



**מרכז רן נאור לחקר הבטיחות בדרכים**  
The Ran Naor Road Safety Research center



**המכון לחקר התחבורה**  
הטכניון - מכון טכנולוגי לישראל  
Technion - Israel Institute of Technology  
Transportation Research Institute

## **פיתוח כלים סטטיסטיים להערכת מספרי הנפגעים קשה בתאונות הדרכים בישראל**

ד"ר ויקטוריה גיטלמן

ד"ר אטי דובא\*

\*המעבדה לסטטיסטיקה

**במימון קרן רן נאור לקידום מחקר בטיחות בדרכים**

**מיסודה של עמותת אור ירוק**



**קרן רן נאור לקידום מחקר בטיחות בדרכים**  
**RAN NAOR FOUNDATION**  
for the advancement of road safety research

יולי 2011, חיפה

דו"ח מחקר מס' S/25/2011

## תקציר

### 1. נושא המחקר

מטרת מחקר זה הייתה בפיתוח כלים סטטיסטיים להערכת מספר כולל (אמיתי) של נפגעים קשה בתאונות הדרכים בישראל. הנחת הבסיס בהערכה זו הינה שאת היקף התופעה - המספר הכולל של הנפגעים קשה בתאונות בשנה מסוימת - יש לאמוד על סמך מספר הנפגעים קשה שנרשמו, בתקופה זו, בבתי החולים. לכן, הכלים הסטטיסטיים שפותחו במחקר אמורים לאפשר חישוב מקדמי תיקון/ניפוח למספרי הנפגעים קשה המדווחים, בשנה מסוימת, על-ידי המשטרה.

קבצי הנתונים למחקר הוכנו ע"י הלשכה המרכזית לסטטיסטיקה (למ"ס) והם התבססו על תוצרי עבודה מקדימה של הלמ"ס - שילוב מידע מבתי החולים (רישום הטראומה הלאומי) עם קובץ נתוני תאונות הדרכים של המשטרה, בשנת 2008. פיתוח הכלים במחקר התבסס על שלשה סוגי נתונים שהם: נתוני הקובץ המשולב "משטרה וטראומה", נתוני קובץ "טראומה בלבד", ונתוני קובץ "משטרה בלבד" (כאשר שני הקבצים האחרונים כללו רשומות שלא נכנסו לקובץ המשולב).

נפגע קשה במחקר הנוכחי מוגדר לפי "הגדרה אדמיניסטרטיבית" שהיא: "אדם שאושפז בעקבות תאונה בבית חולים לתקופה של 24 שעות ויותר", ללא קשר לחומרת הפציעה לפי הסקאלה הרפואית. לכל הנפגעים מרישום הטראומה, ע"י הלמ"ס חושב מאפיין על סמך הגדרה זו, דהיינו הנפגע מסומן כ-"פצוע קשה" כאשר "תאריך שחרור פחות תאריך הגעה לבית חולים" גדול מ-1.

### 2. כלים סטטיסטיים שפותחו במחקר

להערכת מספר נפגעים קשה בתנאי הארץ פותחו שני כלים:

(א) מודל False-Positiven המחשב הסתברות לדיווח שגוי על נפגע קשה ע"י המשטרה והמאפשר ניפוי של מספר נפגעים קשה שזוהו ע"י המשטרה ולא אומתו ע"י רישום הטראומה (הנפגעים קל שבטעות זוהו כנפגעים קשה ע"י המשטרה);

(ב) מודל להערכת ההסתברות להיות מזהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא ברישום הטראומה. מודל זה מאפשר תיקון של תת-הדיווח בנתוני המשטרה - ניפוח מספר הנפגעים קשה שדווחו ע"י המשטרה בעזרת הנתונים מרישום הטראומה.

שני המודלים שנבנו במחקר הם מודלי רגרסיה לוגיסטית.

מודל א': להערכת ההסתברות לדיווח שגוי המודל נעזר בארבעה משתנים מסבירים, לפי הקטגוריות של סוג רכב, סוג תאונה, יום בשבוע ו-יחידת המשטרה. (לפירוט מודל א', ראה פרק 6 של הדו"ח).

