

## הפאנל האינטרנטי ההסתברותי היחיד במדינת ישראל: ייצוגיות, דגימה הסתברותית, ותהליך הגיוס המורחב לפאנל

סיגל אלון, ענת אורן, קארין בלנרו\*

תקציר. הערת מחקר זו מציגה את הפאנל האינטרנטי ההסתברותי היחיד במדינת ישראל, שפותח ונבנה על ידי מכון ב. י. ולוסיל כהן למחקרי דעת קהל באוניברסיטת תל אביב. מתוארים בה ייחודיותו של הפאנל האינטרנטי ההסתברותי ושימושו האפשריים בהשוואה לפאנלים האינטרנטיים הקיימים בישראל, וכן האופן שבו נבנה מדגם הפאנל. הגיוס לפאנל ושימורו הם הליך מורכב הכולל את המרכיבים הבאים: בניית מדגם ייעודי המבוסס על דגימה הסתברותית מתוך מסגרת דגימה – דגימה אקראית של פרטים ממאגר מרשם האוכלוסין; שילוב שיטות גיוס לפאנל (באמצעות הטלפון והדואר) להגברת הייצוגיות; ורענון מתמיד של הפאנל. הליך זה, שנעשה בשקיפות מלאה כנדרש, מקנה לפאנל יתרונות מתודולוגיים במחקר על פני כל הפאנלים האחרים הקיימים כיום בישראל, למשל יכולת לחשב את טעות הדגימה. זהו הפאנל היחיד בישראל העומד בהלימה לסטנדרטים המתודולוגיים הגבוהים ביותר, והוא יכול להחליף סקרי פנים אל פנים או סקרים טלפוניים. איכותו המתודולוגית גבוהה ביותר, ועלותו נמוכה בהשוואה לעלויות של שיטות האיסוף המסורתיות.

מילות מפתח: סקר אינטרנטי, פאנל אינטרנטי הסתברותי, דגימה הסתברותית, מדגם מייצג

### הקדמה

משבר הקורונה יצר צורך בחשיבה יצירתית על דרכים חדשות לעריכת מחקרים תוך הקפדה ושמירה על מצוינות מתודולוגית. במהלך השנים, אף שמדינות רבות בעולם עברו

\* פרופ' סיגל אלון, החוג לסוציולוגיה ואתרופולוגיה ומכון ב. י. ולוסיל כהן למחקרי דעת קהל, אוניברסיטת תל אביב

ד"ר ענת אורן, מכון ב. י. ולוסיל כהן למחקרי דעת קהל

גב' קארין בלנרו, מכון ב. י. ולוסיל כהן למחקרי דעת קהל

תודה לפרופ' דוד שטיינברג מהחוג לסטטיסטיקה באוניברסיטת תל אביב על הייעוץ בבניית המדגם לפאנל האינטרנטי, ותודה לד"ר אירית אדלר, מזל דודו הלוי ולחברות ולחברי הוועדה המדעית של מכון ב. י. ולוסיל כהן למחקרי דעת קהל על הערות והארות לגבי תהליך הבנייה של הפאנל ולגבי הבדיקות המתודולוגיות.

לאיסוף נתונים אינטרנטי, בישראל לא היה אפשר לאסוף נתונים במדגם הסתברותי מייצג באמצעות שאלונים למילוי עצמי באינטרנט כפי שאפשר היה לעשות באיסוף נתונים טלפוני או פנים אל פנים. לכן איסוף נתונים אינטרנטי ברמה מתודולוגית גבוהה נעשה רק כאשר היו מסגרות דגימה סגורות, למשל עובדים בארגון, מאושפזים, בוגרי אוניברסיטה, מקבלי מענקים וכדומה. משבר הקורונה חידד את הבעיה, שכן במהלכו לא הייתה אפשרות לערוך מחקרים פנים אל פנים. בחודש מאי 2020, מכון ב. י. ולוסיל כהן למחקרי דעת קהל<sup>1</sup> הרים את הכפפה והחליט להשקיע (ועדיין משקיע גם כיום) משאבים כספיים אדירים בהיקף מאות אלפי שקלים בבניית פאנל אינטרנטי מייצג, היחיד במדינת ישראל, שמבוסס על מדגם פרטים רנדומלי וייעודי ממרשם האוכלוסין (מאגר משרד הפנים). בסקירה זו נעמוד על הבעייתיות בפאנלים האינטרנטיים הקיימים בישראל ונתאר את ייחודיותו של הפאנל האינטרנטי ההסתברותי ואת שימושי האפשריים. בחלקה השני של הסקירה נדגים את אופן בנייתו של מדגם הפאנל באמצעות שני בסיסי נתונים. הפאנל האינטרנטי ההסתברותי של מכון כהן נבדל מפאנלים שקיימים כיום בישראל בשני היבטים מתודולוגיים חשובים. ראשית, הוא נבנה מתוך מדגם ייעודי, המשמש לגיוס המשתתפים בפאנל ומבוסס על דגימה רנדומלית מתוך מסגרת דגימה. שנית, הגיוס לפאנל נעשה באמצעות שילוב של שיטות גיוס: סקר גיוס טלפוני ייעודי וגיוס באמצעות דואר. הליך גיוס זה – דגימה מתוך מסגרת דגימה ושילוב שיטות גיוס – עומד בהלימה לסטנדרטים הסתברותיים בינלאומיים באיכות המתודולוגית הגבוהה ביותר במחקרי אינטרנט, כמו ה־NORC (National Opinion Research Center) באוניברסיטת שיקגו ומחקרי מכון פו האמריקאי (Pew Research Center, Kaczmarek et al., 2019; Keeter, 2019; Keeter & McGeeney, 2015)<sup>2</sup>.

