

הערכת אפקטיביות

של תוכניות בשוק העבודה

אסנת ליפשיץ וטלי לרום*

נייר מדיניות לקראת הכנס השנתי של מכון אהרן - טיוטה להערות



* פרופ' אסנת ליפשיץ היא חברת סגל בבית ספר טיומקין לכלכלה באוניברסיטת רייכמן ויועצת למכון אהרן למדיניות כלכלית. ד"ר טלי לרום היא מנהלת מחקר בכירה במכון אהרן למדיניות כלכלית, אוניברסיטת רייכמן.

הערכת אפקטיביות של תוכניות בשוק העבודה

מדי שנה מושקעים בישראל מאות מיליוני שקלים בתוכניות תעסוקה, הכשרה מקצועית והשכלה עבור אוכלוסיות יעד שונות. אולם ללא הערכת אפקטיביות קפדנית, אין דרך לדעת אילו מהתוכניות אכן משיגות את יעדיהן - ואילו מהוות בזבוז משאבים, ובפרט משאבים ציבוריים. תוכנית שאינה אפקטיבית אינה רק הוצאה מיותרת מאחר ויש לה גם עלות אלטרנטיבית, והמשאבים שהושקעו בה היו יכולים להיות מופנים לתוכנית שתניב תשואה חיובית למשתתפים ולמשק כולו. נוסף על כך, ללא מדידה שיטתית נוצר מצב שבו החלטות מדיניות מתקבלות על בסיס אינטואיציה, לחצים פוליטיים או אינרציה - ולא על בסיס ראיות.

נייר זה מציג מסגרת סדורה להערכת אפקטיביות של תוכניות תעסוקה והכשרה בישראל, כחלק ממדיניות ציבורית מבוססת ראיות. האתגר המרכזי בהערכת אפקטיביות הוא לבודד את ההשפעה הסיבתית של התוכנית מגורמים אחרים, ולשם כך נדרשות שיטות אמידה מתקדמות. המסמך מציג את השיטות המובילות בתחום והמקובלות בעולם: ניסוי אקראי מבוקר (RCT), ניסוי טבעי, שיטות התאמה (Matching), (PSM), הפרש הפרשים (DiD), עיצוב אי-רציפות גרגרית (RDD), ומחקר אירוע (Event Study), תוך הבאת דוגמאות לשימוש בשיטות אלו בישראל ודיון ביתרונות ובחסרונות של כל אחת מהן. בנוסף, אנחנו מדגישים את החשיבות של בחינת ההשפעה הדינמית של תוכניות לאורך זמן ושל ניתוח עלות-תועלת אישי וחברתי כבסיס לתעדוף תוכניות והקצאת משאבים.

על מנת ליישם מדיניות מבוססת ראיות בישראל מוצע מודל לביצוע מחקרי אפקטיביות תוך שימוש בנתונים המנהליים של הלשכה המרכזית לסטטיסטיקה, עם או ללא מיזוג לנתוני המשתתפים בתוכנית. מודל זה מאפשר בניית קבוצת השוואה ומעקב אחר משתני התוצאה הרלבנטיים - השכלה, תעסוקה ושכר - גם כאשר ניסוי אקראי מבוקר אינו אפשרי לביצוע. הנייר מציג דוגמה ליישום מודל זה במחקר האפקטיביות לתוכנית "קימם" לשילוב צעירים מהחברה הערבית באקדמיה.

הערכה שיטתית תוך שימוש בשיטות אמידה מתקדמות היא תנאי הכרחי לכך שהשקעה בכלל, והשקעה ציבורית בפרט, בתוכניות תעסוקה והכשרה תניב את התשואה המרבית למשתתפים ולמשק. על מנת לבצע הערכת אפקטיביות שיטתית אנו ממליצים על יצירת תשתית מוסדית ומחקרית מתאימה, תוך דגש על איסוף והנגשה של נתונים איכותיים, מפורטים ועדכניים ושיתוף פעולה בין גופים ממשלתיים.

תוכן העניינים

1. רקע ומוטיבציה.....	4
2. האתגר בהערכת אפקטיביות.....	6
3. שיטות לאמידת אפקטיביות.....	8
3.1 ניסוי אקראי מבוקר (Randomized Controlled Trial - RCT).....	8
3.2 ניסוי טבעי (Natural Experiment).....	9
3.3 שיטות התאמה.....	10
3.4 הפרש ההפרשים (Difference-in-Differences - DiD).....	12
3.5 עיצוב אי-רציפות רגרסיבית (Regression Discontinuity Design - RDD).....	13
3.6 בחינת ההשפעה הדינמית של תוכניות התערבות.....	14
4. עלות-תועלת אישית וחברתית.....	18
5. מודל לביצוע מחקרי אפקטיביות בישראל ללא RCT תוך שימוש בנתונים מנהליים.....	19
6. דוגמה לביצוע מחקר אפקטיביות בישראל ללא RCT וללא מיזוג נתונים מנהליים : תוכנית "קימם" ..	21
7. המלצות מדיניות.....	24
מקורות.....	26
נספח א : הנתונים הדרושים.....	28
נספח ב : הגדרת המשתנים במחקר "קימם".....	30

1. רקע ומוטיבציה

"We can leverage good science to improve the effectiveness of policies that serve the poor worldwide...the accumulation of experimental results eventually paints a picture that helps make sense of the world, and guide policy. It is the accretion of results that makes sense and justifies the whole enterprise".

אסתר דופלו, זוכת פרס נובל בכלכלה על מחקרי אפקטיביות, מתוך נאום הזכייה בפרס.

בעשורים האחרונים הולכת ומתרחבת ההכרה בחשיבותם של מחקרי אפקטיביות כבסיס למדיניות ציבורית מבוססת-ראיות (Evidence-Based Policy) הערכת אפקטיביות מאפשרת לבחון האם תוכניות ותהליכי התערבות אכן משיגים את יעדיהם, באיזו מידה הם תורמים לרווחת המשתתפים ולצמיחה הכלכלית, ומהו יחס העלות-תועלת שלהם לחברה. בעולם הכלכלי והחברתי, מדובר באחת מגישות המחקר המרכזיות - תחום מוביל בספרות האמפירית המודרנית - אשר זיכה חוקרים בפרס נובל בכלכלה והוביל לשינוי משמעותי באופן שבו ממשלות וגופים בינלאומיים מתכננים ומעריכים תוכניות ציבוריות. בישראל, כמו במדינות אחרות, חיזוק היכולת למדוד ולהעריך אפקטיביות של תוכניות תעסוקה, הכשרה והשכלה הוא תנאי הכרחי לשיפור מתמיד של המדיניות, למיקוד משאבים בכלים יעילים ולשקיפות ציבורית גבוהה יותר.

שוק העבודה בישראל מתאפיין ברמת תעסוקה גבוהה ואבטלה נמוכה, ויחד עם זאת שני אתגרים משמעותיים ניצבים בפני המשק: (1) קבוצות שנמצאות עדיין ברמת תעסוקה נמוכה - גברים חרדים, גברים ערבים ונשים ערביות, ואנשים עם מוגבלות; (2) פרוץ ושכר נמוכים, בין היתר כתוצאה ממחסור בהון אנושי ומיומנויות בעיקר בקרב פרטים שאינם פונים ללימודים אקדמיים. רמות השכר והיקפי התעסוקה מובילים לכך שהכנסות חלק ממשקי הבית, ובפרט משקי הבית החרדים והערבים, הן עדיין נמוכות ולא ממצות את הפוטנציאל הכלכלי של אוכלוסיות אלו. משבר הקורונה והמלחמה חידדו את האתגרים הללו ואת הצורך במדיניות כוללת לתעסוקה והון אנושי.

על רקע אתגרי שוק העבודה המליצה וועדת התעסוקה 2030 על הצבת יעדים שאפתניים לשיעור התעסוקה ואיכותה, ועל שורה של כלים להשגת היעדים ובראשם רפורמה במערכת ההכשרה המקצועית ותוכניות תעסוקה לאוכלוסיות יעד. המלצות הוועדה אומצו על ידי הממשלה ומיושמות על ידי זרוע העבודה, ובמקביל מופעלות תוכניות תעסוקה, השכלה והכשרה רבות אחרות על ידי גופים בממשלה (שירות התעסוקה, משרד הכלכלה, המשרד לשוויון חברתי, רשות החדשנות, השירות האזרחי ועוד) ובמגזר השלישי. תוכניות אלו כרוכות בהשקעה ציבורית ניכרת, המגיעה למאות מיליוני שקלים בשנה. ואולם, כיום מידת האפקטיביות של מרבית הכלים הללו אינה ידועה. לתוכניות רבות אמנם בוצעו ומתבצעים מחקרים מלווים של איסוף נתונים, ובחלק קטן נעשה שימוש בקבוצות ביקורת ובניסויי אקראי מבוקר (RCT), אך מרביתם לא מתבצעים באופן המאפשר זיהוי נכון של האפקטיביות כפי שהמליצה וועדת התעסוקה 2030 ומפורט בנייר זה. משמעות הדבר היא שהקצאת משאבים ציבוריים בהיקף גדול נעשית ללא בסיס ראייתי מספק: תוכניות שאינן אפקטיביות עשויות להמשיך לפעול ולצרוך תקציבים על חשבון תוכניות שתרומתן גבוהה יותר; תוכניות שדורשות התאמות אינן מזוהות בזמן; ותוכניות חדשות מושקות ללא מנגנון מובנה ללמידה ושיפור. בהיעדר הערכה שיטתית, נוצר למעשה כשל שוק מוסדי - שבו אין מנגנון תיקון עצמי המבטיח שמשאבי הציבור מופנים לשימושים היעילים ביותר.

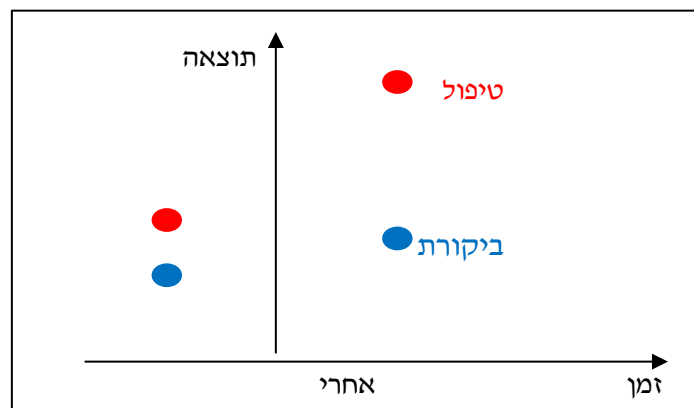
על מנת לשדרג ולטייב את הפעילות בכלל, ופעילות הממשלה בפרט, בתחום התעסוקה, לשמר ולפתח את הכלים היעילים ביותר להשגת היעדים ולהקצות את המשאבים באופן אופטימלי, ישנה חשיבות מעשית וכלכלית ממדרגה ראשונה בפיתוח מדיניות סדורה להערכת תוכניות. יש לבחון תוכניות (התערבויות) קיימות וחדשות בהתאם לאפקטיביות שלהן, כאשר המונח "אפקטיביות" מתייחס למידה שבה התוכנית משיגה את יעדיה ותורמת להשגת יעדי שיעור ואיכות התעסוקה, ובפרט ביחס למשאבים שהושקעו בה - קרי הפער בין התועלת והעלות של התוכנית למשק. לשם כך מציג נייר זה מסגרת כללית לבחינת ההשפעה של תוכניות הכשרה ותעסוקה אשר נועדו להגדיל את הזדמנויות התעסוקה או את פוטנציאל ההשתכרות של המשתתפים. פרק 2 מתאר את האתגרים בבחינת אפקטיביות, ופרק 3 מציג את השיטות המתודולוגיות המובילות להערכת אפקטיביות - ניסוי אקראי מבוקר, ניסוי טבעי, שיטות התאמה, הפרש הפרשים, עיצוב אי-רציפות רגרסיבית ומחקר אירוע - לצד דיון בבחינת ההשפעה הדינמית של תוכניות לאורך זמן, בליווי דוגמאות יישומיות מישראל. פרק 4 מתאר שימוש בתוצאות מחקרי אפקטיביות להערכת עלות-תועלת אישית וחברתית. פרק 5 מציג מודל לביצוע מחקרי אפקטיביות בישראל תוך שימוש בנתונים מנהליים של הלשכה המרכזית לסטטיסטיקה, עם או ללא מיזוג נתונים, ופרק 6 מציג דוגמה ליישום מודל זה במחקר האפקטיביות לתוכנית "קימם" לצעירים בחברה הערבית. פרק 7 מסכם בהמלצות מדיניות להקמת תשתית מוסדית לביצוע מחקרי אפקטיביות באופן שיטתי, לקביעת סטנדרטים למחקרי הערכה כתנאי למימון ממשלתי ולהנגשת נתונים מנהליים לצורך מחקר.