מודל ב': לפיתוח מודל זה נדרש לאתר גורמים המשפיעים על ההסתברות של נפגע קשה הנמצא בקובץ טראומה, להיות מזהה כנפגע קשה ע"י המשטרה (לאחר ניכוי המקרים של ה- False Positive, בעזרת מודל א'). זאת הסתברות מותנית של תפיסה ע"י המשטרה בתנאי שהייתה תפיסה ע"י טראומה. הסתברות זו מכונה כ"הסתברות תפיסה מחדש" ע"י המשטרה ( recapture probability), ואת הגורמים המשפיעים עליה יש למצוא באמצעות מודל הקושר בין הגורמים להסתברות.

מכיוון שהמודל מתבסס על תצפיות מקבצי הטראומה, כאשר בסוף התהליך המודל צריך להיות מופעל על נתוני המשטרה, היה צורך באיתור מאפייני הנפגעים המשותפים לשני סוגי הקבצים. בשלב הראשון, נבחנו קשרים בין מאפייני הנפגעים קשה בטראומה ובמשטרה ונמצאו משתנים כאלה שעשויים להיות משמעותיים בקביעת ההסתברות לכידה ע"י המשטרה, כאשר הם ניתנים לשחזור מלא בנתוני הטראומה על סמך הנתונים בקובץ המשטרה. בשלב השני, משתנים מלאכותיים כאלה (הקבצות קטגוריות של מאפייני הנפגעים) שנמצאו בשלב הראשון שימשו לצורך בחירת המודל שעל-פיו נאמדו הסתברויות הלכידה ע"י המשטרה. לבחינת הקשרים בין מאפייני הנפגעים בטראומה ובמשטרה שימש הקובץ המשולב; לפיתוח מודל מסביר להסתברות להיות מדווח כנפגע קשה ע"י המשטרה שימשו נתונים משני קבצים: הקובץ המשולב וקובץ טראומה בלבד (סה"כ 5153 רשומות על נפגעים קשה לפי הטראומה).

במודל שנבנה במחקר, להערכת הסתברות הלכידה ע"י המשטרה משמשים מאפיינים אלה: מגדר הנפגע, קבוצת גיל הנפגע, יחידת המשטרה, סוג דרך, סוג תאונה, סוג נפגע, וקטגוריות נפגע  $I_{cat}$  המשקפות בעיקר את סוג הרכב בו נסע הנפגע. (לפירוט מודל ב', ראה פרק 6 של הדו"ח).

### 3. יצירת אמד למספר כולל של נפגעים קשה על סמך דיווח המשטרה

כדי לקבל אמד למספר הכולל של נפגעים קשה, בעזרת הכלים שפותחו במחקר, על סמך קובץ המשטרה יש לבצע שתי פעולות:

(א) לתקן את מקרי ה-False Positive - מקרים של דיווח שגוי ע"י המשטרה, בעזרת מודל א';

(ב) לנפח את מספר המקרים שנותר על סמך הסתברויות הלכידה ע"י המשטרה, בעזרת מודל ב'.

למעשה, לכל רשומה של נפגע קשה שדווח ע"י המשטרה, מחושב ערך  $p_{final}$  המהווה "מספר כולל" (או אמיתי) של נפגעים קשה עם מאפיינים אלה שאמור להיות בקובץ המשטרה. ערך ה- $p_{final}$  מוערך באופן הבא:

$$p_{final} = p_{TP} * (1/p_{catch})$$

$$p_{TP} = 1 - p_{FP}$$

כאשר

$p_{TP}$  - הסתברות לדיווח נכון ע"י המשטרה (TP - True Positive);

$p_{FP}$  - הסתברות לדיווח שגוי ע"י המשטרה (FP - False Positive), לפי מודל א';

$p_{catch}$  - הסתברות לכידה ע"י המשטרה, לפי מודל ב'.

הגדלים הללו ( $p_{final}$ ) מסוכמים על פני כל הנפגעים קשה שבקובץ המשטרה וכך מתקבל אמד למספר הכולל של נפגעים קשה, בשנה מסוימת.

למעשה, ערך ה- $p_{TP}$  מהווה מקדם תיקון, בעוד שערך ה- $(1/p_{catch})$  מהווה מקדם ניפוח - לכל רשומה של נפגעים קשה בקובץ המשטרה.