מבחינה מתודולוגית, הפאנל האינטרנטי ההסתברותי הוא תחליף ראוי לסקרים פנים אל פנים ולסקרים טלפוניים מכיוון שהוא מבוסס על אותם עקרונות דגימה. יתרונו בכך שהוא מאפשר לקבל נתונים באיכות גבוהה, במהירות, במצבים רבים (למשל בעת סגר) ובעלות נמוכה מזו של שיטות איסוף מסורתיות (Cernat and Revilla, 2020). נוסף על כך, הפאנל מאפשר לערוך ניסויים מורכבים על מדגם מייצג של האוכלוסייה שנבנה מתוך מסגרת דגימה, וכך להסיק מסקנות מתוצאות הניסוי לגבי האוכלוסייה הכללית. כדי שהפאנל לא יכלול "משיבים מקצועיים" מקפיד מכון כהן על רענון המשתתפים בפאנל. משתתפים מגויסים לפאנל האינטרנטי באופן שוטף מאז מאי 2020 ועד היום, כך שהפאנל גדל כל הזמן ואף מתרענן. משיבים המבקשים לצאת מהפאנל מוסרים ממנו. כיום

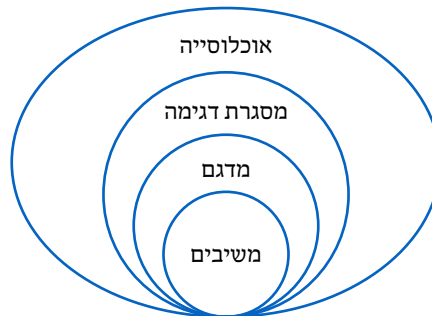
1 מכון ב. י. ולוסיל כהן למחקרי דעת קהל הוא מכון אוניברסיטאי המשלב בין פיתוח אקדמי של תאוריה ומתודולוגיה של סקרים ובין עריכת מחקרי דעת קהל בלתי תלויים ברמה מדעית גבוהה ביותר. המכון פועל במסגרת הפקולטה למדעי החברה, בשיתוף החוג לסטטיסטיקה וחקר ביצועים, בכל הקשור לבניית מדגמי אוכלוסייה לסקרים שונים על פי הדרישות הייחודיות של כל סקר. מטרתו של המכון הן לערוך סקרים עבור חוקרים מהאקדמיה ועבור מוסדות ציבור; לפתח תוכנית של סקרי עמדות תקופתיים בנושאים בעלי עניין ציבורי בחברה הישראלית; לבנות מאגרי נתונים שיאפשרו לעקוב לאורך זמן אחר מגמות בדעת הקהל בישראל; לערוך ניסויים במתודולוגיות של סקרים כדי לתרום לאיכותם של מחקרי דעת קהל בישראל ולפתחם; ולהעביר סדנאות מתודולוגיות בנושאים הקשורים לאיסוף ולניתוח נתוני סקרים.

2 לתיאור פאנלים נוספים מבוססי דגימה הסתברותית ראו Callegaro et al., 2014.

מונה הפאנל יותר מ-10,000 משתתפים יהודים וכ-1,000 משתתפים ערבים. המטרה היא להגיע למספר קבוע של 12,000 משיבים שיש בהם תחלופה. מאז בנייתו של הפאנל נערכו בעזרתו סקרים רבים ומגוונים. כך למשל, לראשונה מאז הצטרפה ישראל לפרויקט הבינלאומי למחקרים חברתיים (ISSP),<sup>3</sup> איסוף הנתונים של מודולת 2019 בנושא אי-שוויון חברתי נעשה בישראל באמצעות הפאנל האינטרנטי ההסתברותי של מכון כהן (עד כה נעשה איסוף הנתונים בישראל באמצעות ראיונות פנים אל פנים). כמו כן, במסגרת מחקר הבחירות הישראלי (INES, Israel National Elections Studies) נערכו סקרי פאנל בשתי נקודות זמן (לפני ואחרי הבחירות). עד כה איסוף הנתונים בפרויקט נעשה באמצעות ראיונות טלפוניים. סקרי פאנל אחרים נערכו בהפרש של שנה בין שתי נקודות הזמן.

### מהי דגימה הסתברותית?

דגימה הסתברותית (probabilistic sampling) היא שם כולל לפרוצדורות לבחירת מדגם, שבהן היחידות שבמדגם נבחרות רנדומלית. לדגימה הסתברותית נדרשת מסגרת דגימה, שהיא רשימת כל יחידות האוכלוסייה. יחידות האוכלוסייה שבמדגם נבחרות רנדומלית מתוך מסגרת הדגימה, ולכל יחידה יש סיכוי חיובי ויודע להיכלל במדגם. לעומת זאת, כאשר בחירת המדגם אינה נעשית באופן רנדומלי או שהסיכוי לבחירה אינו ידוע, הדגימה היא לא הסתברותית. היתרון העיקרי של דגימה הסתברותית הוא מידה גבוהה של ייצוגיות המדגם (representation) את האוכלוסייה. זהו התוקף החיצוני (external validity) של הסקר, כלומר היכולת להכליל מהמדגם אל כלל האוכלוסייה אגב אומדן הטעות הסטטיסטית (Cornesse et al., 2020).



איור 1. דגימה הסתברותית

3 פרויקט ISSP (International Social Survey Program) הוא תוכנית בינלאומית למחקרים חברתיים. זוהי קבוצת עבודה המורכבת מנציגים של מוסדות מחקר ממדינות שונות המשתפות פעולה בעריכת סקרי דעת קהל שנתיים, במטרה לשפר את המחקר ההשוואתי בין מדינות. תחילתה של התוכנית בשנת 1985 בשיתוף פעולה מחקרי בין גרמניה, בריטניה, ארצות הברית ואוסטרליה. כיום משתתפות בפרויקט 45 מדינות. ישראל הצטרפה לתוכנית בשנת 1989 ושימשה מזכירת הפרויקט בשנים 2008-2014.