2. האתגר בהערכת אפקטיביות

האתגר המרכזי בהערכת אפקטיביות הוא לבודד ולזהות את השפעת ההתערבות הנחקרת על משתני התוצאה (למשל השכר) מיתר ההשפעות הרלוונטיות ומשינויים על פני זמן. באופן תיאורטי ההשפעה על הפרט היא ההפרש בין התוצאה שלו עם המדיניות ובלעדיה, אולם באופן מעשי לא ניתן לצפות **באותו פרט באותו זמן** עם המדיניות ובלעדיה. חשוב לזכור שלא די בהשוואת התוצאות של פרטים אחרי שעברו את ההתערבות לתוצאות שלהם לפני, שכן יתכן שבזמן זה קרו שינויים נוספים שאינם קשורים להתערבות; ומאידך לא די בהשוואת התוצאות של פרטים שעברו את ההתערבות לתוצאות של פרטים שלא עברו אותה, שכן יתכן שקיימים הבדלים נוספים בין הקבוצות שאינם קשורים להתערבות. לכן הערכת ההשפעה נעשית על ידי השוואה בין קבוצת אנשים שעברו את ההתערבות (קבוצת טיפול) לקבוצה דומה של אנשים שלא עברו התערבות (קבוצת ביקורת או קבוצת השוואה),¹ ולשם כך נדרשים נתונים על **שתי הקבוצות** לפחות **בשתי נקודות זמן** - לפני השינוי ואחריו. ההשפעה נטו של ההתערבות במקרה זה נאמדת באמצעות הפרש ההפרשים (difference in differences).

איור 1 מציג דוגמה פשוטה להמחשה. בדוגמה זו חלה עלייה בין התקופות בקבוצת הביקורת (הנקודה הכחולה הימנית לעומת הנקודה הכחולה השמאלית), ולכן ההפרש בין התקופות בקבוצת הטיפול הוא אומד מוטא כלפי מעלה להשפעת המדיניות, כלומר מעריך אותה ביתר. בנוסף נקודת הפתיחה של קבוצת הטיפול גבוהה מזו של קבוצת הביקורת (הנקודה האדומה השמאלית לעומת הנקודה הכחולה השמאלית), ולכן גם ההפרש בין הקבוצות בתקופה השנייה הוא אומד מוטא כלפי מעלה. ההשפעה נטו של ההתערבות - הפרש ההפרשים - היא ההפרש בין התקופות בקבוצת הטיפול פחות ההפרש בין התקופות בקבוצת הביקורת, או לחילופין ההפרש בין הקבוצות לאחר הטיפול פחות ההפרש בין הקבוצות לפני הטיפול.

איור 1: אילוסטרציה לבחינת שינוי מדיניות



¹ כאשר קבוצת האנשים הדומים למשתתפי התוכנית נבחרת באופן אקראי היא נקראת קבוצת ביקורת, בשאר המקרים היא נקראת קבוצת השוואה.

שתי הנחות עומדות בבסיס הערכת האפקטיביות באופן זה: (1) ללא התערבות, השינוי על פני זמן בקבוצת הטיפול הוא זהה לשינוי על פני זמן בקבוצת הביקורת (parallel trends); (2) ההשתייכות של פרטים לקבוצת הטיפול היא מקרית או נקבעת רק על פי המאפיינים הידועים שלהם, ואינה תוצאה של בחירה של הפרטים (selection). הנחות אלו הן חשובות במיוחד כאשר מדובר בהערכת אפקטיביות של תוכנית פיילוט ומטרת ההערכה היא לקבוע האם ליישם את התוכנית בקנה מידה רחב יותר. במקרה זה יש לוודא כי להערכת האפקטיביות יש תוקף חיצוני, כלומר שההשפעה הנאמדת עבור פרטים בקבוצת הטיפול שהשתתפו בתוכנית בקנה מידה קטן היא אכן ההשפעה הצפויה עבור פרטים שישתתפו בתוכנית בעתיד אם היא תיושם עבור האוכלוסייה כולה.

3. שיטות לאמידת אפקטיביות

ישנן מספר שיטות לאמידת האפקטיביות של התערבות. פרק זה מציג סיכום קצר של הגישות הנפוצות, עם התייחסות לסוג הנתונים הדרוש. נתחיל את הפרק עם תאור השיטה האופטימלית לבדיקת אפקטיביות ומציאת השפעה סיבתית, שיטת הניסוי האקראי המבוקר. בשיטה זו קבוצת הטיפול וקבוצת הביקורת נבחרות באופן אקראי. שיטה זו היא אמנם השיטה האופטימלית למדידת אפקטיביות, אך היא קשה עד בלתי אפשרית ליישום במקרים רבים כפי שיוסבר בהמשך. המשך הפרק יעסקו במדידת אפקטיביות כאשר לא נתונה קבוצת ביקורת או שקבוצת הטיפול לא נבחרה אקראית, במקרה זה נתאר מספר שיטות חלופיות ליצירת קבוצת הביקורת ולמדידת השפעת ההתערבות ATT.

3.1. ניסוי אקראי מבוקר (Randomized Controlled Trial - RCT)

ניסוי אקראי מבוקר נחשב במקרים רבים לשיטה המועדפת להערכת השפעת התערבות. בשיטה זו אוכלוסיית המועמדים לתוכנית מחולקת מראש באופן אקראי לשתי קבוצות - אלו שישתתפו בתוכנית (קבוצת הטיפול) ואלו שלא ישתתפו בה (קבוצת הביקורת). את הגרלת המועמדים לקבוצות ניתן לבצע על כל האוכלוסייה יחד או בחלוקה לחתכים, אם מעוניינים להגדיל את הסיכוי שפרטים מחתך מסויים יקבלו טיפול (מסיבות אתיות או על מנת להשיג יעדי מדיניות המתייחסים לקבוצות ספציפיות). כאשר ההגרלה היא אכן אקראית, מספר הפרטים בכל אחת מהקבוצות הוא גדול דיו ותפעול התוכנית נעשה באופן יעיל וללא סלקציה, שיטה זו מאפשרת לזהות בצורה נקייה את ההשפעה. חשוב לציין כי אין חובה שכל מי שהוקצה לתוכנית ישתתף בה בפועל, כל עוד ההשתתפות אינה סלקטיבית (כלומר גם החלוקה בין משתתפים ולא משתתפים בקבוצת הטיפול היא אקראית). דוגמא בולטת לשימוש בניסוי אקראי מבוקר היא עבודתם של שלוסר ושנן המעריכה את השפעת תוכנית "מעגלי תעסוקה" על התעסוקה וההכנסות של מקבלי הבטחת הכנסה, המתוארת בהמשך.

אולם עריכת ניסוי אקראי מבוקר אינה תמיד אפשרית או מתאימה, בפרט בתוכניות שההשתתפות בהן היא וולונטרית ואי השתתפות אינה כרוכה בסנקציה כלשהי. במקרה כזה לא ניתן לזהות את השפעת הטיפול אלא רק את השפעת הכוונה לטפל (intention to treat) ואת השפעת הטיפול על הפרטים ש"מצייתים" ומשתתפים בתוכנית. במקרים אחרים, עצם הסירוב לתת טיפול לפרטים שהוגרלו לקבוצת הביקורת עלול לגרום נזק ולכן יישום של ניסוי מבוקר יגרום יותר נזק מתועלת. כך למשל, מרכזי ההכוון התעסוקתי פונים באופן אקטיבי לפרטים הנמצאים מחוץ לשוק העבודה במטרה לסייע להם להשתלב בו, ונדרשים לשמור על מוניטין בקהילה שבה הם פועלים על מנת שיוכלו להמשיך ולמשוך משתתפים חדשים. החלטה אקראית על אי מתן שירות לחלק מהפונים לקבלת שירות במרכזים עלולה לפגוע במוניטין שלהם ובמטרה שלשמה הוקמו. בנוסף, קשה להשתמש בכלי של ניסוי אקראי מבוקר כאשר מספר המשתתפים הפוטנציאליים קטן או כאשר מדובר בטיפול שמותאם באופן אישי ומגוון הטיפולים גדול ביחס לגודל האוכלוסייה.

יתרונות: אופטימלי לזיהוי סיבתיות.

חסרונות: יקר, לא אתי, מצריך תכנון מוקדם ושיתוף פעולה מלא (לא ניתן להשתמש בו עבור תכניות התערבות וולנטריות).

דוגמא ליישום בישראל: תוכנית מעגלי תעסוקה

מעגלי תעסוקה היא תכנית השמה רב מגזרית בפריסה ארצית של שירות התעסוקה הישראלי. התכנית החלה במרץ 2014 כחלק משיתוף פעולה בין שירות התעסוקה וגיוינט ישראל - תב"ת ומטרתה היתה סיוע למשתתפי התכנית להשתלב בשוק העבודה, להתקדם בקריירה ולהתמיד בה, תוך ניתוק התלות בקצבאות ומניעת הכניסה לאבטלה עמוקה. התכנית יועדה לתובעי ותובעות הבטחת הכנסה, ודורשי עבודה חדשים ומחדשי רישום בלשכות התעסוקה. במסגרת התכנית המשתתפים זכו לליווי תעסוקתי הכולל אימון אישי וקבוצתי, סיוע בחיפוש משרות מתאימות וכלים נוספים לרכישת מיומנויות הנדרשות בשוק העבודה. כל זאת בנוסף להתייצבות והליווי בלשכה. התכנית לוותה במחקר RCT חיצוני על ידי צוות חוקרים מאוניברסיטת תל אביב, אנליה שלוסר וינון שן. המחקר נערך בין מרץ 2014 - מרץ 2016 והשתתפו בו כ- 15,000 תובעי הבטחת הכנסה. ההשתתפות בתכנית נקבעה עפ"י הקצאה רנדומלית על בסיס שבועי, כאשר התבצעה הרחבה הדרגתית של הלשכות המשתתפות. משתני התוצאה שנבדקו במחקר הם תשלומי ביטוח לאומי, תעסוקה, השתתפות בכח העבודה. המחקר התבסס על נתונים מנהליים משירות התעסוקה וכן על נתונים שנאספו בסקר טלפוני.

3.2. ניסוי טבעי (Natural Experiment)

במקרים שבהם לא ניתן לערוך ניסוי אקראי מבוקר, ניתן לעיתים להעריך את ההשפעה של תוכנית או מדיניות על ידי שימוש באירועים אקראיים או חיצוניים שהתרחשו במציאות - מה שמכונה "ניסוי טבעי". בניסויים טבעיים, מערך המחקר מנצל שינויים רגולטוריים, חוקים, או נסיבות חיצוניות שאינן תלויות בהחלטות הפרט או החוקרים, אך יצרו באופן טבעי הבחנה בין קבוצת אוכלוסייה שהושפעה מהשינוי (קבוצת הטיפול) לבין קבוצה דומה שלא הושפעה (קבוצת הביקורת). כאשר מתקיימים תנאים של הקצאה אקראית או אקראיות-למחצה, ניתן להתייחס לשינוי זה כאל "ניסוי" במובן הסיבתי, גם אם הוא לא תוכנן מראש על ידי החוקר.