בשיטה זו בוצעה הערכה של המספר הכולל של נפגעים קשה בשנת 2008 והתקבלו תוצאות כלהלן:

מספר הנפגעים קשה שדווח ע"י המשטרה היה  $2201=N$ ;

בעקבות הפעלת מודל א' והסתברויות ה-TP, מספר הנפגעים קשה תוקן ל- $2111=N_{TP}$ ;

בעקבות הפעלת מודל ב' של הסתברות הלכידה ע"י המשטרה המספר הכולל של נפגעים קשה עלה ל-  $N_{final} = 6709$ .

נמצא שבממוצע, 96% מהנפגעים קשה של המשטרה היו מדווחים נכון, כאשר המספר האמיתי של נפגעים קשה בתאונות היה גבוה פי 3.05 לעומת דיווח המשטרה.

#### 4. סיכום

הכלים הסטטיסטיים שפותחו במחקר מאפשרים לבצע הערכות של המספר הכולל של נפגעים קשה בתאונות הדרכים בישראל על סמך נתוני המשטרה. אומדנים אלה ניתן לקבל עבור אוכלוסיות שונות של נפגעים קשה וכמו כן, לבצע הערכות דומות של המספר הכולל של נפגעים קשה עבור שנים אחרות. עם זאת, שימוש בכלים ובמקדמי תיקון/ניפוח שפותחו במחקר לתיקון מספרי הנפגעים קשה בשנים האחרות מתבסס על הנחת דמיון בין מצב הדיווח שהיה בשנת 2008 ובשנים האחרות.

שימוש באומדנים של המספר הכולל של נפגעים קשה בתאונות הדרכים בישראל עשוי לתרום להגברת הדיוקים בהערכות בטיחות שונות ובתכנון פעילויות התערבות. בין היתר, ניתן לבצע הערכות רטרואקטיביות של המספרים הכוללים של נפגעים קשה בתאונות ולבחון את מגמות השינוי לאורך זמן על סמך בסיס מתוקן זה.

## תוכן עינינים

6	1. מבוא.....
6	1.1. רקע למחקר ומטרתו.....
8	1.2. הנתונים ששימשו למחקר.....
9	1.3. תוכן הדו"ח.....
10	2. הצגת מתודולוגית ה-Capture-Recapture.....
14	3. חקר מקדים של הנתונים.....
14	3.1. נתוני קובץ A המשולב, כולל יו"ש.....
15	3.2. נתוני קובץ A המשולב, ללא יו"ש.....
17	3.3. הערכות ראשוניות של מספר הנפגעים קשה.....
23	4. מודל ה-False Positive - ניפוי נפגעים קל בקרב נפגעים קשה של המשטרה.....
23	4.1. בחינת הנתונים.....
24	4.2. פיתוח המודל.....
25	4.3. ממצאי המודל.....
	5. מודל לחישוב ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה - לניפוח נפגעים קשה של המשטרה.....
28	5.1. כללי.....
28	5.2. הנתונים ששימשו לבניית המודל.....
30	5.3. בחינת משתנים מסבירים.....
40	5.4. בנית מודל להסתברות לכידה ע"י המשטרה.....
44	6. סיכום: יצירת אמד למספר נפגעים קשה בתאונות הדרכים על סמך דיווח המשטרה.....
44	6.1. כלים סטטיסטיים שפותחו במחקר.....
48	6.2. יצירת אמד למספר כולל של נפגעים קשה על סמך דיווח המשטרה.....
51	מראי מקום.....
53	נספח א': ממצאים מסקר ספרות.....
60	נספח ב': מבנה קובץ A - נתונים משולבים.....
66	נספח ג': שיטות ה-CART וה-GBM.....
67	נספח ד': הערכת רווחי הסמך לאמד ההסתברות ל-FP.....
69	נספח ה': ריכוז משתנים לבניית המודל להסתברות לכידה ע"י המשטרה.....
73	נספח ו': כללים להשלמת ערכים חסרים במאפייני הנפגעים.....
74	נספח ז': שיטה ליצירת רווח סמך למספר הנפגעים הכולל.....

## 1. מבוא

### 1.1. רקע למחקר ומטרות

בישראל, בדומה למדינות רבות אחרות, הסטטיסטיקה הרשמית של תאונות הדרכים וחלק ניכר מניתוחים והערכות בתחום הבטיחות בדרכים מתבססים על נתוני התאונות שנאספים ע"י המשטרה. הדיווחים על הרוגים בתאונות הדרכים נחשבים למוגדרים היטב ומדויקים יחסית. לעומת זאת, כאשר מדובר במספרי נפגעים ברמות חומרה נמוכות יותר, קיימת תופעה של תת-דיווח (כגון: Derriks and Mak, 2007). תופעה זו והשלכותיה על ביצוע הערכות ובחירת פעילויות התערבות בתחום הבטיחות בדרכים נדונה בספרות המקצועית לאורך מספר עשורים - ראה סקר ספרות בנספח א'.

