (1990). Baseline comparability in clinical trials: Prevention of "Poststudy Anxiety". *Drug Information Journal*, 24, 541-548.
- Gratton, C., & Jones, I. (2010). *Research methods for sports studies*. London, UK: Routledge.
- Hoffman, J. R. (2006). *Norms for fitness, performance, and health*. Champaign, IL: Human Kinetics.

- Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, *52*, 19-26.
- Lin, Y., Zhu, M., & Su, Z. (2015). The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials*, *45*, 21-25.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P., . . . Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, *340*, c869.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.
- Taves, D. R. (1974). Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics*, *15*, 443-453.
- Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2015). *Research methods in physical activity* (7th ed.). Champaign, IL: Human Kinetics.
- Vickers, A. J. (2009). *What is a p-value anyway? 34 stories to help you actually understand statistics*. London, UK: Pearson.
- Warner, R. M. (2008). *Applied statistics - from bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.

נספח 1:

קוד R לחישוב הסיכוי לקבלת אי-שוויון בין שתי קבוצות במשתנה אחד
(הקוד המודגש הוא זה שהחוקר יכול לשנות בהתאם לצרכיו)

```
# checking the chances of differences in a measured variable at baseline
library(dplyr)
i=0
count=0
#repetitions =how many times to run the simulation
#expected_score_range = what is the scores you expect to see in the measured variable
#participants_per_group =how many participants per group
#percent_diff = percent of difference allowed
repetitions=1000
expected_score_range=10:80
participants_per_group = 15
percent_diff=.2
vec_x1=c()
vec_x2=c()
while (i<=repetitions) {
  x1=mean(sample(expected_score_range, participants_per_group, replace=T))
  x2=mean(sample(expected_score_range, participants_per_group, replace=T))
  percent_diff_allowed=percent_diff
  if (abs(x1-x2)>percent_diff_allowed*max(c(x1,x2))) {
    count=count+1
    vec_x1=append(vec_x1, x1)
    vec_x2=append(vec_x2, x2)
    cat("x1=", x1, " x2=", x2, "\n")
  }
  i=i+1
}
cat("number of pairs with difference above ", percent_diff_allowed*100, "%: ", count)
cat("percentage of unequal pairs: ", count/repetitions*100, "%")
data=data.frame(vec_x1,vec_x2)
data=data %>% mutate(difference=abs(vec_x1-vec_x2))
cat ("Range of differences between samples: ", range(data$difference))
```

נספח 2:

קוד R לחישוב התמורות לסדרות של ארבע תמורות, שש תמורות, או לבחירה אקראית של סדרות התמורות (הקוד המודגש הוא זה שיכול החוקר לשנות בהתאם לצרכיו. יש לבחור, לאחר מכן, את שורת הקוד המתאימה לחישוב בו החוקר מעוניין)

```
# Permuted block design - using the blockrand package
library(blockrand)
number_of_groups=2
participants_number=48
group_names=c("aerobic", "control")
#example for: 48 participants, 2 groups, blocksize of 4
blockrand(participants_number, number_of_groups, levels=group_names, block.sizes =
c(2,2))
#example for: 48 participants, 2 groups, blocksize of 6
blockrand(participants_number, number_of_groups, levels=group_names, block.sizes =
c(3,3))
#example for: 48 participants, 2 groups, random blocksize of 4 or 6
blockrand(participants_number, number_of_groups, levels=group_names, block.sizes =
c(2,3))
#example for: 48 participants, 2 groups, random blocksize of 2, 4 or 6
blockrand(participants_number, number_of_groups, levels=group_names, block.sizes =
c(3))
```