דוגמה קלאסית לכך היא עבודתו של Angrist (1990) שבחנה את ההשפעה של השירות הצבאי על ההכנסות לטווח הארוך בארצות הברית. אנגריסט ניצל את הגרלות הגיוס ("draft lottery") שהתקיימו בתקופת מלחמת וייטנאם, שבמסגרתן הוקצה לכל גבר צעיר מספר רנדומלי שהכתיב את סיכויי גיוסו. מאחר שההגרלה הייתה אקראית, ניתן היה להשוות בין מי שהוגרלו לגיוס (קבוצת טיפול) לבין מי שלא הוגרלו (קבוצת ביקורת), וליחס את ההבדלים בהכנסות בעתיד להשפעת השירות הצבאי בלבד. המחקר התבסס על נתונים מנהליים של הביטוח הלאומי האמריקאי, שאפשרו מעקב ארוך טווח אחר ההכנסות, והציג תוצאה נגד אינטואיציה: שירות צבאי לא העלה, ולעיתים אף הפחית, את ההכנסות לטווח הארוך. תנאי האקראיות בהגרלת הגיוס הפכו את המחקר לאחד המוקדים הראשונים של שימוש בניסויים טבעיים בכלכלה אמפירית.

מחקר מכונן נוסף הוא זה של Card and Krueger (1995) שבחנו את השפעת העלאת שכר המינימום על שיעורי התעסוקה. החוקרים ניצלו שינוי רגולטורי חיצוני - העלאת שכר המינימום במדינת ניו ג'רזי בשנת 1992 - והשוו את השינוי בתעסוקה בענף המזון המהיר במדינה זו לעומת מדינת פנסילבניה השכנה, שבה שכר המינימום נותר ללא שינוי. ההנחה הייתה ששתי המדינות דומות במאפייני השוק ובמגמות הכלכליות, כך שניתן לראות בהבדל בשיעורי התעסוקה לאחר השינוי כהשפעת ההתערבות. תנאי הניסוי הטבעי התקיים משום שהשינוי בשכר המינימום התרחש במיקום אחד בלבד ובאופן בלתי תלוי בהחלטות המעסיקים או העובדים. בכך הצליחו החוקרים לבדוד את ההשפעה הסיבתית של שכר מינימום על תעסוקה, ולהפריך מיתוס רווח לפיו העלאת שכר מינימום בהכרח מפחיתה תעסוקה.

באופן כללי, ניסויים טבעיים מבוססים על הזדמנויות מחקריות שבהן ההיסטוריה או המדיניות הציבורית יצרו באופן מקרי תנאי "ניסוי" במציאות. הצלחת השיטה תלויה במידת האקראיות של האירוע, בדמיון בין קבוצות ההשוואה, ובזמינות נתונים איכותיים לפני ואחרי ההתערבות.

יתרונות: מנצל אירועים חיצוניים או שינויים רגולטוריים כאילו היו ניסוי אקראי, מאפשר לזהות השפעות סיבתיות גם ללא הקצאה ניסויית.

חסרונות: רגיש להנחות לגבי האקראיות של האירוע ודורש זהות במגמות בין הקבוצות; לעיתים קשה להוכיח שתנאי הניסוי הטבעי אכן מתקיימים במלואם. תלוי במזל.

דוגמה ליישום בישראל: איכות בתי ספר

דוגמה ישראלית בולטת היא עבודתם של Gould, Lavy and Paserman (2004) שבחנו את השפעת איכות בית הספר על הישגי עולים מאתיופיה בישראל. החוקרים ניצלו מדיניות הקצאה חצי-אקראית של עולים בין מרכזי קליטה, שלא בוצעה על פי מאפייניהם האישיים אלא בהתאם לשיקולים לוגיסטיים של קליטה. מאחר שההקצאה לא נבעה מבחירה של העולים עצמם, ניתן לראות בה מעין ניסוי טבעי, שבו "הטיפול" הוא החשיפה לבתי ספר באיכות שונה בהתאם למיקום המרכז. החוקרים השתמשו בנתוני מבחנים וציונים לאורך זמן כדי לאמוד את השפעת איכות בית הספר על הישגי התלמידים, ומצאו השפעה משמעותית ומתמשכת של תנאי החינוך הראשוניים על תוצאות עתידיות. תנאי הניסוי הטבעי התקיים משום שההקצאה הראשונית בין מרכזי הקליטה הייתה חיצונית לפרטים ולא תלויה במאפייניהם.

3.3. שיטות התאמה

3.3.1. בניית קבוצת השוואה על ידי התאמת מאפיינים נצפים (Selection on)

(Observables)

כאשר לא ניתן לערוך ניסוי מבוקר או לנצל אירוע חיצוני לצורך ניסוי טבעי, ניתן להעריך את האפקטיביות של תוכנית באמצעות **בניית קבוצת השוואה מלאכותית** הדומה ככל האפשר למשתתפים בתוכנית במאפייניהם הנצפים. גישה זו נשענת על ההנחה כי ההשתתפות בתוכנית מוסברת על ידי מאפיינים נצפים בלבד - כלומר, שאין משתנים סמויים (בלתי נצפים) המשפיעים הן על ההשתתפות בתוכנית והן על התוצאה הנמדדת. במסגרת גישה זו (Rubin, 1973), החוקר מבצע **בקרה על משתני רקע רלוונטיים** - כגון גיל, מין, השכלה, מצב משפחתי, הכנסה קודמת, השכלת הורים, אזור מגורים, או ניסיון תעסוקתי - כדי להתאים לכל משתתף בתוכנית פרט דומה שלא השתתף בה - "תאום מלאכותי". כיון שקבוצת ההשוואה נבנתה כך שהיא זהה במאפיינים לקבוצת הטיפול, ניתן לפרש את ההבדל הממוצע בתוצאות בין שתי הקבוצות כ-ATT (Average Treatment Effect on the Treated).

ככל שההתאמה מבוססת על מספר רב ומדויק יותר של מאפיינים רלוונטיים, כך ההערכה של האפקטיביות תהיה אמינה יותר. עם זאת, קיים תמיד החשש שהמשתנים שנבחרו אינם מכסים את כל ההבדלים המהותיים בין הקבוצות. בפרט, כאשר משתנים כמו מוטיבציה, כישורים רכים, או רשתות חברתיות אינם נמדדים - עלולה להתקבל הטיית בחירה (selection bias). לכן, השיטה מתאימה בעיקר למקרים שבהם ישנם נתונים עשירים ומפורטים על משתתפי התוכנית ועל כלל האוכלוסייה.

יתרונות: מאפשרת לזהות השפעות סיבתיות בהיעדר ניסוי אקראי; גמישה ליישום במגוון תוכניות; מתבססת על נתונים קיימים.

חסרונות: תלויה באיכות ורוחב הנתונים; אינה מתמודדת עם הטיית הנובעות ממשתנים בלתי נצפים; דורשת הנחת Selection on Observables שקשה לעיתים להצדיק אמפירית.

דוגמא ליישום בישראל: תוכנית קימס לחברה הערבית

בישראל, גישה זו מיושמת במחקר אפקטיביות לתוכנית "קימס" לחברה הערבית, שנועדה לשפר את שילובם של צעירים ערבים באקדמיה בישראל. החוקרים יצרו קבוצת ביקורת על בסיס נתונים מנהליים של הלמ"ס, תוך התאמת משתתפי התוכנית לפרטים דומים במאפייניהם אך שלא נטלו בה חלק. ניתוח זה אפשר לבחון את ההשפעה של השתתפות בתוכנית על שיעורי הפונים ללימודים אקדמיים בישראל תוך שמירה על דמיון מרבי במאפייני רקע. להרחבה על מחקר זה ראו פרק 6.

3.3.2. בניית קבוצת השוואה על ידי התאמת הסתברויות לקבלת טיפול (Propensity Score Matching)

שיטה זו מהווה הרחבה מתוחכמת של עקרון ההתאמה (matching) ומתמקדת בהתאמת ההסתברויות להשתתפות בתוכנית במקום התאמת מאפיינים אחד לאחד. הרעיון המרכזי הוא שבמקום למצוא "תאום מלאכותי" לכל משתתף, די למצוא פרט שלא השתתף בתוכנית אך ההסתברות שלו להשתתף בתוכנית (propensity score) דומה לזו של המשתתף.

ההסתברות להשתתפות נאמדת באמצעות מודל בחירה (לוגיט או פרוביט), שבו המשתנה התלוי הוא ההשתתפות בתוכנית, והמשתנים המסבירים הם מאפייני רקע נצפים - כמו גיל, השכלה, ניסיון תעסוקתי, אזור מגורים ועוד. לאחר חישוב ההסתברויות לכל פרט באוכלוסייה, נבנית קבוצת ביקורת על בסיס פרטים שלא השתתפו בתוכנית אך בעלי הסתברות דומה להשתתף בה. השוואת התוצאות בין שתי הקבוצות מאפשרת לאמוד את השפעת ההשתתפות בהנחה של common support - כלומר, קיים טווח משותף רחב מספיק של הסתברויות לשתי הקבוצות.

שיטה זו מצמצמת את בעיית הממדים (curse of dimensionality) האופיינית לשיטות התאמה ישירה ומאפשרת שימוש יעיל גם במקרים שבהם קיימים משתנים רבים או כאשר קבוצת הביקורת קטנה יחסית. עם זאת, גם כאן מדובר בהנחת בחירה על סמך משתנים נצפים בלבד, ולכן ההערכה עשויה להיות מוטת אם קיימים משתנים סמויים המשפיעים על ההשתתפות או על התוצאה.

יתרונות: גמישה יחסית ליישום; מאפשרת יצירת קבוצת ביקורת גם במדגמים קטנים; מבוססת על הסתברות כוללת להשתתפות במקום על התאמה ישירה.

חסרונות: דורשת הנחת common support רחבה; אינה מתמודדת עם משתנים בלתי נצפים; תוצאותיה תלויות מאוד באיכות הנתונים ובניסוח מודל הבחירה.

דוגמא ליישום בישראל: תוכנית אורות לתעסוקה

שיטה זו שימשה במחקר ההערכה של תוכנית "אורות לתעסוקה" (Wisconsin Plan) שבוצע על ידי מכון מאיירס-ג'וניט-ברוקדייל. החוקרים בנו קבוצת ביקורת של תובעי הבטחת הכנסה שלא השתתפו בתוכנית, לאחר שאמדו את ההסתברות להשתתף בה לפי מאפייני רקע סוציו-דמוגרפיים. באמצעות השוואה בין תוצאות שתי הקבוצות (שיעורי תעסוקה, הכנסה, תלות בקצבאות), הם הצליחו לאמוד את ההשפעה של התוכנית על שיפור בתעסוקה ובשכר. תנאי ה-common support התקיים הודות למדגם גדול ולנתונים מנהליים מפורטים שאפשרו לבצע התאמה איכותית.

ראוי לציין כי השוואת שתי הדוגמאות - מעגלי תעסוקה (סעיף 3.1) ואורות לתעסוקה - ממחישה נקודה חשובה לגבי עלות מחקר האפקטיביות עצמו. מחקר האפקטיביות של תוכנית מעגלי תעסוקה התבסס על RCT, שיטה הדורשת תכנון מוקדם, תיאום מלא עם הגוף המפעיל, הקצאה אקראית של משתתפים ומעקב צמוד - תהליך יקר ומורכב. לעומת זאת, מחקר האפקטיביות של אורות לתעסוקה התבסס על PSM תוך שימוש בנתונים מנהליים קיימים - שיטה זולה משמעותית שאינה דורשת התערבות בתפעול התוכנית. חרף ההבדל בעלות המחקר ובמורכבותו, שני המחקרים סיפקו תובנות בעלות ערך לקובעי מדיניות. יתרה מכך, תוצאותיהם ניתנות להשוואה: השפעת מעגלי תעסוקה על שיעור התעסוקה באוכלוסיית בעלי ותק עד שנה בהבטחת הכנסה (15 נקודות אחוז) גבוהה מההשפעה של אורות לתעסוקה באוכלוסייה זו (8 נקודות אחוז), בעוד שעלות תוכנית מעגלי תעסוקה (כ-1,400 ש"ח למשתתף לשנה) נמוכה בסדר גודל מעלות אורות לתעסוקה (כ-12,000 ש"ח למשתתף). תובנה זו - שניתן להשיג תוצאות דומות או טובות יותר בעלות נמוכה יותר - לא הייתה מתאפשרת ללא מחקרי אפקטיביות לשתי התוכניות, ומדגימה כי גם מחקרי אפקטיביות שאינם מבוססים על RCT, ושעלותם נמוכה, מספקים בסיס ראייתי בעל ערך רב להקצאת משאבים ציבוריים.