## דגימה לא הסתברותית והפאנלים הקיימים כיום בישראל

דגימה לא הסתברותית (non-probabilistic sampling) היא כזאת שבה המדגם אינו נבחר רנדומלית מתוך האוכלוסייה, או שהסיכוי לבחירה של יחידה במדגם אינו ידוע (Elliott & Valliant, 2017). בשנים האחרונות, בהיעדר פאנל אינטרנטי המבוסס על דגימה הסתברותית, גדל מאוד השימוש של מכוני מחקר שוק בסקרים לא הסתברותיים למטרות איסוף נתונים. במקרים רבים אלה הם סקרים אינטרנטיים המבוססים על דגימה מתוך פאנל קבוע של אנשים, שפנו ביוזמתם ונרשמו לפאנל, ובמשך תקופה מסוימת הם ממלאים שאלונים לסקרים שונים עבור תשלום או עבור הטבות עם שווי כספי. כלומר, המשיבים בחרו את עצמם לפאנל ולא נדגמו באופן רנדומלי מתוך מסגרת דגימה הכוללת את כל הפרטים באוכלוסייה. בפאנלים אלו נוטים להשתתף מובטלים רבים יותר ומשתמשי אינטרנט רבים יותר בהשוואה לחלקם היחסי באוכלוסייה, וכ-30% מהאנשים שייכים ליותר מפאנל אחד (Pennay et al., 2018). פאנלים אלו מבטיחים מדגם מייצג, אך בפועל הם מספקים מדגם מכסות מתוך מאגר האנשים שנרשמו לפאנל<sup>4</sup>. השאלה היא, כמובן, את מי מייצג מדגם שכזה. כאשר המשיבים הם מי שפנו בעצמם לחברות הפאנלים, הדגימה היא לא הסתברותית (דגימת נוחות), שכן בחירת המדגם אינה נעשית באופן רנדומלי והסיכוי לבחירה אינו ידוע. המדגם נבנה על פי מכסות (תאים) חברתיות כלכליות, למשל מכסות של אזור גאוגרפי, מגדר וגיל, אשר משקפות (ולא מייצגות) את ההתפלגויות באוכלוסייה. כלומר, אף שבפאנלים הקיימים כיום בישראל מבטיחים מדגם מייצג, בפועל מדובר בקבוצה של אנשים שלא ידוע כיצד נבחרו ומהם באמת מאפייניהם הדמוגרפיים. משיב שציין שגילו 65 – מאפיין שמגדיל את הסיכוי להיכלל בפאנל, בשל מיעוט במשתתפים מבוגרים – יכול בעצם להיות בן 28. מפעילי הפאנל דואגים באופן מלאכותי שהמאפיינים העיקריים של הנדגמים (לרוב אלה הם מאפיינים נגישים, כגון גיל ומגדר) יתפלגו באופן דומה להתפלגותם באוכלוסייה. כאשר מעריכים תוצאות סקרים מפאנלים לא הסתברותיים, חשוב לדעת שאי-אפשר להסיק ממדגם כזה מסקנות לגבי כלל האוכלוסייה ואין בסיס מדעי לחישוב מדדים סטטיסטיים בסיסיים כגון רווח בר סמך או הטעות הסטטיסטית הנובעת מכך שהאומדנים חושבו על מדגם ולא על האוכלוסייה. כל החסרונות הללו פוגעים במהימנות המחקר, ומתוך כך ביכולת הפרסם בכתבי עת מקצועיים (Callegaro et al., 2014).

## בניית הפאנל האינטרנטי ההסתברותי של מכון כהן

במשך השנים לא היה אפשר לערוך בישראל איסוף נתונים אינטרנטי במדגם הסתברותי מייצג. משבר הקורונה הקשה עוד יותר על איסוף הנתונים, מכיוון שלא הייתה אפשרות לערוך סקרים פנים אל פנים. לפיכך החל אז מכון כהן בבנייתו של פאנל אינטרנטי הסתברותי מייצג, היחיד מסוגו בישראל.

4 מדגם מכסות (quota sampling), להבדיל ממדגם שכבות, הוא מדגם נוחות (הנחשב גם כסוג של דגימה מתואמת) הכולל מכסות חברתיות כלכליות שמשקפות את התפלגויותיהן באוכלוסייה.

הפאנל האינטרנטי מבוסס על דגימה הסתברותית: בקרב האוכלוסייה היהודית מתבצעת דגימה רנדומלית של פרטים מתוך מאגר מרשם האוכלוסין (שאינו כולל מספרי טלפון) ואיתור מספרי טלפון לפרטים הנדגמים. בקרב האוכלוסייה הערבית לא ניתן לאתר מספרי טלפון של פרטים,<sup>5</sup> ולכן מתבצעת דגימה רנדומלית של משקי בית מתוך מאגר בתי האב בישראל (הכולל מספרי טלפון, ומתוכם נדגם רנדומלית מספר טלפון אחד בכל משק בית). הפאנל האינטרנטי ההסתברותי של מכוני כהן נבדל מפאנלים שקיימים כיום בישראל בשני היבטים מתודולוגיים חשובים. ראשית, המדגם המשמש לגיוס המשתתפים בפאנל הוא מדגם ייעודי המבוסס על דגימה רנדומלית מתוך מסגרת דגימה, באוכלוסייה היהודית ממאגר מרשם האוכלוסין ובאוכלוסייה הערבית ממאגר בתי האב בישראל. כך מתקיים תנאי מכריע בדגימה הסתברותית: דגימה אקראית מתוך מסגרת דגימה. שנית, הגיוס לפאנל נעשה באמצעות שילוב בין סקר טלפוני ייעודי ובין גיוס באמצעות דואר.<sup>6</sup> באמצעות הליך גיוס זה – דגימה רנדומלית מתוך מסגרת דגימה ושילוב של שיטות גיוס – נבנה פאנל אינטרנטי הסתברותי המייצג את האוכלוסייה בישראל.

יתרון מרכזי בדגימה מתוך מאגר מרשם האוכלוסין הוא הנתונים הדמוגרפיים הקיימים עבור כלל הנדגמים ("מאותרים" ו"לא מאותרים"); "מסכימים" להשתתף בפאנל ו"לא מסכימים". נתונים אלו מאפשרים לבצע בדיקות מתודולוגיות ומעקב אחר ייצוגיות התפלגות המאפיינים הדמוגרפיים בעת בניית הפאנל. אי לכך בדיקות מתודולוגיות אלו מתאפשרות רק בקרב האוכלוסייה היהודית. מאמר זה מתאר את תהליך הבדיקות המתודולוגיות והממצאים.