בישראל, המודעת לתופעת תת-הדיווח על מספרי הנפגעים בתאונות הדרכים עלתה בעשור האחרון, הודות לזמינות ואפשרויות השוואה בין מאגרי מידע שונים, בנוסף לקובצי המשטרה, וביניהם: רישום הטרומה הלאומי (ITR - Israeli National Trauma Registry); מאגר נתונים של חברות הביטוח. בהערכות שבוצעו בישראל (מתת, 2004) נתגלו אי-התאמות בין מספרים ומאפייני הנפגעים בתאונות המדווחים באמצעות קבצי "ת"ד" של המשטרה (הסטטיסטיקה הרשמית) לבין המספרים ופרטי הנפגעים שהתקבלו מאיחוד המקורות, כאשר בין הסוגיות בולטות היה הפרש ניכר בין מספר נפגעים קשה בתאונות המדווח על סמך קבצי המשטרה לבין מספר נפגעים קשה לפי רישום הטרומה הלאומי.

יתרה מזו, במחקר אשר השווה את המגמות לאורך זמן במספרי הנפגעים קשה לפי רישום הטרומה הלאומית לעומת דיווחי המשטרה (פלג, אהרונסון-דניאל, 2004) נמצאו הן הפרשים ניכרים בין המספרים המדווחים בשני המקורות והן מגמות הפוכות לאורך זמן: עליה במספר הנפגעים קשה לפי רישום הטרומה וירידה לפי דיווחי המשטרה. לדוגמא, בשנת 1998 המשטרה רשמה 3,721 פצועים קשה (=המאושפזים ליותר מ-24 שעות) שהיווה 39% מהנתון המקביל של משרד הבריאות (מאגר בתי החולים) והיה כ-6% מעל הנתון המקביל מרישום הטרומה. לעומת זאת, בשנת 2002, המשטרה רשמה 2,559 פצועים קשה והמספר היווה רק 29% מהנתון המקביל של משרד הבריאות ורק 64% מנתוני רישום הטרומה, כאשר הצגה שיטתית של הנתונים מרישום הטרומה מול נתוני המשטרה הדגישה פער הולך וגדל לאורך השנים.

במחקר המשך שנערך ע"י הלשכה המרכזית לסטטיסטיקה - למ"ס (CBS, 2005) ואשר השווה בין נתוני המשטרה וה-ITR התקבל אומדן כי המספר הכולל של הנפגעים קשה (=המאושפזים בעקבות התאונה ליותר מ-24 שעות) צריך להיות בין 3,700 ל-4,600, במקום 2,576 שדווח ע"י המשטרה.

בכל ההערכות הנ"ל ההשוואה בין הנפגעים קשה לפי המקורות השונים התבססה על "הגדרה אדמיניסטרטיבית" של נפגע קשה - אדם שאושפז בעקבות תאונה בבית חולים לתקופה של 24 שעות ויותר. הגדרת פצוע קשה לפי סקאלה רפואית (כגון: ISS - Injury Severity Score) שימשה לביחנות עומק של הנתונים אך לא להשוואה מספרי הנפגעים בין נתוני המשטרה ונתוני הטרומה.

יצוין כי הפער בין מספר הנפגעים המדווח ע"י המשטרה לבין המקורות הרפואיים הינו עובדה מוכרת בכל העולם. לדוגמא, בבדיקות מקורות מידע מקבילים שנערכו במדינות האיחוד האירופי נמצא

שבדרך כלל יותר נפגעי התאונות נרשמים בבתי החולים כאשר, בממוצע, כ-30% מהנפגעים קשה וכ-60% מהנפגעים קל אינם מופיעים בנתוני המשטרה (ETSC, 1999). תת-הדיווח בנתוני המשטרה ניכר יותר בקרב הנפגעים בתאונות רכב יחיד, הולכי רגל ורוכבי אופניים. באופן דומה, בהשוואת נתוני התאונות הרשמיים עם רשומות בתי החולים בנורבגיה נמצא שהסטיסטיקה הרשמית משקפת 46% מהולכי הרגל, 42% מרוכבי אופניים, 53% מהנוסעים ברכב וכשליש בלבד מרוכבי רכב דו-גלגלי שנפגעו בתאונות וטופלו בבתי החולים (Elvik and Vaa, 2004). (ראה דוגמאות נוספות בנספח א').