3.4. הפרש ההפרשים (Difference-in-Differences - DiD)

שיטת הפרש ההפרשים (Difference-in-Differences) היא אחת מהשיטות הנפוצות ביותר להערכת אפקטיביות של תוכניות ומדיניות ציבורית כאשר לא ניתן לבצע הקצאה אקראית של המשתתפים. מטרתה לזהות את ההשפעה הסיבתית של ההתערבות על משתנה התוצאה (כגון שכר, תעסוקה או הישגים לימודיים), תוך נטרול שינויים כלליים שהתרחשו במקביל במערכת כולה. העיקרון הבסיסי פשוט: נבחן את השינוי בממוצע משתנה התוצאה **בקבוצת הטיפול לפני ואחרי ההתערבות, ונשווה אותו לשינוי המקביל בקבוצת הביקורת** שלא השתתפה בהתערבות. ההבדל בין שני השינויים - כלומר, "הפרש ההפרשים" - מייצג את ההשפעה נטו של ההתערבות, בהנחה שהשינוי בקבוצת הביקורת משקף את מה שהיה קורה לקבוצת הטיפול אילו לא הייתה נחשפת להתערבות.

תנאים לתוקף השיטה

כדי שהאומד של DiD יהיה בעל משמעות סיבתית, נדרשת עמידה בכמה הנחות מרכזיות:

1. מגמות מקבילות (Parallel Trends): ההנחה החשובה ביותר היא כי לולא ההתערבות, מגמת השינוי במשתנה התוצאה בקבוצת הטיפול הייתה זהה לזו שבקבוצת הביקורת. כלומר, שני הקווים היו "מקבילים" לאורך זמן. הפרה של הנחה זו תוביל להטיה באמידה, ולכן נהוג לבדוק אמפירית האם המגמות אכן דומות בתקופה שלפני ההתערבות.
2. היעדר השפעות עקיפות (No Spillover Effects): ההתערבות לא משפיעה באופן עקיף גם על קבוצת הביקורת. לדוגמה, אם תוכנית תעסוקה באזור מסוים משפיעה על ביקוש לעובדים גם באזור סמוך, השוואת שני האזורים לא תבודד את האפקט הנקי.
3. שינויים חיצוניים זהים: כל השפעה חיצונית (כמו מיתון, שינוי מיסוי, או רפורמה מערכתית אחרת) משפיעה באופן דומה על שתי הקבוצות, כך שהשפעתה מנוטרלת בהפרש.

שיטת DiD מבוססת על שימוש בנתוני **פאנל או נתוני חתך חוזר** (repeated cross-sections) הכוללים לפחות שתי נקודות זמן: לפני ההתערבות ואחריה. יתרונה המרכזי הוא בכך שניתן ליישמה גם כאשר ההקצאה לטיפול אינה אקראית, כל עוד יש קבוצת השוואה דומה דיה ומידע זמין לאורך זמן.

דוגמה קלאסית לשימוש בשיטה היא עבודתם של Card ו-Krueger (1995) שהזכרה לעיל, אשר השוו את השינוי בשיעורי התעסוקה בענף המזון המהיר בניו-ג'רזי (שבה הועלה שכר המינימום) לפנסילבניה (שבה לא שונה החוק). באמצעות השוואת השינויים בין המדינות לפני ואחרי הרפורמה, הצליחו החוקרים לבדוד את ההשפעה של העלאת שכר מינימום מתנודות מאקרו-כלכליות כלליות.

בעשור האחרון פותחו הרחבות לשיטת DiD המאפשרות לטפל במצבים מורכבים יותר, כמו הטמעה מדורגת של תוכניות (staggered adoption) או טיפולים משתנים בזמן. הרחבות אלו מאפשרות לחשב ממוצע השפעה משוקללים לאורך זמן (Weighted Average Treatment Effects) ולהתמודד עם מגבלות של גישות קלאסיות. יישומים עדכניים משתמשים גם במודלים של event study הבוחנים את התפתחות ההשפעה לאורך השנים שלאחר ההתערבות, וכך בוחנים את יציבות התוצאות ואת קיום המגמות המקבילות לפני האירוע. להרחבה על שיטות אלו ראו סעיף 3.6.

יתרונות: שיטה פשוטה יחסית; מאפשרת זיהוי סיבתי גם בהיעדר ניסוי אקראי; ניתנת ליישום על נתוני מנהליים קיימים; רבת שימוש במחקרי מדיניות.

חסרונות: דורשת נתונים לאורך זמן על שתי קבוצות; תלויה בהנחת המגמות המקבילות; רגישה לשינויים חיצוניים שאינם נצפים.

דוגמא ליישום בישראל: תוכניות "עתודה מדעית-טכנולוגית" ו"לתת חמש"

עבודתם של זוסמן ומעגן (2018) שבחנו את השפעת תוכניות "עתודה מדעית-טכנולוגית" ו"לתת חמש" על הישגים לימודיים ובחירה במקצועות מדעיים בקרב תלמידי תיכון. החוקרים השוו את השינוי בשיעור התלמידים הלומדים חמש יחידות מתמטיקה בבתי ספר שהשתתפו בתוכניות לעומת בתי ספר שלא השתתפו בהן, לפני ואחרי תחילת ההתערבות. הנתונים כללו סדרות זמן ארוכות של נתוני בגריות מבית הספר והלמ"ס, שאפשרו לוודא את קיום מגמות מקבילות טרם ההתערבות.

המחקר מצא כי שיעור הלומדים מתמטיקה בהיקף חמש יחידות עלה משמעותית בבתי הספר שהשתתפו בתוכניות, ומצא עדויות להשפעה חיובית גם על תחומי מדע נוספים. תנאי השיטה התקיימו הודות לכך שהבחירה של בתי הספר בתוכנית נקבעה ברמה המערכתית (ולא ברמת התלמידים), והייתה מדורגת על פני זמן - דבר שאפשר להבחין בין קבוצות טיפול וביקורת מובהקות.

3.5. עיצוב אי-רציפות רגרסיבית (Regression Discontinuity Design - RDD)

שיטת עיצוב אי-רציפות רגרסיבית (Regression Discontinuity Design - RDD) מנצלת מצבים שבהם הזכאות לתוכנית או להתערבות נקבעת על פי **סף מוגדר של משתנה רציף** - כגון גיל, הכנסה, ציון מבחן או גודל כיתה. הרעיון המרכזי הוא שפרטים הנמצאים מעט מעל הסף ומעט מתחתיו דומים מאוד זה לזה בכל מאפייניהם, ולכן ההבדל היחיד ביניהם הוא עצם החשיפה להתערבות. השוואת התוצאות של שתי הקבוצות הסמוכות לסף מאפשרת לאמוד את ההשפעה הסיבתית של ההתערבות באופן נקי מהטיות בחירה. השיטה מחייבת מספר תנאים: (1) הסף קובע את הזכאות באופן חד וברור (או לפחות משפיע משמעותית על ההסתברות להשתתף, במקרה של Fuzzy RDD); (2) הפרטים אינם יכולים לבחור באיזה צד של הסף להימצא (אין מניפולציה); (3) כל המאפיינים האחרים של הפרטים משתנים באופן חלק סביב הסף. כאשר תנאים אלו מתקיימים, שיטת RDD נחשבת לאחת השיטות האמינות ביותר לזיהוי סיבתי, ולעיתים מכונה "הקרובה ביותר לניסוי מבוקר" מבין השיטות הלא-ניסוייות.

השיטה רלוונטית במיוחד להערכת תוכניות בשוק העבודה הישראלי, שכן תוכניות רבות מבוססות על ספי זכאות: גיל, רמת הכנסה, ציון מבחן סף, מספר ילדים, או גודל יישוב. בכל מקרה שבו הזכאות להתערבות נקבעת לפי ערך סף של משתנה מדיד, ניתן ליישם RDD.

יתרונות: זיהוי סיבתי חזק מאוד הקרוב לניסוי מבוקר; אינה דורשת הקצאה אקראית; ניתנת לאימות אמפירי באמצעות בדיקות שקיפות סביב הסף.

חסרונות: מוגבלת למצבים שבהם קיים סף זכאות ברור; ההשפעה הנאמדת היא מקומית - תקפה רק עבור פרטים הסמוכים לסף ואינה בהכרח ניתנת להכללה לכלל האוכלוסייה; דורשת מספר גדול של תצפיות סביב הסף.

דוגמא ליישום בישראל: גודל כיתות לימוד

עבודתם הפורצת של Angrist and Lavy (1999) היא אחד המחקרים המכוננים בשימוש בשיטת RDD בכלכלה. החוקרים ניצלו את "חוק מימונידס" הקובע שגודל כיתה מקסימלי הוא 40 תלמידים - כך שכאשר מספר התלמידים בשכבה עולה על 40, בית הספר חייב לפצל לשתי כיתות. קפיצה זו בגודל הכיתה סביב סף 40 התלמידים (וכפולותיו) יצרה שונות אקראית-למחצה בגודל הכיתות, שאפשרה לאמוד את ההשפעה הסיבתית של גודל כיתה על ציוני מבחנים. המחקר מצא כי הקטנת כיתות שיפרה משמעותית את הישגי התלמידים, במיוחד בבתי ספר המשרתים אוכלוסיות חלשות.

מחקר ישראלי נוסף המיישם שיטה זו הוא עבודתם של Battistin and Meroni (2016) שבחנו את ההשפעה של ספי זכאות לתוכניות סיוע ברשויות מקומיות בישראל על תוצאות תעסוקתיות וחינוכיות. השיטה רלוונטית גם להערכת תוכניות שבהן ספי הכנסה או גיל קובעים זכאות, כגון תוכניות סבסוד מעונות, קורסי הכשרה מקצועית מותנים בגיל, או סיוע למובטלים ארוכי טווח.

3.6. בחינת ההשפעה הדינמית של תוכניות התערבות

השיטות שתוארו בסעיפים הקודמים מניחות לרוב כי השפעת ההתערבות ניתנת לסיכום באומדן יחיד. אולם בפועל, השפעתן של תוכניות תעסוקה והכשרה משתנה לאורך הזמן: תוכנית הכשרה עשויה להציג השפעה שלילית בטווח הקצר (בתקופת הלימודים, "אפקט נעילה") אך חיובית וגדלה בטווח הארוך; תוכנית השמה עשויה להציג דפוס הפוך של השפעה מיידית הדועכת עם הזמן; ומדיניות רגולטורית עשויה להציג השפעה קבועה. הבחנה בין דפוסים אלו חיונית, שכן תוכנית שנראית "כושלת" בטווח הקצר עשויה להיות מוצלחת ביותר בטווח הארוך - ולהיפך.