כדי לבצע בדיקות אלו בעת בניית הפאנל השתמש מכוני כהן בנתונים משני מדגמים של האוכלוסייה היהודית. מדגם I נבנה במאי 2020 וכלל 16,000 נדגמים. העבודה עליו נעשתה בחודשים מאי-ספטמבר 2020 והוא שימש לבדיקת התפלגויות המאפיינים הדמוגרפיים כאשר הגיוס נעשה רק באמצעות סקר טלפוני. מדגם II נבנה בספטמבר 2020 וכלל אף הוא 16,000 נדגמים. העבודה עליו נעשתה בין ספטמבר 2020 לאפריל 2021 והוא שימש לבדיקת התפלגויות המאפיינים הדמוגרפיים כאשר הגיוס נעשה הן באמצעות סקר גיוס טלפוני והן באמצעות גיוס דואר.

אפשר לומר שתקן הזהב (gold standard) של פאנלים הסתברותיים הוא פאנל AmeriSpeak של NORC (Greaves, 2017). פאנל זה נבנה כבר לפני כעשור מתוך מסגרת דגימה ובשילוב שיטות גיוס. בשלב הראשון נשלחת לנדגמים ערכת גיוס בדואר. בשלב השני, מי שלא ענה זוכה לשתי תזכורות נוספות באמצעות גלויה, וכן לתזכורת שלישית טלפונית אם אותר מספר הטלפון שלו. לבסוף, בשלב השלישי, נערך ביקור פנים אל פנים במשקי בית של נדגמים שלא נוצר איתם קשר.

5 כאשר דוגמים מתוך מאגר מרשם הפנים, עבור מרבית האוכלוסייה הערבית אין כתובת מדויקת ומצוין רק שם היישוב. נוסף על כך, באותו יישוב אפשר למצוא פרטים רבים בעלי אותו שם פרטי ושם משפחה. שילוב של שני גורמים אלו מביא לשיעור נמוך מאוד של איתור טלפוני באוכלוסייה הערבית, עד כדי אחוזים בודדים בלבד, ולכן דגימה ממרשם האוכלוסין תוביל להטיה בייצוגיות המדגם.

6 באוכלוסייה הערבית הגיוס לפאנל התבצע באמצעות סקר גיוס טלפוני, ללא שלב הגיוס באמצעות הדואר, מאותן סיבות שצוינו לעיל.

במכון כהן היה ברור כי השימוש בדואר ישראל בעייתי באופן כללי, ובפרט בפנייה ראשונית לנדגמים, ולכן הוחלט כי הפנייה הראשונה תיעשה באמצעות הטלפון וכי היקף ההטיה של שיטה זו ייבדק באמצעות בדיקות מתודולוגיות לגבי ייצוגיות מאפייני המשתתפים. לפי הצורך תיעשה בשלב מאוחר יותר פנייה נוספת באמצעות הדואר, במטרה למצות את שיעור ההשבה ועל מנת לשפר את ייצוגיות הפאנל, תוך התבססות על מסגרת הדגימה ושמידה על עקרון הדגימה הרנדומלית. בסך הכול נבנה הפאנל בארבעה שלבים, המפורטים להלן.

**שלב ראשון: בניית מדגם פרטים רנדומלי ממרשם האוכלוסין ואיתור מספרי טלפון של הנדגמים**

תחילה, כאמור, נדגמו פרטים רנדומלית ממרשם האוכלוסין, ומספרי הטלפון שלהם אותרו בעזרת חברה לאיתור מספרי טלפון. השאלה המתבקשת היא כמובן מיהם הנדגמים שאותר להם לפחות מספר טלפון אחד (להלן "המאותרים") ומי הם הנדגמים שלא אותר להם מספר טלפון ("הלא מאותרים"). היתרון שבדגימה מתוך מרשם האוכלוסין הוא הנתונים הזמינים לכלל הפרטים במדגם לגבי מגדר, גיל ואשכול חברתי-כלכלי – הן למאותרים, הן ללא מאותרים. לכ-80% מהנדגמים אותרו מספרי טלפון אישיים או מספרי טלפון של קרובי משפחה, ולכאורה היה אפשר לגייס אותם באמצעות הטלפון. בפועל, בעת ההתקשרות למאותרים נמצא כי שיעור האיתור היה נמוך יותר, בשל שיעור גדול של מספרי טלפון לא תקינים ובשל טעויות בזיהוי הנדגם. לפיכך הוחלט להרחיב את תהליך הגיוס ולהוסיף גיוס באמצעות דואר.

טבלה 1 מציגה את התפלגות המגדר, הגיל והאשכול החברתי-כלכלי בקרב כלל הפרטים במדגם, בקרב המאותרים ובקרב הלא מאותרים, בהשוואה למאפייני האוכלוסייה היהודית הבוגרת בישראל (הלמ"ס, 2019, לוח 2.3). אפשר לראות כי הן התפלגות המגדר הן התפלגות הגיל בקרב כלל הנדגמים לסקר הגיוס דומות מאוד להתפלגויות בכלל האוכלוסייה. כמו כן, התפלגות המגדר והגיל בקרב המאותרים והלא מאותרים גם יחד דומה מאוד להתפלגויות בכלל האוכלוסייה. בקרב הלא מאותרים קיים פער קטן בקבוצת הגיל המבוגרת ביותר: שיעור בני 70 ומעלה בקרב הלא מאותרים הוא 17.3% לעומת 13.8% בקרב המאותרים, ובכלל האוכלוסייה הוא עומד על 13.1%. התפלגות האשכול החברתי-כלכלי בקרב המאותרים דומה לזו של כלל הנדגמים במדגם ההסתברותי, ואילו בקרב הלא מאותרים נמצא שיעור גבוה יותר של האשכולות הנמוכים ביותר, על חשבון האשכולות הגבוהים.