ע"פ המחקרים שנערכו במספר מדינות (כגון: בריטניה, הולנד, צרפת, אוסטרליה - ראה נספח א'), להערכת מספר כולל של הנפגעים קשה משמשים קבצי נתונים משולבים אשר נבנו בעזרת הצלבה של מאגר מידע משטרתי ומאגר מידע רפואי. ניפוח מספר הנפגעים קשה מבוצע על סמך מספרי המאושפזים בעקבות תאונות הדרכים שנמצאו במאגר הרפואי ולא נמצאו במאגר המשטרה (כאשר מספרים אלה מהווים תוספת למספר הנפגעים שנמצאו הן במאגר המשטרה והן במאגר הרפואי, או למספר הנפגעים שדווח ע"י המשטרה). זאת, מתוך הנחה כי מספר המאושפזים בעקבות התאונות שנמצא בבתי החולים משקף את ההיקף האמיתי של התופעה, ולכן, יש לתקן את מספרי הנפגעים קשה של המשטרה בהתאם.

בתהליך הערכה זו (קבלת המספר הכולל של הנפגעים קשה) קיימים שני שלבים: (1) ביצוע הצלבה בין רשימות הנפגעים ממאגרי מידע שונים, (2) פיתוח מקדמי תיקון/ניפוח למספרי הנפגעים. ליצירת הקובץ המשולב, ברוב המחקרים בחו"ל שימשו שיטות הצלבה מבוססות הסתברות אשר נעזרו במספר פרטים מזהים של הנפגעים כגון: שם, מגדר, גיל, זמן פגיעה ומיקום הפגיעה/בית חולים בו טופל הנפגע. בישראל, החל משנת 2008, הלשכה המרכזית לסטיסטיקה (למ"ס) מבצעת שילוב מידע מבתי החולים (באמצעות רישום הטראומה הלאומי) עם קובץ נתוני תאונות הדרכים של המשטרה. קישור נתוני רישום הטראומה לנתוני המשטרה נערך ע"פ מספרי תעודת זהות, שנת לידה של הנפגע, תאריך התאונה, שם משפחה ושם פרטי של הנפגע. כתוצאה מתהליך זה, כל שנה, מתקבלים רשומות הנפגעים בתאונות שנמצאו הן בקבצי המשטרה והן ברישום הטראומה, וגם רשומות הנפגעים שנמצאו בטראומה או במשטרה, בלבד. קיום מאגר משולב זה יצר בסיס לפיתוח כלים להערכת מספר כולל של הנפגעים קשה בתאונות הדרכים בישראל.

לפיתוח מקדמי תיקון/ניפוח למספרי הנפגעים שדווחו ע"י המשטרה על מנת לקבל אומדנים למספרם הכולל (האמיתי), ככלל, משמשת שיטת Capture-Recapture ("לכידה-לכידה חוזרת") - ראה נספח א'. יישום השיטה מותנה בקיום מספר תנאים כגון: הסתברות שווה ללכידה של כל מקרה; אי-תלות בין המאגרים השונים; נכונות הצלבה בין המאגרים השונים וכד'. לרוב, נתוני תאונות דרכים לא עומדים בהנחות בסיס אלה ולכן, מבוצעת חלוקה של מאגר הנפגעים לקבוצות יותר הומוגניות אשר יש סיכוי שהן כן מקיימות את התנאים הדרושים ליישום השיטה. מאידך, אותו התהליך מאפשר זיהוי של מאפייני הנפגעים שמשפיעים על הסתברותם להופיע במאגרי המידע המוצלבים, מה שמביא ליצירת מקדי תיקון/ניפוח שונים בהתאם לנסיבות פגיעה שונות. להערכת ההסתברות להימצא במאגר המשולב (או במילים אחרות, להיות מדווח ע"י המשטרה כנפגע קשה) ניתן להיעזר בשיטות סטיסטיטיות לפיתוח מודלים מסבירים רב-פרמטריים - ראה דוגמא במחקר של Amoros et al (2006) שנערך בצרפת.