קיימים שני כלים עיקריים לבחינת ההשפעה הדינמית, הנבדלים במטרתם וביישומם. מחקר אירוע (Event Study) הוא בעיקרו כלי ויזואלי ואבחוני: הוא משרטט את נתיב ההשפעה לאורך זמן באמצעות מקדמים נפרדים לכל תקופה, ומשמש לשתי מטרות - לבדוק אם הנחת המגמות המקבילות מתקיימת (בתקופה שלפני ההתערבות) ולתאר את הדינמיקה של ההשפעה (בתקופה שלאחריה). שיטה זו מתאימה במיוחד כאשר כל קבוצת הטיפול נחשפת להתערבות בו-זמנית. לעומת זאת, הפרש הפרשים דינמי (Dynamic DiD) הוא מסגרת אמידה שפותחה כדי להתמודד עם מצב של הטמעה מדורגת (staggered adoption) - כאשר קבוצות שונות נכנסות לתוכנית בזמנים שונים. במצב כזה, ניתוח Event Study סטנדרטי עלול לתת תוצאות מוטות, ונדרשים אומדנים מתוקנים המפרידים בין קבוצות לפי מועד כניסתן. בפועל, ניתן לשלב בין הגישות: להשתמש ב-Event Study כדי לשרטט גרפית את נתיב ההשפעה, ובאומדני Dynamic DiD כדי לחשב את ההשפעות המתוקנות כאשר ההטמעה מדורגת.

3.6.1. מחקר אירוע (Event Study)

שיטת מחקר אירוע (Event Study) היא הרחבה חשובה של שיטת הפרש ההפרשים, המאפשרת לבחון את התפתחות ההשפעה של התערבות לאורך זמן - הן לפני ההתערבות והן אחריה. בעוד ששיטת DiD הקלאסית מספקת אומד יחיד של השפעת ההתערבות, מחקר אירוע מציג את ההשפעה בכל תקופה בנפרד ביחס לתקופת בסיס (בדרך כלל התקופה שלפני ההתערבות).

היתרון המרכזי של השיטה הוא כפול: ראשית, בחינת המקדמים בתקופות שלפני ההתערבות מאפשרת לבדוק את תקפות הנחת המגמות המקבילות - אם המקדמים בתקופות שלפני ההתערבות קרובים לאפס ואינם מובהקים, הדבר מחזק את האמון שלא היו הבדלים שיטתיים בין הקבוצות קודם להתערבות. שנית, המקדמים בתקופות שלאחר ההתערבות מאפשרים לבחון את הדינמיקה של ההשפעה - האם היא מיידית או מתגבשת בהדרגה, האם היא מתמשכת או דועכת, וכיצד היא מתפתחת לאורך תקופות שונות.

בהקשר של תוכניות תעסוקה והכשרה, שיטת מחקר אירוע חיונית במיוחד: היא מאפשרת לזהות את תקופת ה"נעילה" (lock-in effect) שבה משתתפים בתוכנית הכשרה חווים ירידה זמנית בתעסוקה ובשכר (בעת השתתפותם בקורס), ולעקוב אחר ההתאוששות וההשפעה החיובית שמתגבשת רק לאחר סיום ההכשרה. ללא מחקר אירוע, הערכת אפקטיביות המתבססת על ממוצע יחיד עלולה לערב את ההשפעה השלילית הזמנית עם ההשפעה החיובית ארוכת הטווח, ובכך להוביל לתוצאה מוטת.

יתרונות: מאפשרת לבדוק את הנחת המגמות המקבילות באופן אמפירי; חושפת את הדינמיקה של ההשפעה לאורך זמן; מאפשרת לזהות השפעות מושהות ותקופות "נעילה"; מתאימה גם להטמעה מדורגת של תוכניות.

חסרונות: דורשת נתוני פאנל עם מספר תקופות רב לפני ואחרי ההתערבות; רגישה לבחירת תקופת הבסיס ולאופן הגדרת חלון הזמן; בהטמעה מדורגת עלולה לסבול מהטיות הנובעות משימוש בקבוצות שכבר טופלו כביקורת.

דוגמא ליישום בישראל: תוכנית "מענק עבודה" (מס הכנסה שלילי)

עבודתם של Brender and Strawczynski (2019) שבחנו את השפעת תוכנית "מענק עבודה" (מס הכנסה שלילי) על שיעורי התעסוקה וההכנסות של עובדים בשכר נמוך בישראל. החוקרים השתמשו בגישה מחקר אירוע כדי לעקוב אחר השינויים בתעסוקה ובשכר של הזכאים לתוכנית לאורך מספר שנים לפני ואחרי הפעלתה. ניתוח מחקר האירוע אפשר להם לבדוק את הנחת המגמות המקבילות בתקופה שלפני הפעלת התוכנית, ולהצביע על עלייה הדרגתית בשיעורי התעסוקה בקרב הזכאים בשנים שלאחר תחילתה.

מחקר ישראלי נוסף שמשלב שיטת מחקר אירוע הוא עבודתם של Lavy and Schlosser (2005) שבחנו את ההשפעה של תוכניות תמריצים לבתי ספר על שיעורי הברגרות, תוך מעקב אחר ההשפעה לאורך מספר שנים לפני ואחרי הפעלת התוכנית. ניתוח מחקר האירוע איפשר להבחין בין שינויים מידיים להשפעות מצטברות ולהעריך את קיימות ההשפעה לאורך זמן.

3.6.2. הפרש ההפרשים הדינמי (Dynamic DiD)

שיטת DiD הקלאסית (סעיף 3.4) מניחה שההשפעה של ההתערבות קבועה לאורך זמן ומספקת אומדן ממוצע יחיד. בפועל, הנחה זו לעיתים קרובות אינה מתקיימת: השפעת תוכנית הכשרה עשויה להתחזק עם הזמן ככל שהמשתתפים מיישמים את המיומנויות שרכשו, ואילו השפעת תוכנית השמה עלולה לדעוך. שיטת הפרש ההפרשים הדינמי מרחיבה את מסגרת ה-DiD הקלאסית ומאפשרת לאמוד השפעה נפרדת לכל תקופה לאחר ההתערבות. באופן פורמלי, במקום לכלול במודל משתנה דמי יחיד לתקופה שלאחר ההתערבות, נכללים משתנים דמיים נפרדים לכל תקופת זמן ביחס לנקודת ההתערבות (למשל: שנה אחרי, שנתיים אחרי, שלוש שנים אחרי). המקדמים של משתנים אלו מאפשרים לשרטט את "נתיב ההשפעה" של התוכנית לאורך זמן.

בעשור האחרון התגלה כי היישום הסטנדרטי של DiD דינמי באמצעות רגרסיה עם אפקטים קבועים דו-כיווניים (Two-Way Fixed Effects - TWFE) עלול להניב אומדים מוטעים כאשר מתקיימים שני תנאים בו-זמנית: ההשפעה משתנה לאורך זמן (heterogeneous treatment effects), וההטמעה של התוכנית היא מדורגת (staggered adoption) - כלומר, קבוצות שונות נחשפות להתערבות בזמנים שונים. במצב כזה, אומדן ה-TWFE הקלאסי משתמש באופן מרומז בקבוצות שכבר טופלו כ"ביקורת" עבור קבוצות שטופלו מאוחר יותר, מה שעלול ליצור הטיה ואף היפוך סימן של ההשפעה הנאמדת.

הספרות העדכנית פיתחה אומדנים חלופיים המתמודדים עם בעיה זו. Callaway and Sant'Anna (2021) מציעים לחשב השפעה ממוצעת נפרדת לכל "קבוצה" (המוגדרת לפי מועד תחילת הטיפול) ולכל תקופת זמן ביחס לנקודת ההתערבות, ולאחר מכן לשקלל את ההשפעות הספציפיות לאומדן כולל. de Chaisemartin and D'Haultfoeulle (2020) מציעים גישה דומה המתמקדת בהשפעה המיידית של שינוי במצב הטיפול. Sun and Abraham (2021) מפתחים אומדן שמתמודד ספציפית עם הטיה בניתוח event study כאשר ההטמעה מדורגת. הכלל המנחה הוא שבכל ההשוואות, קבוצת הביקורת צריכה להיות מורכבת אך ורק מיחידות שטרם טופלו ("not-yet-treated") או שלעולם לא יטופלו ("never-treated").

שיטה זו דורשת נתוני פאנל (מעקב אחר אותם פרטים או יחידות לאורך זמן) עם מספר תקופות רב מספיק הן לפני ההתערבות (לצורך בדיקת מגמות מקבילות) והן אחריה (לצורך שרטוט נתיב ההשפעה). בהקשר של תוכניות תעסוקה, נדרשים לפחות נתונים שנתיים על תעסוקה ושכר לשלוש שנים לפני ההתערבות ולשלוש שנים לאחריה. כאשר נתונים כאלו זמינים, ניתן להבחין בבירור בין תקופת "נעילה" (lock-in) לבין ההשפעה ארוכת הטווח.

יתרונות: מאפשרת לעקוב אחר התפתחות ההשפעה לאורך זמן; מתמודדת עם הטמעה מדורגת של תוכניות; מאפשרת לזהות אפקט "נעילה" והשפעות מושהות; מספקת בסיס להחלטות מדיניות מושכלות לגבי עיתוי המדידה.

חסרונות: דורשת נתוני פאנל עם מספר רב של תקופות; מורכבת יותר מבחינה חישובית; דורשת בחירות מתודולוגיות (הגדרת קבוצת הביקורת, שקלול) שעשויות להשפיע על התוצאות; דורשת מדגם גדול מספיק לאמידת השפעות ספציפיות לכל קבוצה ותקופה.

דוגמה ליישום בישראל: תוכנית "מעגלי תעסוקה"

תוכנית "מעגלי תעסוקה" (סעיף 3.1) מהווה דוגמה ישראלית מובהקת למקרה שבו שיטת Dynamic DiD רלוונטית. התוכנית הוטמעה בהדרגה (staggered adoption): לשכות תעסוקה שונות הצטרפו לתוכנית בזמנים שונים, כך שבכל נקודת זמן חלק מהלשכות כבר הפעילו את התוכנית ואחרות טרם החלו. מבנה זה מאפשר לנצל את השונות בעיתוי ההטמעה כדי לאמוד את ההשפעה הדינמית, תוך שימוש בלשכות שטרם הצטרפו כקבוצת ביקורת. יתרה מכך, תוכניות כמו "קימס" מצריכות מעקב רב-שנתי כדי לבחון האם ההשפעה על כניסה ללימודים אקדמיים מתרגמת גם להשפעה על סיום תואר ועל שילוב בשוק העבודה. שימוש באומדנים דינמיים מאפשר לזהות את הנקודה שבה ההשפעה מגיעה לשיאה, האם היא מתמשכת, ובאיזה שלב כדאי לבצע הערכת ביניים שתספק מידע מהימן לקובעי מדיניות.

4. עלות-תועלת אישית וחברתית

התועלת מתוכנית התערבות נובעת מהשפעת ההשתתפות בה על הזדמנויות התעסוקה ועל השכר בהווה ובעתיד. עבור הפרט, התועלת מושפעת גם מהעדפותיו ומהוויתור על הכנסה פוטנציאלית ועל פנאי בתקופה שבה הוא משתתף בתוכנית. לפיכך **הרווח האישי** שמפיק הפרט מקיומה של התוכנית הוא השינוי בערך הנוכחי של תוחלת התועלת שלו על פני אופק חייו כאשר התוכנית קיימת, אל מול המצב שבו היא אינה קיימת. אם מדובר בתוכנית שההשתתפות בה היא חובה, כגון תוכנית "מרווחה לעבודה" למקבלי הבטחת הכנסה אשר נדרשים להשתתף בה כתנאי להמשך קבלת הקצבה, הרי שהרווח האישי יכול להיות שלילי; במילים אחרות, יתכן שקיומה של תוכנית החובה מרע את מצבם של פרטים שבחרו להשתתף בה, כמו גם של פרטים שבחרו לא להשתתף בה ("נקנסו" על כך). לעומת זאת אם מדובר בתוכנית שההשתתפות בה היא וולנטרית, כגון תוכנית הכשרה מקצועית, הרי שהרווח האישי הוא בהכרח חיובי לפרטים שבחרו להשתתף בה ואפס לפרטים שבחרו לא להשתתף בה.