טבלה 1. התפלגות מגדר, גיל ואשכול חברתי-כלכלי בקרב כלל הנדגמים, המאותרים והלא מאותרים, בהשוואה למאפייני האוכלוסייה היהודית בישראל

| מגדר              | גברים | כלל האוכלוסייה היהודית הבישראל | כלל מדגם משרד הפנים לסקר גיוס | טיוב (איתור מספרי טלפון) |            |
|-------------------|-------|--------------------------------|-------------------------------|--------------------------|------------|
|                   |       |                                |                               | מאותרים                  | לא מאותרים |
| מגדר              | גברים | 48%                            | 49.5%                         | 49.7%                    | 48.9%      |
|                   | נשים  | 52%                            | 50.5%                         | 50.3%                    | 51.1%      |
| גיל               | 29-18 | 23.1%                          | 21.9%                         | 21.7%                    | 22.5%      |
|                   | 39-30 | 19.5%                          | 19.0%                         | 19.4%                    | 17.9%      |
|                   | 49-40 | 17.4%                          | 17.3%                         | 17.7%                    | 16.1%      |
|                   | 59-50 | 13.7%                          | 13.9%                         | 14.1%                    | 13.2%      |
|                   | 69-60 | 13.2%                          | 13.1%                         | 13.2%                    | 13.0%      |
|                   | +70   | 13.1%                          | 14.8%                         | 13.8%                    | 17.3%      |
| אשכול חברתי-כלכלי | 3-1   | -                              | 21.9%                         | 19.3%                    | 29.4%      |
|                   | 6-4   | -                              | 32.0%                         | 31.9%                    | 32.3%      |
|                   | 8-7   | -                              | 36.8%                         | 38.4%                    | 32.3%      |
|                   | 10-9  | -                              | 9.3%                          | 10.4%                    | 6%         |

\* נתונים מתוך מדגם I

### שלב שני: עריכת סקר גיוס טלפוני בקרב המאותרים

סקר גיוס טלפוני נערך בקרב כלל המאותרים במטרה לקבל את הסכמתם להשתתף בפאנל. ראיונות הגיוס נערכו רק עם האדם שנדגם, ולא עם איש במקומו. סוקרי המכון וידאו שהם מראיינים את האדם שנדגם ממרשם האוכלוסין וזיהו אותו על פי שם פרטי ושם משפחה. אם נמצא שמספר הטלפון אינו של האדם שנדגם אלא של קרוב משפחה, ביקש המראיין מקרוב המשפחה אמצעי התקשורת עם הנדגם. אם קרוב משפחה, או כל אדם אחר, ביקש להשתתף אף הוא בפאנל האינטרנטי ההסתברותי, המראיין התנצל והסביר כי רק האדם שנדגם יכול להשתתף בפאנל. מראייני המכון עמלו לאתר את הנדגם בהתקשרויות חוזרות, בימים שונים ובשעות שונות, במטרה למצות את המדגם. שיעור ההיענות בסקר הטלפוני היה גבוה ועמד על 45% (מתוך המאותרים שעיימם נוצר קשר אישי, או עם אדם מטעמם, או שלא הייתה תשובה).<sup>7</sup>

7 זהו שיעור היענות גבוה יחסית למקובל בסקרים טלפוניים. מחקרים אמפיריים מלמדים כי איסוף טלפוני סטנדרטי בקרב כלל הציבור הישראלי - התקשרות במשך חמישה ימים, בשעות ובימים שונים, עד ארבע חזרות לכל משק בית, ופניית שכנוע אחת למסרבים - מביא לשיעורי היענות של 25%-20% (Miller et al., 2000).

לצד שיעור ההיענות הגבוה בסקר הגיוס נשאלת השאלה מיהם הנדגמים שענו לסקר הגיוס (להלן "המרוויינים"), בהשוואה לנדגמים שלא רואיינו ("הלא מרוויינים") ובהשוואה לכלל האוכלוסייה. טבלה 2 מציגה את התפלגות המגדר, הגיל והאשכול חברתי-כלכלי בקרב המרוויינים ובקרב הלא מרוויינים, בהשוואה למאפייני האוכלוסייה (הלמ"ס, 2019, לוח 2.3).

טבלה 2. התפלגות מגדר, גיל ואשכול חברתי-כלכלי בקרב מרווייני סקר הגיוס ובקרב הלא מרוויינים, בהשוואה למאפייני האוכלוסייה היהודית בישראל

| סקר גיוס | כלל האוכלוסייה היהודית הבוגרת בישראל |         | גברים | נשים  | גיל               |
|----------|--------------------------------------|---------|-------|-------|-------------------|
|          | לא רואיינו                           | רואיינו |       |       |                   |
|          | 50.2%                                | 49.4%   | 48%   |       | מגדר              |
|          | 49.8%                                | 50.6%   | 52%   |       |                   |
|          | 22.2%                                | 20.1%   | 23.1% | 29-18 |                   |
|          | 20.1%                                | 17.8%   | 19.5% | 39-30 |                   |
|          | 17.9%                                | 17.0%   | 17.4% | 49-40 |                   |
|          | 14.1%                                | 14.5%   | 13.7% | 59-50 |                   |
|          | 12.8%                                | 14.8%   | 13.2% | 69-60 |                   |
|          | 12.9%                                | 15.8%   | 13.1% | +70   |                   |
|          | 21.2%                                | 14.5%   | -     | 3-1   | אשכול חברתי-כלכלי |
|          | 33.5%                                | 27.8%   | -     | 6-4   |                   |
|          | 35.8%                                | 44.8%   | -     | 8-7   |                   |
|          | 9.5%                                 | 12.9%   | -     | 10-9  |                   |

\* נתונים מתוך מדגם I

בטבלה 2 אפשר לראות כי התפלגות המגדר והתפלגות הגיל, בקרב המרוויינים ובקרב הלא מרוויינים, דומות מאוד להתפלגויות בכלל האוכלוסייה. לגבי התפלגות האשכול חברתי-כלכלי, אפשר לראות כי המרוויינים נוטים להשתייך לאשכולות גבוהים יותר בהשוואה ללא מרוויינים.