בהסתמך על ממצאי הספרות המקצועית ומצב הנתונים בישראל, מטרת המחקר הנוכחי הייתה בפיתוח כלים סטטיסטיים להערכת מספר כולל (אמיתי) של נפגעים קשה בתאונות הדרכים בישראל. הכלים הסטטיסטיים שפותחו במחקר מאפשרים חישוב מקדמי תיקון/ניפוח למספרי הנפגעים קשה המדווחים על-ידי המשטרה. המחקר התבסס על נתוני הנפגעים בתאונות בשנת 2008. שימוש בכלים ובמקדמי הניפוח שפותחו במחקר לתיקון מספרי הנפגעים קשה בשנים האחרות מתבסס על הנחת דמיון בין מצב הדיווח שהיה בשנת 2008 ובשנים האחרות.

במחקר פותחו כלים לתיקון מספרי הנפגעים קשה בכל הארץ, לרבות התאונות שהתרחשו במחוזות יהודה ושומרון. עם זאת, כפי שיוסבר בהמשך הדו"ח, הכלים שפותחו במחקר ישימים גם לתיקון מספרי הנפגעים קשה בגבולות הקו הירוק, ללא יו"ש, ובחיתוכי מידע אחרים.

נפגע קשה במחקר הנוכחי מוגדר לפי ה"גדרה האדמיניסטרטיבית"<sup>1</sup>: "אדם שאושפז בעקבות תאונה בבית חולים לתקופה של 24 שעות ויותר", ללא קשר לחומרת הפגיעה לפי הסקאלה הרפואית (ISS). גם לנפגעים מקובץ טראומה, ע"י הלמ"ס חושב מאפיין על סמך הגדרה זו: PazuaKasheTrauma = 1, דהיינו מסומן "פצוע קשה" ע"פ הגדרה אדמיניסטרטיבית כאשר "תאריך שחרור פחות תאריך הגעה לבית חולים" גדול מ-1.

## 1.2. הנתונים ששימשו למחקר

קבצי הנתונים למחקר זה הוכנו ע"י הלמ"ס והם התבססו על תוצרי עבודה מקדימה של הלמ"ס - שילוב מידע מבתי חולים (רישום הטראומה) עם קובץ נתוני תאונות הדרכים של המשטרה, בשנת 2008<sup>2</sup>. פיתוח הכלים במחקר התבסס על שלשה סוגי נתונים שהם: נתוני הקובץ המשולב "משטרה וטראומה", נתוני קובץ "טראומה בלבד", ונתוני קובץ "משטרה בלבד". קבצי הנתונים שנמסרו ע"י הלמ"ס למחקר זה:

קובץ A ("משטרה וטראומה") - המכיל 5,343 רשומות של נפגעים שנמצאו גם בקובץ המשטרה וגם בקובץ הטראומה;

קובץ B ("טראומה בלבד") - המכיל 2,364 רשומות של נפגעים שנמצאו רק בקובץ הטראומה ולא נמצאו להן רשומות מקבילות בקובץ המשטרה;

קובץ C ("משטרה בלבד") - המכיל 528 רשומות של פצועים קשה שנמצאו רק בקובץ המשטרה ולא נמצאו להן רשומות מקבילות בקובץ הטראומה.

כמו כן, נמסר קובץ D המכיל 145 רשומות של הרוגים שנמצאו רק בקובץ המשטרה ולא נמצאו להן רשומות מקבילות בקובץ הטראומה.

יש לציין כי: (א) הנתונים מקובץ הטראומה מתייחסים ל-17 בתי החולים בלבד (מתוך 23 בתי החולים במדינה. ע"פ הערכות, קובץ הטראומה מכסה יותר מ-80% מהפצועים בתאונות הדרכים); (ב) בקבצים שנמסרו למחקר ישנם נפגעים מכל הארץ, לרבות במחוזות יו"ש.