תחת ההנחה שעליית השכר משקפת את עליית התפוקה של הפרט כתוצאה ישירה של התוכנית שהשתתף בה, **התשואה החברתית** מהתוכנית מחושבת על ידי חיסור העלויות שלה מהעלייה הצפויה בשכר הפרט בפועל, קרי הערך הנוכחי של זרם ההכנסות הצפויות משכר על פני אופק חייו כאשר התוכנית קיימת אל מול המצב שבו היא אינה קיימת.² סך ההוצאה הממשלתית עבור התוכנית כולל תשלומי העברה למשתתפים בתוכנית, אם הם זכאים לכך (למשל דמי אבטלה או דמי קיום); תשלומי העברה לפרטים שאינם עובדים ואינם משתתפים בתוכנית, אם הם זכאים לכך (למשל קצבת הבטחת הכנסה); והעלות הישירה של התוכנית (למשל קורס הכשרה או רכזי תוכנית תעסוקה).

נדגיש כי הרווח האישי והתשואה החברתית מההשתתפות בתוכנית תלויים הן בפרמטרים הנאמדים והן בהחלטותיו הצפויות של כל פרט בהינתן פרמטרים אלו. על כן כל אחת משיטות האמידה שתוארו בסעיף הקודם צריכות לספק הסתברויות מותנות לסלקציה (סינון) של הפרטים לתוכנית, כמו גם הסתברויות מותנות לסלקציה שלהם לתעסוקה לאחר ההשתתפות בתוכנית. הסתברויות אלו הן שקובעות במישרין את התועלת האישית והחברתית הצפויה מההשתתפות בתוכנית ומעצם קיומה.

ניתוח עלות-תועלת על פי מסגרת זו, תחת הנחות מפשטות, מראה כי **מערכת הכשרות מקצועיות הנותנת תשואה של 6%** לשנת לימוד היא בעלת תשואה חברתית חיובית.

² אם מניחים שלמשק יש פונקציית ייצור שבה עבודה והון הם גורמי ייצור משלימים אז התשואה החברתית היא גדולה יותר, שכן במצב זה העלייה בהון אנושי ובתעסוקה כתוצאה מהתוכנית מביאים גם לעלייה בהשקעה בהון.

5. מודל לביצוע מחקרי אפקטיביות בישראל ללא RCT תוך שימוש

בנתונים מנהליים

כפי שתואר בפרק 3, שיטות שונות להערכת אפקטיביות דורשות סוגים שונים של נתונים ותנאים. בפועל, עבור מרבית תוכניות התעסוקה וההכשרה בישראל, ניסוי אקראי מבוקר (RCT) אינו ישים מהסיבות שפורטו בסעיף 3.1 - ההשתתפות וולונטרית, מספר המשתתפים קטן, או שקיימים שיקולים אתיים. פרק זה מציע מודל מעשי לביצוע מחקרי אפקטיביות בישראל ללא RCT, תוך שימוש בנתונים המנהליים של הלשכה המרכזית לסטטיסטיקה. מוצגים שני מסלולים: מודל עם מיזוג נתונים (המועדף מבחינה מתודולוגית) ומודל ללא מיזוג (ישים גם כאשר מיזוג אינו אפשרי). שני המודלים מתבססים על גישת selection on observables (סעיף 3.3.1) לבניית קבוצת השוואה, ובמידת האפשר משלבים אמידת הפרש הפרשים (סעיף 3.4). דוגמה מפורטת ליישום המודל ללא מיזוג מוצגת בפרק 6.

שני המודלים חולקים שלבי מחקר משותפים, והם נבדלים בעיקר באופן בניית קבוצת ההשוואה וביכולות הניתוח שכל אחד מהם מאפשר. שלבי המחקר המשותפים לשני המודלים:

1. איסוף נתונים מלא לגבי משתתפי התוכנית, הכולל נתוני רקע ומשתני תוצאה בטרם ההתערבות (ראו נספח א לרשימת משתנים כדוגמה).
2. איסוף נתוני תוצאה של משתתפי התוכנית לאחר ההתערבות (שנה, שנתיים או שלוש מסיום ההתערבות, בהתאם למטרות המחקר).
3. איתור המשתתפים המרכזיים שמנבאים השתתפות בתוכנית (selection on observables).
4. בניית קבוצת השוואה מהנתונים המנהליים, כמפורט בהמשך עבור כל אחד משני המודלים.
5. אמידת השפעת התוכנית כהפרש בין ממוצע משתנה התוצאה בקבוצת הטיפול לממוצע משתנה התוצאה בקבוצת ההשוואה.

מודל עם מיזוג נתונים מנהליים

מודל זה מתבסס על האפשרות לבצע זיווג (מיזוג) בין נתוני משתתפי התוכנית לנתונים המנהליים של הלשכה המרכזית לסטטיסטיקה ברמת הפרט. זיווג זה מאפשר שני יתרונות משמעותיים שאינם זמינים במודל ללא מיזוג: ראשית, מעקב ארוך טווח אחר משתני התוצאה של קבוצת הטיפול באמצעות הנתונים המנהליים (השכלה, תעסוקה, שכר), ללא תלות בסקרי מעקב שעלותם גבוהה ושיעור ההיענות בהם עלול להיות נמוך. שנית, אפשרות לאמוד מודל PSM (סעיף 3.3.2) על הקובץ המאוחד של קבוצת הביקורת וקבוצת הטיפול, מה שמאפשר שימוש במספר רב יותר של משתני בקרה ובניית קבוצת השוואה איכותית יותר.

בשלב בניית קבוצת ההשוואה (שלב 4), החוקר משתמש באלגוריתם התאמה (matching) על הקובץ המאוחד כדי למצוא לכל משתתף בתוכנית "תאום מלאכותי" בעל מאפיינים דומים ככל האפשר. כאשר קיימים נתוני תוצאה גם לפני ההתערבות, ניתן לשלב אמידת הפרש הפרשים (סעיף 3.4) ובכך לנטרל הבדלים קבועים בין הקבוצות שאינם נצפים.

מודל זה הוא המועדף מבחינה מתודולוגית, אולם יישומו מותנה בקבלת אישור למיזוג נתונים מהלמ"ס, תהליך שהוא כיום ארוך ומורכב. מודל זה ישים בעיקר עבור **תוכניות ממשלתיות** שבהן הגוף המפעיל יכול לבצע את תהליך המיזוג בשיתוף הלמ"ס, ועבור תוכניות גדולות דיין שמצדיקות את ההשקעה בתהליך.

מודל ללא מיזוג נתונים מנהליים

מודל זה מתבסס על זמינות וחוקיות השימוש במשתנים המנהליים של הלמ"ס ללא מיזוג של הנתונים ברמת הפרט. במודל זה, בניית קבוצת ההשוואה (שלב 4) נעשית באופן שונה: לכל פרט i בקבוצת הטיפול, מאתרים בנתונים המנהליים את כל הפרטים שזהים לו בתכונות הנבחרות (בכפוף למספר מינימלי של תצפיות מותאמות בהתאם לדרישות הלמ"ס לשמירה על אנונימיות). קבוצת הפרטים הזהים מוגדרת כ-Ai, ומחושב עבורם ממוצע משתנה התוצאה. השפעת התוכנית נאמדת כהפרש בין ממוצעי משתני התוצאה של קבוצת הטיפול לאלו של קבוצת ההשוואה המותאמות.

היתרון המרכזי של מודל זה הוא מעשי: ניתן לבצע אותו ללא תהליך מיזוג נתונים, ולכן הוא ישים גם עבור **תוכניות של המגזר השלישי** ועבור תוכניות קטנות יחסית שבהן RCT אינו רלוונטי. מודל זה יושם בפועל במחקר האפקטיביות של תוכנית קימם המתואר בפרק 6.

מגבלות המודל ללא מיזוג ודרכי התמודדות

למודל ללא מיזוג מספר מגבלות שיש להכיר:

- מגבלת מספר משתני הבקרה: ניתן להשתמש רק במספר מוגבל של משתני רקע לבניית קבוצת ההשוואה. דרישות הלמ"ס לשמירה על אנונימיות (מספר מינימלי של תצפיות בכל תא) מגבילות את מספר המשתנים שניתן לשלב, שכן כל משתנה נוסף מקטין את מספר התצפיות בכל שילוב. כתוצאה מכך, בנוסף להנחת ה-selection on observables, לא ניתן לשלוט בכל המשתנים הנצפים שהיו יכולים לשפר את ההתאמה. מגבלה זו מדגימה את הצורך בהנגשת נתונים ברזולוציה גבוהה יותר, כפי שמפורט בהמלצות (פרק 7).
 - היעדר אפשרות ל-PSM: ללא מיזוג, לא ניתן לאמוד מודל Propensity Score Matching (סעיף 3.3.2) על הקובץ המשותף, שכן אין גישה לנתונים ברמת הפרט של שתי הקבוצות יחד. לכן ההתאמה מוגבלת לשיטת matching ישירה על משתנים נצפים.
 - מגבלת מעקב ארוך טווח: ללא מיזוג, נתוני התוצאה של קבוצת הטיפול לאחר ההתערבות צריכים להיאסף באופן עצמאי (באמצעות סקרי מעקב או דיווח של מפעילי התוכנית), ולא ניתן להשתמש בנתונים המנהליים למעקב אוטומטי אחר תעסוקה ושכר.
- חרף מגבלות אלו, המודל ללא מיזוג מהווה כלי מעשי בעל ערך, בפרט כאשר הוא משלב מספר אסטרטגיות להתמודדות עם הטיית הבחירה, כגון בניית מספר קבוצות השוואה בעלות רמות סלקטיביות שונות (כפי שיודגם בפרק 6). זמינות של נתונים מנהליים מפורטים היא תנאי הכרחי לשיפור איכות המחקר בשני המודלים, והמלצות בנושא זה מפורטות בפרק 7.

6. דוגמה לביצוע מחקר אפקטיביות בישראל ללא RCT וללא מיזוג נתונים

מנהליים: תוכנית "קימם"

פרק זה מציג דוגמה מפורטת ליישום המודל שהוצע בפרק 5 לביצוע מחקר אפקטיביות בישראל ללא RCT וללא מיזוג נתונים מנהליים. הדוגמה ממחישה כיצד ניתן ליישם את שיטת ההתאמה על משתנים נצפים (selection on observables, סעיף 3.3.1) תוך שימוש בנתונים מנהליים של הלמ"ס ללא מיזוג נתונים, ומדגימה הן את הפוטנציאל של גישה זו והן את המגבלות המעשיות שלה.

תוכנית "קימם" הינה שנת מעבר לצעירים וצעירות בעלי פוטנציאל אקדמי מהחברה הערבית, כלומר מסיימי י"ב בעלי תעודת בגרות המאפשרת לימודים אקדמיים. חזון התוכנית הוא לסייע לצעירים וצעירות אלו להשתלב בלימודים אקדמיים בישראל בתחומים המתאפיינים בביקוש ובתשואה גבוהים ולסיים בהצלחה את הלימודים, המהווים כלי מרכזי לתעסוקה איכותית ומוביליות חברתית, באמצעות תוכניות שנת מעבר איכותיות המקנות כלים וכישורים ומבצעות אבחון והכוון לימודי ותעסוקתי. ההתערבות כוללת שלושה שלבים: שלב א' במהלך כיתה י"ב המורכב מ-6-8 מפגשים, שעיקרם חשיפה ראשונית לאקדמיה ולתחומי לימוד; שלב ב' במהלך השנה שלאחר י"ב שהוא תוכנית מלאה של 5 ימים בשבוע הכוללת הכוון לימודי ותעסוקתי אישי וקבוצתי, הכנה למבחני קבלה, לימוד עברית, חשיפה לאקדמיה ולתחומי לימוד, קורסים אקדמיים וחזיון כישורים רגשיים-חברתיים; ושלב ג' במהלך השנים שלאחר סיום התוכנית הכוללת תמיכה והתמודדות עם צרכים שעולים מהשטח, חיבור למענים ושירותים תומכים עבור משתתפים שהתחילו לימודים אקדמיים והמשך ליווי לקראת תחילת לימודים למשתתפים שטרם התחילו.

