

השימוש במאגרי מידע מנהליים, דוגמת קמיליון, לצורך ביצוע מחקרים רפואיים, רווח מאד.

כאשר בודקים השערות על מאגרים מאין אלו, נוהגים לחשב ולצטט את רמת המובהקות של ההשערה (p-value), כמדד ליכולת להשליך מהמאגר המסויים, לאוכלוסיות אחרות. יש כשל גדול בשיטה סטטיסטית זו: רמת המובהקות נבנתה, אך ורק, למצב בו מעוניינים לבדוק השערה על אוכלוסייה ידועה (לדוגמא – כלל רשימת המצביעים בישראל). לצורך כך, מגרילים מדגם מקרי ומייצג מהאוכלוסייה, מה שמבטיח את האפשרות להשליך מסקנות מהמדגם לאוכלוסייה הכללית.

מאגר מידע מנהלי אינו מהווה מדגם, ובוודאי לא מדגם מייצג, מאוכלוסייה כלשהי; כך שלא ברור מה המשמעות של ערכי מובהקות, שמחושבים על נתונים ממאגרי מידע. חלופה אפשרית לרמת מובהקות, שמתאימה גם למאגרי מידע, היא Generalizability Index (מדד הכללה) (ראו בקישור הרצ"ב)

<https://www.ipr.northwestern.edu/workshops/past-workshops/cluster-randomized-trials-in-education/2016/password-protected/files/2015/Session%2010%20Tipton%202014.pdf>

הרעיון הכללי, הוא לחשב, באמצעות כלים סטטיסטיים, את ה"מרחק" (מבחינת נתוני רקע), בין המאגר עליו נערך המחקר, לבין מאגר מידע אחר (לדוגמא נתונים מבית חולים אחר). ככל שמרחק זה קטן יותר, עולה הסיכוי לכך שמסקנות ממאגר המחקר, יהיו נכונות גם למאגר האחר.

שימו לב, יש צורך לחשב מדד זה, עבור כל מאגר אליו מעוניינים להשוות. מכיוון שלכל מאגר, יכולים להיות מאפיינים שונים.

לאור משברים באמינות של מחקרים רפואיים, בעידן המידע, כדאי לשקול זניחת השימוש ברמת המובהקות, לטובת שיטות מתאימות יותר.