### שלב שלישי: הזמנתם של מרווייני סקר הגיוס הטלפוני להשתתף בפאנל האינטרנטי

בסקר הגיוס נשאלים המרוויינים תחילה לגבי מאפיינים כמו גיל, מין, השכלה, דת ומידת דתיות. הם נשאלים גם לגבי שימוש במחשב ובאינטרנט, עבודה ותעסוקה, ותחושות ומצב רוח בעקבות משבר הקורונה. איסוף המידע הדמוגרפי חשוב, ומטרתו לוודא את זהותם של הנדגמים ואת מאפייניהם הדמוגרפיים על מנת לאפשר את בדיקת ייצוגיותו של הפאנל ואת הצעדים הנדרשים למיצוי מסגרת הדגימה. בתום הריאיון הקצר, המרוויינים מוזמנים להיכלל בפאנל האינטרנטי ולהשתתף בסקרים פעמים אחדות בשנה תמורת תגמול כספי עבור כל השתתפות בסקר. כ-65% מהמרוויינים, שהם 22% מכלל הנדגמים, מסכימים להשתתף בפאנל האינטרנטי.



הערה על מחקר

טבלה 3 מציגה התפלגות של מאפיינים דמוגרפיים (מגדר, גיל, השכלה ורמת דתיות), בקרב משתתפי סקר הגיוס המסכימים להיכלל בפאנל ובקרב הלא מסכימים, בהשוואה למאפייני האוכלוסייה. מהטבלה אפשר לראות כי בקרב המסכימים להיכלל בפאנל, התפלגות המגדר והתפלגות הגיל דומות מאוד להתפלגויות באוכלוסייה, למעט ייצוג חסר לבני 70 ומעלה.

טבלה 3. התפלגות מאפיינים דמוגרפיים בקרב כלל משתתפי סקר הגיוס ובקרב מסכימים ולא מסכימים להיכלל בפאנל, בהשוואה למאפייני האוכלוסייה היהודית הבוגרת בישראל

| סקר גיוס                |                      | כלל המשתתפים בסקר | כלל האוכלוסייה היהודית הבוגרת בישראל |                       |             |
|-------------------------|----------------------|-------------------|--------------------------------------|-----------------------|-------------|
| לא מסכימים להיכלל בפאנל | מסכימים להיכלל בפאנל |                   |                                      |                       |             |
| 49.7%                   | 49.2%                | 49.4%             | 48%                                  | גברים                 | מגדר        |
| 50.3%                   | 50.8%                | 50.6%             | 52%                                  | נשים                  |             |
| 13.0%                   | 24.8%                | 20.1%             | 23.1%                                | 29-18                 | גיל         |
| 14.1%                   | 20.2%                | 17.8%             | 19.5%                                | 39-30                 |             |
| 15.3%                   | 18.0%                | 17.0%             | 17.4%                                | 49-40                 |             |
| 13.4%                   | 15.3%                | 14.5%             | 13.7%                                | 59-50                 |             |
| 17.9%                   | 12.8%                | 14.8%             | 13.2%                                | 69-60                 |             |
| 26.3%                   | 9.0%                 | 15.8%             | 13.1%                                | +70                   |             |
| 21.6%                   | 11.2%                | 15.3%             | 28.8%                                | עד תיכונית, ללא בגרות | השכלה*      |
| 19.3%                   | 22.8%                | 21.4%             | 20.2%                                | תיכונית עם בגרות      |             |
| 22.0%                   | 20.8%                | 21.3%             | 15.5%                                | על-תיכונית            |             |
| 37.0%                   | 45.2%                | 42.0%             | 35.5%                                | אקדמית                |             |
| 10.9%                   | 6.5%                 | 8.2%              | 10.1%                                | דתי מאוד, חרדי        | רמת דתיות** |
| 10.2%                   | 12.2%                | 11.4%             | 11.3%                                | דתי                   |             |
| 12.5%                   | 10.7%                | 11.4%             | 12.8%                                | מסורתי דתי            |             |
| 15.4%                   | 15.1%                | 15.2%             | 22.1%                                | מסורתי לא כל כך דתי   |             |
| 51.0%                   | 55.5%                | 53.8%             | 43.2%                                | חילוני                |             |
| -                       | -                    | -                 | 0.4%                                 | לא ידוע               |             |