הגדרת משתני התוצאה

מחקר האפקטיביות של קימם תוכנן כמחקר דינמי הבוחן את ההשפעה בארבע תחנות לאורך זמן, משנה אחת לאחר סיום התוכנית ועד להשתלבות בשוק העבודה. עיצוב רב-שלבי זה מאפשר לבחון את התפתחות ההשפעה, לזהות שלבים שבהם נדרש חיזוק ולספק תוצאות ביניים לקובעי מדיניות עוד לפני שניתן לאמוד את ההשפעה הסופית על תעסוקה ושכר.

1. שלב 1 - שנה מסיום התוכנית: שיעור הפרטים שהתחילו שנה א' במוסד אקדמי בישראל, התפלגות לפי סוג מוסד (אוניברסיטה, מכללה, מכללה לחינוך) והתפלגות תחום הלימוד לפי ארבע קטגוריות (מדעי החברה, חינוך, מקצועות טכנולוגיים, מקצועות פרא-רפואיים, לפירוט ראו נספח ב).
2. שלב 2 - שנתיים מסיום התוכנית: משתני התוצאה של שלב 1, בתוספת שיעור הנשירה במעבר משנה א' לשנה ב'.
3. שלב 3 - סיום לימודים בזמן תקני: שיעור הפרטים שסיימו לימודים, התפלגות בוגרים לפי מוסד ותחום, ושיעור נשירה.
4. שלב 4 - השתלבות בשוק העבודה: אמידת ההשפעה על תעסוקה ושכר ועדכון ניתוח עלות-תועלת.

מתודולוגיה: יישום המודל ללא מיזוג

מכיוון שניסויי אקראי מבוקר וניסויי טבעי אינם ישימים במקרה של תוכנית קימם (ההשתתפות וולונטרית ולא מוגדר סף זכאות חיצוני), שיטת המחקר התבססה על גישת selection on observables (סעיף 3.3.1), בהתאם למודל ללא מיזוג שהוצע בפרק 5. קבוצת ההשוואה נבנתה מתוך הנתונים המנהליים של הלמ"ס, מפרטים ערבים בשנתונים 1997-1999 עד גיל 21, תוך התאמה ("תאומים מלאכותיים") על פי שבעה מאפייני רקע: מגדר, גיל, מקום מגורים (צפון/מרכז), השכלת ההורים, הכנסת משק הבית, מספר אחים, והרכב תעודת הבגרות (8 קטגוריות של שילובי יחידות לימוד במתמטיקה, אנגלית ועברית, לפירוט ראו נספח ב). מגבלת הלמ"ס על מספר מינימלי של תצפיות בכל תא (לצורך שמירה על אנונימיות) הגבילה את מספר משתני הבקרה שניתן להשתמש בהם. לפיכך, למרות שנתוני משתתפי התוכנית כללו משתנים נוספים (כגון ציוני בגרות, ציון פסיכומטרי), לא ניתן היה לשלוט בהם. מגבלה זו ממחישה את אחד האתגרים המרכזיים של המודל ללא מיזוג שתואר בפרק 5.

כדי להתמודד עם הטיית בחירה עצמית ולתחום את טווח ההשפעה הנאמדת, נבנו שתי קבוצות השוואה: **קבוצת השוואה בסיסית** כללה פרטים הזהים בשבעת המאפיינים ללא דרישה נוספת. קבוצה זו משמשת חסם תחתון לאפקטיביות, שכן היא כוללת גם פרטים שלא ביטאו כוונה ללמוד (לא ניגשו לפסיכומטרי או יע"ל) ולכן צפויים להציג שיעורי למידה נמוכים יותר, מה שמגדיל את הפער הנאמד מול קבוצת הטיפול. **קבוצת השוואה סלקטיבית** הוסיפה תנאי שהפרטים ניגשו לפסיכומטרי ו/או יע"ל, כלומר הביעו מוטיבציה ללמוד במוסד אקדמי ישראלי. קבוצה זו משמשת חסם עליון לאפקטיביות, שכן היא מורכבת מפרטים בעלי מוטיבציה דומה יותר למשתתפי התוכנית, ולכן הפער הנאמד קטן יותר. חשוב לציין שמעל 60% ממשתתפי קימם לא ניגשו לפסיכומטרי או יע"ל לפני כניסתם לתוכנית, ועידוד הגשה לבחינות אלו היה בעצמו אחד מיעדי ההתערבות. לפיכך, ההשוואה מול הקבוצה הסלקטיבית, המורכבת מנבחנים בלבד, מחמירה את תנאי ההשוואה מעבר לנדרש ומספקת אומד שמרני להשפעת התוכנית.

תובנות מתודולוגיות מהדוגמה

הניסיון ביישום מחקר האפקטיביות לתוכנית קימם מדגים מספר סוגיות מתודולוגיות שרלוונטיות לכל מחקר אפקטיביות המבוסס על נתונים מנהליים בישראל. ראשית, הפער בשנתונים בין קבוצת הטיפול (שנתון 2000-2002) לקבוצת ההשוואה (שנתון 1997-1999) נובע ממגבלת זמינות נתוני הלמ"ס ומהווה איום פוטנציאלי על תקפות ההשוואה, שכן שינויים מערכתיים (כגון מגמות בהשכלה או השפעת הקורונה והמלחמה) עשויים להשפיע על התוצאות. מגבלה זו מדגישה את הצורך בשיפור זמינות הנתונים המנהליים כפי שמפורט בהמלצות (פרק 7). שנית, עיצוב המחקר כרצף של ארבע תחנות מדידה ממחיש את החשיבות של בחינת ההשפעה הדינמית (סעיף 3.6). ממצאי הביניים של התחנה הראשונה מצביעים על תרומה חיובית, אולם השאלות המרכזיות ביותר מבחינת עלות-תועלת חברתית (האם ההשתלבות באקדמיה מתורגמת לסיום תואר ולתעסוקה איכותית?) יענו רק בתחנות הבאות.

נקודה זו מעלה אתגר מהותי של מחקרי אפקטיביות בתחום ההכשרה וההשכלה: פרק הזמן הנדרש עד לקבלת תוצאות מלאות הוא ארוך. במקרה של תוכנית קימס, תוצאות מלאות על סיום תואר יתקבלו רק כארבע שנים לאחר סיום התוכנית, ותוצאות על השתלבות בשוק העבודה - מאוחר עוד יותר. פער זמנים זה בין ההשקעה בתוכנית לבין היכולת לאמוד את תוצאותיה מלווה לעיתים קרובות בתסכול מצד מקבלי החלטות, הנדרשים לקבל החלטות תקציביות בטרם התבררה התמונה המלאה. זהו חיסרון מובנה של מחקרי אפקטיביות שאין דרך להתגבר עליו, שכן אין תחליף למדידה של תוצאות בפועל לאורך זמן. עם זאת, בדיוק מסיבה זו חיוני לעצב מחקרי אפקטיביות עם תחנות ביניים, כפי שנעשה במחקר קימס. תוצאות ביניים מאפשרות לזהות מגמות בשלבים מוקדמים, לבצע דיוקים והתאמות בתוכנית עוד במהלך פעילותה, ולספק לקובעי מדיניות בסיס ראייתי חלקי אך בעל ערך לקבלת החלטות - גם כאשר התמונה המלאה טרם התבררה.

7. המלצות מדיניות

על בסיס הסקירה המתודולוגית, הניסיון המחקרי שנצבר בישראל והמגבלות שזוהו, להלן המלצות ליצירת תשתית מוסדית ומחקרית שתאפשר הערכת אפקטיביות שיטתית של תוכניות בשוק העבודה.

הקמת גוף מקצועי להערכת אפקטיביות של תוכניות בשוק העבודה

מומלץ להקים גוף מקצועי עצמאי שתפקידו לרכז, לתאם ולפקח על הערכות אפקטיביות של תוכניות התערבות בשוק העבודה. גוף זה יפעל כרגולטור מקצועי של מחקרי אפקטיביות ויהיה אחראי על מספר תפקידים מרכזיים: קביעת סטנדרטים מתודולוגיים למחקרי אפקטיביות, בחינה ואישור של תוכניות מחקר מלוות, ריכוז וסינתזה של ממצאים ממחקרי הערכה שונים, וייעוץ לקובעי מדיניות על בסיס הראיות שנצברו.

מודל אפשרי הוא רשת מרכזי "מה עובד" (What Works Centres) שהוקמה בבריטניה, הכוללת גופים כגון *What Works Centre for Local Economic Growth* המתמחה בהערכת תוכניות לצמיחה כלכלית מקומית ותעסוקה. מודל נוסף הוא *J-PAL* (*Abdul Latif Jameel Poverty Action Lab*) שמקורו ב-MIT ופועל כיום ברחבי העולם, ומתמחה בעריכת ניסויים אקראיים מבוקרים ובתרגום ממצאיהם למדיניות. גוף ישראלי מקביל יוכל לשלב את היתרונות של שני המודלים: מחד, קביעת סטנדרטים ופרסום סינתזות מחקריות כמו ברשת הבריטית; ומאידך, ליווי ישיר של מחקרי אפקטיביות ובניית שיתופי פעולה עם גופי ממשלה, כמו ב-J-PAL.

הגוף יוכל לפעול במסגרת אקדמית קיימת (כגון מכון מחקר באוניברסיטה), כגוף ממשלתי עצמאי, או כשילוב ביניהם. עצמאותו המקצועית חיונית להבטחת מהימנות הממצאים ומניעת מצבים שבהם הגוף המפעיל את התוכנית הוא גם המעריך שלה.

קביעת סטנדרט למחקרי אפקטיביות כתנאי למימון ממשלתי

מומלץ לקבוע כי כל תוכנית תעסוקה או הכשרה המבקשת מימון ממשלתי, או הממומנת כבר על ידי המדינה, תידרש לכלול מרכיב של הערכת אפקטיביות העומד בסטנדרטים מתודולוגיים מוגדרים. דרישה זו תחול הן על תוכניות חדשות המבקשות מימון לראשונה והן על תוכניות קיימות המבקשות להמשיך ולקבל מימון ציבורי.

הסטנדרט המוצע יכלול את הרכיבים הבאים: הגדרה ברורה של משתני התוצאה שייבדקו ולוחות הזמנים למדידתם; בחירת שיטת אמידה מתאימה מתוך השיטות שתוארו בנייר זה, תוך נימוק הבחירה; תכנון מוקדם של בניית קבוצת השוואה או ביקורת, כולל הגדרת המאפיינים להתאמה; איסוף נתוני רקע מלאים על המשתתפים טרם תחילת ההתערבות (ראו נספח א); ותכנית מעקב אחר משתני תוצאה לאורך זמן, הכוללת לפחות שתי נקודות מדידה לאחר סיום ההתערבות.

עבור **תוכניות חדשות** - יש לשקול שילוב ניסוי אקראי מבוקר כבר בשלב הפיילוט, ככל שהדבר אפשרי ואתי. תוכניות שבהן RCT אינו ישים (למשל תוכניות וולונטריות או תוכניות עם מספר משתתפים קטן) ידרשו להציג תכנית מחקר חלופית המבוססת על אחת השיטות המפורטות בנייר זה. עבור **תוכניות קיימות** - יש לבצע הערכת אפקטיביות רטרוספקטיבית תוך שימוש בנתונים מנהליים, לפי המודל המוצע בפרק 5. המשך מימון ממשלתי של תוכניות שאינן מציגות עדויות לאפקטיביות צריך להיבחן מחדש, תוך הקצאת המשאבים לתוכניות שהוכחו כאפקטיביות או לפיתוח חלופות חדשות.

הנגשת נתונים מנהליים של הלמ"ס לצורך מחקרי אפקטיביות

כפי שעולה מהמודל המוצע (פרק 5) ומהניסיון עם מחקר קימם (פרק 6), אחד החסמים המרכזיים לביצוע מחקרי אפקטיביות בישראל הוא הנגישות המוגבלת לנתונים המנהליים של הלשכה המרכזית לסטטיסטיקה. בפרט, שימוש בשיטות כמו Propensity Score Matching דורש גישה לנתונים ברמת הפרט על מאפייני רקע מפורטים, הן עבור משתתפי התוכנית והן עבור האוכלוסייה הכללית - וזאת על מנת לבנות קבוצת ביקורת איכותית המבוססת על הסתברויות להשתתפות. המלצותינו בתחום זה הן:

- אפשר מיזוג נתונים למטרות מחקר אפקטיביות. יש ליצור מסלול מוסדר ומהיר יחסית שיאפשר לחוקרים מורשים לבצע מיזוג (זיווג) בין נתוני משתתפי תוכניות לנתונים המנהליים של הלמ"ס, תוך שמירה מלאה על סודיות סטטיסטית. מיזוג כזה חיוני לביצוע PSM איכותי ולמעקב ארוך טווח אחר משתני תוצאה. כיום, תהליך קבלת אישור למיזוג נתונים הוא ארוך ומורכב, מה שמרתיע חוקרים ומעכב מחקרים.
 - הגדלת הרזולוציה של הנתונים הנגישים ללא מיזוג. כפי שתואר בפרק 5, מחקרי אפקטיביות ללא מיזוג מוגבלים במספר משתני הרקע שניתן לשלוט בהם, בשל דרישות הלמ"ס לשמירה על אנונימיות (מספר מינימלי של תצפיות בכל תא). מומלץ לבחון מנגנונים שיאפשרו שימוש במשתנים נוספים תוך עמידה בדרישות הסודיות - למשל, הרחבת השימוש בחדרי מחקר מאובטחים של הלמ"ס שבהם ניתן לעבוד עם נתונים ברזולוציה גבוהה יותר מבלי להוציאם מהמערכת.
 - הרחבת מנגנוני העברת מידע מנהלי ללמ"ס. בהמשך להמלצות וועדת התעסוקה 2030, יש לייעל ולהרחיב את מנגנוני העברת המידע המנהלי מהמוסד לביטוח לאומי, רשות המיסים, משרד החינוך, המל"ג, מה"ט, האגף להכשרה מקצועית וגופים נוספים ללשכה המרכזית לסטטיסטיקה. בפרט, יש לוודא שנתוני ההשתתפות בתוכניות תעסוקה והכשרה מועברים באופן שוטף ומובנה, כך שניתן יהיה לשלבם עם נתוני תעסוקה, שכר והשכלה לצורך מחקרי אפקטיביות.
 - הוספת שאלות על השתתפות בתוכניות לסקרי הלמ"ס. מומלץ להוסיף לסקר כוח אדם ולסקרים אחרים של הלמ"ס שאלות על השתתפות בהכשרות מקצועיות ובתוכניות תעסוקה. מידע זה, בשילוב עם נתוני תעסוקה ושכר הקיימים כבר בסקרים, יאפשר ביצוע מחקרי אפקטיביות ברמת המדגם הארצי ויסייע בזיהוי דפוסים חוצי-תוכניות.
- יישום המלצות אלו יצמצם משמעותית את הפער שקיים כיום בין הפוטנציאל המחקרי הגלום בנתונים המנהליים של ישראל לבין השימוש בפועל בנתונים אלו לצורך הערכת אפקטיביות ושיפור מדיניות התעסוקה.

מקורות

1. בראון, ל. (2015), "מדיניות פעילה בשוק העבודה: השפעת התכנית "מעגלי תעסוקה" על קבלת גמלאות", בנק ישראל.
2. משרד העבודה, הרווחה והשירותים החברתיים (2020), "הוועדה לקידום תחום התעסוקה לקראת שנת 2030: דו"ח מסכם".
3. ליפשיץ, א. ולרום, ט. (יפורסם בקרוב), "מחקר אפקטיביות לתוכנית "קימם" - מחזור ראשון: תוצאות אפקטיביות שנתיים מסיום המחזור הראשון" מכון אהרן למדיניות כלכלית.
4. Angrist, J. D. (1990), "Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records", *The American Economic Review*, 313-336.
5. Angrist, J. D. and Lavy, V. (1999), "Using Maimonides' rule to estimate the effect of class size on scholastic achievement", *The Quarterly Journal of Economics* 114(2), 533 - 575.
6. Battistin, E. and Meroni, E. C. (2016), "Should we increase instruction time in low achieving schools? Evidence from Southern Italy", *Economics of Education Review* 55, 39- 56.
7. Brender, A. and Strawczynski, M. (2019), "The effect of the Israeli earned income tax credit on labor supply: A differential analysis", *Israel Economic Review* 17(1), 1 - 46.
8. Callaway, B. and Sant'Anna, P. H. (2021), "Difference-in-differences with multiple time periods", *Journal of Econometrics* 225(2), 200- 230.
9. Card, D. and Krueger, A. B. (1995), "Time-series minimum-wage studies: a meta-analysis", *The American Economic Review* 85(2), 238-243
10. de Chaisemartin, C. and D'Haultfoeulle, X. (2020), "Two-way fixed effects estimators with heterogeneous treatment effects", *American Economic Review* 110(9), 2964-2996.
11. Duflo, E. (2020), "Field experiments and the practice of policy", *American Economic Review* 110(3), 545-571.
12. Gould, E. D., Lavy, V. and Paserman, M. D. (2004), "Immigrating to opportunity: Estimating the effect of school quality using a natural experiment on Ethiopians in Israel", *The Quarterly Journal of Economics* 119(2), 489-526.
13. Lavy, V. and Schlosser, A. (2005), "Targeted remedial education for underperforming teenagers: Costs and benefits", *Journal of the European Economic Association* 3(2 - 3), 431-443.
14. Myers-JDC-Brookdale Institute (2010), "Evaluation of the "Orot LeTaasuka" program using propensity score matching".
15. Rubin, D. B. (1973), "Matching to remove bias in observational studies", *Biometrics* 29(1), 159-183.
16. Schlösser, A. and Shanan, Y. (2016), "Evaluation of the "Employment Circles" program using randomized controlled trial methodology", Tel Aviv University.

17. Sun, L. and Abraham, S. (2021), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects", *Journal of Econometrics* 225(2), 175-199.
18. Todd, P. E. and Wolpin, K. I. (2023), "The best of both worlds: combining randomized controlled trials with structural modeling", *Journal of Economic Literature* 61(1), 41-85.
19. Zusman, T. and Maagan, D. (2018), "Impact of “Atuda Mada’it-Technogit” and “Latet Hamesh” programs on math achievements: A difference-in-differences approach", Israel Ministry of Education.

נספח א: הנתונים הדרושים

לביצוע מחקר אפקטיביות נדרשים נתוני מעקב לאורך זמן על פרטים שהשתתפו ושלא השתתפו בתוכנית, הכולל משתני רקע ומשתני תוצאה בהתאם לסוג התוכנית. נתונים אלו צריכים לכלול מאפיינים דמוגרפיים ואת מירב מאפייני הרקע (כולל הישגים לימודיים וניסיון תעסוקתי טרם ההשתתפות בתוכנית), מאפייני ההשכלה בכל תקופה (כולל סוג מוסד לימודים ותחום לימודים) ומאפייני התעסוקה בכל תקופה (כולל היקף תעסוקה, ענף, משלח יד, שכר חודשי ו/או שכר לשעה). **זמינות נתונים אלו היא קריטית הן לשם בחינת ההשפעה הדינאמית של תוכניות התערבות, והן להתמודדות עם האתגר של מידת הדמיון בין קבוצת הטיפול לקבוצת הביקורת.**

משתני רקע אישי

1. מין
 2. גיל
 3. קבוצת אוכלוסייה (יהודי לא-חרדי, ערבי, חרדי)
 4. מיקום גאוגרפי: מרכז / פריפריה, עיר / כפר (יסודי תיכון)
 5. השכלת ההורים (שנות לימוד, האם אקדמאים)
 6. רמת הכנסה של משק הבית (נמוכה מאוד, נמוכה, ממוצעת, גבוהה, גבוהה מאוד)
 7. תעודת בגרות:
 - זכאות לבגרות
 - מספר יחידות במתמטיקה
 - מספר יחידות באנגלית
- ניתן להרחיב את השאלות לגבי מקצועות נוספות בהתאם לתוכנית ההתערבות:
- מספר יחידות בפיזיקה
 - מספר יחידות במדעי המחשב
 - מספר יחידות בביולוגיה
 - מספר יחידות בכימיה
 - מספר יחידות בעברית
8. ציון פסיכומטרי (מקסילמי עבור אלה שנבחנו מספר פעמים)
9. השכלה על תיכונית (אקדמית או טכנולוגית):
 - סוג מוסד לימודים (אוניברסיטה, מכללה, מכללה טכנולוגית)
 - תחום לימודים / מגמה
10. תעסוקה לפני השתתפות בתוכנית:
 - סטטוס תעסוקתי (עובד במשרה מלאה, עובד במשרה חלקית, מובטל / מחפש עבודה, מחוץ לכוח העבודה)
 - משלח יד במקום העבודה האחרון (תפקיד)
 - ענף כלכלי של מקום העבודה האחרון (תחום פעילות)
 - שם מקום העבודה האחרון
 - שכר במקום העבודה האחרון
 - שנות ניסיון

משתני תוצאה של השכלה

1. האם לומד?
2. לקראת איזה תואר / תעודה
3. שנת לימודים
4. סוג מוסד לימודים (אוניברסיטה, מכללה, מכללה טכנולוגית)
5. שם מוסד הלימודים
6. תחום לימודים / מגמה

משתני תוצאה של תעסוקה ושכר

1. סטטוס תעסוקתי (עובד במשרה מלאה, עובד במשרה חלקית, מובטל / מחפש עבודה, מחוץ לכוח העבודה)
2. משלח יד
3. ענף כלכלי
4. שעות עבודה
5. שכר

נספח ב: הגדרת המשתנים במחקר "קימס"

מצרפי בגריות לבניית קבוצת ההשוואה

- 5 מתמטיקה, 5 אנגלית, 5 עברית
- 5 מתמטיקה, 5 אנגלית, 3 עברית
- 5 מתמטיקה, 4 אנגלית, 5 עברית
- 5 מתמטיקה, 4 אנגלית, 3 עברית
- 4 מתמטיקה, 5 אנגלית, 5 עברית
- 4 מתמטיקה, 5 אנגלית, 3 עברית
- 4 מתמטיקה, 4 אנגלית, 5 עברית
- 4 מתמטיקה, 4 אנגלית, 3 עברית
- 3 מתמטיקה, 5 אנגלית
- 3 מתמטיקה - כל היתר

קיבוץ תחומי לימוד

- מדעי החברה והניהול
 - כלכלה, ניהול, מדעי הייטק
 - חשב שכר, חשבונאות סוג א+ב ויועץ מס
 - כלכלה ומתמטיקה
 - משפטים ומנהל עסקים
 - עסקים וחשבונאות
 - עבודה סוציאלית
- מדעי הרוח וחינוך
 - עיצוב אופנה
 - רוח מדעית
 - מוזיקה
- טכנולוגיה והנדסה
 - ביוכימיה ומדעי המזון
 - הנדסת בנין
 - הנדסה כימית
 - הנדסת נתונים ומידע
 - הנדסת תוכנה
 - הנדסת תעשייה וניהול
 - מדעי המחשב
 - מדעי המחשב ומדעי המוח
 - ארכיטקטורה
 - מתמטיקה וניהול
 - מתמטיקה וביולוגיה
 - הנדסאי תעשייה וניהול
 - התקנות
- רפואה, מקצועות עזר רפואיים וביולוגיה
 - רפואה
 - אופטומטריה
 - ביולוגיה
 - רוקחות
 - הדמיית רנטגן
 - הפרעות בתקשורת
 - רפואת חירום
 - מעבדות רפואיות
 - סיעוד
 - רפואת שיניים