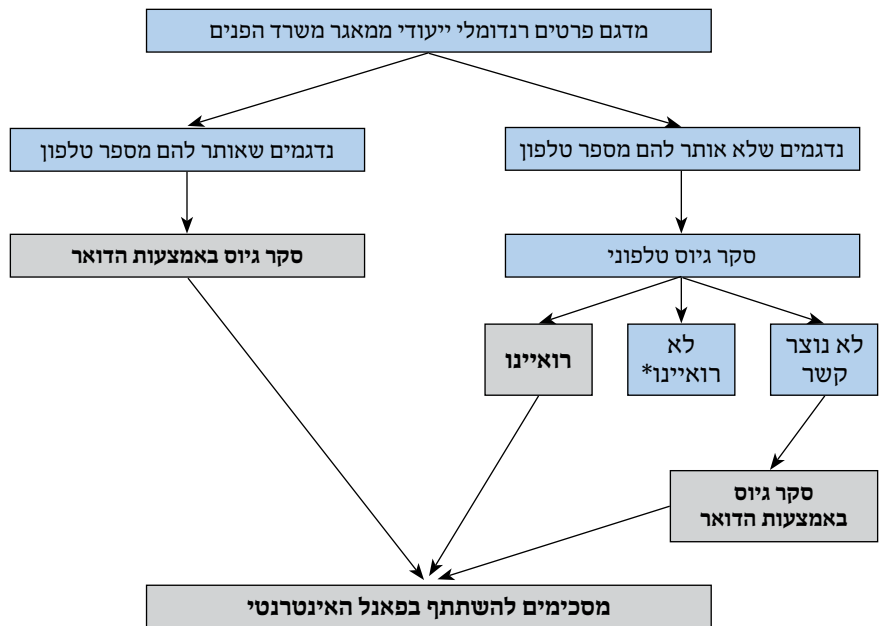
\* נתונים מתוך מדגם I

\*\* מתוך מחולל הלוחות של הסקר החברתי של הלמ"ס לשנת 2019

מהשוואה בין מאפייני ההשכלה בקרב כלל משתתפי סקר הגיוס ובין מאפייני ההשכלה בקרב כלל האוכלוסייה (טבלה 3) עולה כי שיעור בעלי ההשכלה האקדמית גבוה יותר בקרב משתתפי הסקר ואילו שיעור בעלי ההשכלה הנמוכה קטן יותר, האחד על חשבון השני. הבדלים אלו ניכרים אף יותר בקרב המסכימים להיכלל בפאנל. שיעור החרדים בקרב כלל משתתפי סקר הגיוס נמוך מעט מזה שבכלל האוכלוסייה היהודית הבוגרת, ושיעור נמוך יחסית של חרדים נמצא גם בקרב המסכימים להיכלל בפאנל האינטרנטי.

**שלב רביעי: הרחבת תהליך הגיוס - הוספת גיוס באמצעות דואר**

כצפוי, אומדן ההטיה בייצוגיות המגויסים לפאנל במדגם I חידד את הצורך בהרחבת תהליך הגיוס. בשל השיעור הלא מבוטל של נדגמים לא מאותרים - נדגמים שמספרי הטלפון שלהם לא אותרו - וכן בשל הפערים במאפייני משתתפי הפאנל, בעיקר בקבוצות הגיל המבוגרות ובקרב האשכולות החברתיים-כלכליים הנמוכים ביותר, הורחב תהליך הגיוס במדגם II וכלל גם גיוס באמצעות דואר בקרב שתי קבוצות: קבוצת הלא מאותרים, וקבוצת הנדגמים שמספרי הטלפון שלהם אותרו אך לא נוצר עימם קשר. איור 2 מציג את תהליך הגיוס המורחב כפי שנעשה במדגם II.



איור 2. תהליך הגיוס המורחב לפאנל האינטרנטי ההסתברותי: גיוס באמצעות סקר טלפוני ייעודי ובאמצעות הדואר

\* קטגוריית "לא רואיינו" כוללת נדגמים שלא רואיינו מסיבות שונות: סירוב להתראיין, בעיית שפה, אי-כשירות, נדגמים שנמצאו בחו"ל, נדגמים שנפטרו

הערה על מחקר

הגיוס באמצעות הדואר במדגם II הוסיף יותר מ-3% מסכימים להיכלל בפאנל. בסך הכול, בתהליך הגיוס המורחב רבע מתוך כלל הנדגמים המקוריים הסכימו להיכלל בפאנל. לכאורה הגידול בשיעור המסכימים קטן יחסית, אולם ההבדל הזה משמעותי מאוד בשל מאפייניהם הייחודיים, החשובים לייצוגיות המדגם. טבלה 4 מציגה את התפלגויות המגדר, הגיל, ההשכלה ורמת הדתיות בקרב כלל המסכימים ועל פי שיטת הגיוס, בהשוואה למאפייני כלל האוכלוסייה. מהטבלה אפשר לראות כי עיקר תרומתו של הגיוס באמצעות הדואר הוא בקרב חרדים, בקרב בעלי השכלה נמוכה ובקרב צעירים: 17.1% מקרב המסכימים להיכלל בפאנל שגויסו באמצעות הדואר הם חרדים, בהשוואה ל-6.2% שגויסו באמצעות הטלפון. התוספת הזאת מעלה את שיעור החרדים בכלל הפאנל ל-7.6%. כמו כן, 43.6% מקרב המסכימים להשתתף בפאנל שגויסו באמצעות הדואר הם צעירים בני 18-29, בהשוואה ל-21.1% מקרב המסכימים שגויסו באמצעות הטלפון. תוספת זו מעלה את שיעור הצעירים בכלל הפאנל ל-24%. בהמשך לכך, 20.1% מקרב המסכימים להשתתף בפאנל שגויסו באמצעות הדואר הם בעלי השכלה נמוכה, בהשוואה ל-12.4% מקרב המסכימים שגויסו באמצעות הטלפון. תוספת זו מעלה את שיעור בעלי השכלה הנמוכה בכלל הפאנל ל-13.4%, ומורידה את שיעור בעלי השכלה האקדמית ל-43.9%.

טבלה 4. התפלגויות מגדר, גיל, השכלה ורמת דתיות בקרב כלל המסכימים להיכלל בפאנל ועל פי שיטת הגיוס, בהשוואה למאפייני האוכלוסייה היהודית בישראל

| מסכימים בגיוס דואר | מסכימים בגיוס טלפוני | כלל המסכימים להיכלל בפאנל | כלל האוכלוסייה היהודית הבוגרת בישראל |                       |        |
|--------------------|----------------------|---------------------------|--------------------------------------|-----------------------|--------|
| 49.8%              | 47.4%                | 47.7%                     | 48%                                  | גברים                 | מגדר   |
| 50.2%              | 52.6%                | 52.3%                     | 52%                                  | נשים                  |        |
| 43.6%              | 21.1%                | 24.0%                     | 23.1%                                | 29-18                 | גיל    |
| 23.9%              | 17.8%                | 18.6%                     | 19.5%                                | 39-30                 |        |
| 13.1%              | 22.5%                | 21.3%                     | 17.4%                                | 49-40                 |        |
| 8.5%               | 15.1%                | 14.3%                     | 13.7%                                | 59-50                 |        |
| 6.9%               | 14.4%                | 13.5%                     | 13.2%                                | 69-60                 |        |
| 3.9%               | 8.9%                 | 8.3%                      | 13.1%                                | +70                   |        |
| 20.1%              | 12.4%                | 13.4%                     | 28.8%                                | עד תיכונית, ללא בגרות | השכלה* |
| 22.4%              | 20.6%                | 21.0%                     | 20.2%                                | תיכונית עם בגרות      |        |
| 22.2%              | 21.7%                | 21.8%                     | 15.5%                                | על-תיכונית            |        |
| 34.4%              | 45.2%                | 43.9%                     | 35.5%                                | אקדמית                |        |

|       |       |       |       |                     |         |
|-------|-------|-------|-------|---------------------|---------|
| 17.1% | 6.2%  | 7.6%  | 10.1% | דתי מאוד, חרדי      | רמת     |
| 13.6% | 12.6% | 12.7% | 11.3% | דתי                 | דתיות** |
| 9.9%  | 12.5% | 12.2% | 12.8% | מסורתי דתי          |         |
| 12.6% | 13.9% | 13.8% | 22.1% | מסורתי לא כל כך דתי |         |
| 46.7% | 54.8% | 53.7% | 43.2% | חילוני              |         |
| -     | -     | -     | 0.4%  | לא ידוע             |         |

\* נתונים מתוך מדגם II

\*\* מתוך מחולל הלוחות של הסקר החברתי של הלמ"ס לשנת 2019

## סיכום

בשני העשורים האחרונים נבנו בישראל פאנלים לא הסתברותיים רבים, ואלה הפכו לכלי נפוץ לאיסוף נתונים במחקרי שוק. פאנלים כאלה מבטיחים לספק מדגם מייצג בעלות נמוכה, אך בפועל הם מספקים מדגם מכסות מתוך פאנל קבוע של אנשים שפנו ביוזמתם ונרשמו לפאנל, ורבים מהם עונים על סקרים רבים ואף משתייכים ליותר מפאנל אחד. כלומר, כאשר המשיבים פונים בעצמם לחברות הפאנלים מתקבלת דגימה לא הסתברותית, שכן בחירת המדגם אינה נעשית באופן רנדומלי והסיכוי לבחירה אינו ידוע. מתוך הפאנל נבנה מדגם על פי מכסות חברתיות כלכליות (למשל לגבי אזור גאוגרפי, מגדר וגיל) שיוצרות באופן מלאכותי התפלגויות דומות לאלה שבאוכלוסייה, אך המדגם הזה אינו מייצג את האוכלוסייה. הפאנלים הקיימים בישראל מבטיחים מדגם מייצג, בשעה שבפועל המדגם עשוי מקבוצה של אנשים שלא ידוע כיצד נבחרו ומה באמת מאפייניהם הדמוגרפיים. העלות של איסוף הנתונים נמוכה יחסית, אך היא כרוכה במחיר כבד: תוצאות הסקרים מתבססות על מעין קופסה שחורה, אי־אפשר לדעת את מי מייצגים הממצאים, אי־אפשר לחשב את הטעות הסטטיסטית, אי־אפשר להסיק מסקנות מן המדגם ולהכליל מהן אל כלל האוכלוסייה, והיכולת לפרסם מחקרים כאלה בכתבי עת מקצועיים מוגבלת מאוד.

לעומת זאת, הליך הגיוס המורכב לפאנל האינטרנטי ההסתברותי של מכון כהן – בניית מדגם ייעודי מתוך מסגרת דגימה, שילוב שתי שיטות הגיוס, שקיפות מלאה ורענון משתפי הפאנל – יוצר יתרונות מתודולוגיים למחקרים אקדמיים, והוא הפאנל היחיד בישראל העומד בהלימה לסטנדרטים הסתברותיים באיכות המתודולוגית הגבוהה בעולם. לפיכך יכול הפאנל של מכון כהן להחליף סקרים פנים אל פנים או סקרים טלפוניים, ובה בעת לשמור על איכות מתודולוגית גבוהה ביותר ועל עלות נמוכה בהשוואה לעלות שיטות האיסוף המסורתיות.

## מקורות

- הלמ"ס. (2019). שנתון סטטיסטי לישראל 2019 (מס' 70).
- Callegaro, Mario, Reg Baker, Jelke Bethlehem, Anja S. Goritz, Jon A. Krosnick, & Paul J. Lavrakas. (2014). Online panel research: History, concepts, applications and a look at the future. In Mario Calegario, Reg Baker, Jelke Bethlehem, Anja S. Goritz, Jon A. Krosnick, & Paul J. Lavrakas (eds.), *Online Panel Research: A Data Quality Perspective* (pp. 1–22). Wiley.
- Cernat, Alexandru, and Melanie Revilla. (2020). Moving from face-to-face to a web panel: Impacts on measurement quality. *Journal of Survey Statistics and Methodology*, 9(4).
- Cornesse, Carina, Anneleis G. Blom, David Dutwin, Jon A. Krosnick, Edith D. De Leeuw, Stephane Legleye, Josh Pasek, Darren Pennay, Benjamin Phillips, Joseph W. Sakshaug, Bella Struminskaya, & Alexander Wenz. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology*, 8, 4–36.
- Elliott, Michael R., & Richard Valliant. (2017). Inference for nonprobability samples. *Statistical Science*, 32(2), 249–264.
- Greaves, Lara. (2017). *An investigation into the feasibility of an online national probability panel study in New Zealand*. The University of Oakland.
- Kaczmirek, Lars, Benjamin Phillips, Daren Pennay, Paul J. Lavrakas, & Dina Neiger. (2019). *Building a probability-based online panel: Life in Australia* (CSR Methods Series, 2/2019). Center for Social Research and Methods, Australian National University.
- Keeter, Scott. (2019, February 27). *Growing and improving Pew Research Center's American Trends Panel*. Pew Research Center.
- Keeter, Scott, & Kylie McGeeney. (2015, September 22). *Coverage error in internet surveys: Who web-only surveys miss and how that affects results*. Pew Research Center.
- Miller, Scott, Carolyn E. Miller, Andrew Kohut, & Robert Groves. (2000). Consequences of reducing non-response in a large national telephone survey. *Public Opinion Quarterly*, 64(2), 125–148.
- Pennay, Daren W., Dina Neiger, Paul J. Lavrakas, & Kim Borg. (2018). *The online panels benchmarking study: A Total Survey Error comparison of findings from probability-based surveys and nonprobability online panel surveys in Australia* (CSR and SRC Methods Paper, 2/2018). Center for Social Research and Methods, Australian National University.