<sup>1</sup> ראה פרסום למ"ס "תאונות דרכים עם נפגעים 2008 חלק א'" וכד'.  
<sup>2</sup> בנספח 3 בפרסום הלמ"ס "תאונות דרכים עם נפגעים 2008 חלק א'" ניתן למצוא הסברים ומתודולוגיה לשילוב שבוע. ראה: [http://www.cbs.gov.il/publications/acci08/pdf/app3\\_h.pdf](http://www.cbs.gov.il/publications/acci08/pdf/app3_h.pdf)



### 1.3. תוכן הדו"ח

בדו"ח זה מוצגים יסודות המחקר אשר שימשו לפיתוח המודלים להערכת מקדמי התיקון למספר נפגעים קשה, הכלים הסטטיסטיים שפותחו במחקר ותוצאות הערכה בעזרת כלים אלה, לשנת 2008.

עקרונות שיטת ה-Capture-Recapture והנחות שעל הנתונים לקיים, כדי שאפשר יהיה להפעיל את השיטה, מתוארים בפרק 2.

פרק 3 מביא ממצאים מחקירה התחלתית של הנתונים בשלושת קבצי המחקר אשר רמזו על מרכיבי השיטה הנדרשת להערכת מספר נפגעים קשה בתנאי הארץ, לרבות חישוב ראשוני של אומדני הנפגעים קשה, ע"פ הגישות שפותחו בחו"ל.

השיטה להערכת מספר נפגעים קשה בתנאי הארץ כוללת שני כלים:

(א) מודל False-Positive המאפשר ניכוי של מספר נפגעים קשה שזוהו ע"י המשטרה ולא אומתו ע"י רישום הטראומה (הנפגעים קל שבטעות זוהו כנפגעים קשה ע"י המשטרה);

(ב) מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא בקובץ טראומה.

המודל הראשון וממצאיו מוצגים בפרק 4 של הדו"ח. המודל השני וממצאיו מוצגים בפרק 5.

פרק 6 מביא סיכום לכלים הסטטיסטיים שפותחו במחקר והאומדן לנפגעים קשה בשנת 2008.

## 2. הצגת מתודולוגית ה-Capture-Recapture

המטרה של טכניקת ה-Capture-Recapture (CR) הינה לאמוד את מספרם הכולל של הנפגעים. כלומר, לאמוד את סך הנפגעים קשה, שהינו סכום הנפגעים קשה שזוהו על-ידי המשטרה והטראומה, אלו שזוהו ע"י טראומה בלבד, אלו שזוהו על ידי משטרה בלבד ואלו שלא נמצאים לא במשטרה ולא בטראומה. כדי להשתמש בשיטה זאת חייבות להתקיים שתי הנחות. ההנחה הראשונה היא שההסתברות ללכידה (capture) ע"י מקור מסוים שווה עבור כל פרט באוכלוסיה. ההנחה השנייה החייבת להתקיים היא הנחה המכונה הנחת ההומוגניות. מכיוון שהנחה זאת מאוד חשובה, נציג ונסביר אותה במספר דרכים. נעיר כאן שאם הנחה זאת לא מתקיימת, יש לחלק את הנתונים לתת-קבוצות הומוגניות ולהפעיל את השיטה עליהן. בהמשך נציג שיטה לביצוע חלוקה כזאת.

נגדיר:

$P$  = נתוני משטרה (צהוב)

$H$  = נתוני טראומה (כתום)

$\underline{P}$  = נתוני לא משטרה

$\underline{H}$  = נתוני לא טראומה

מכאן:

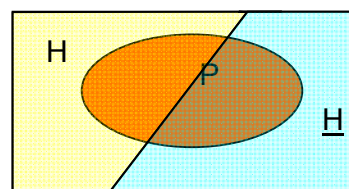
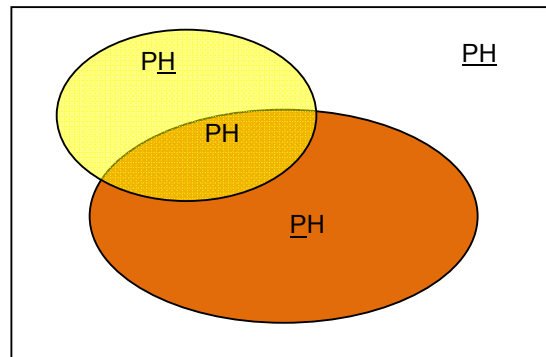
$PH$  = נתונים שגם במשטרה וגם בטראומה (הקובץ המשולב)

$\underline{PH}$  = נתונים שבטראומה בלבד

$\underline{PH}$  = נתונים שבמשטרה בלבד

$\underline{PH}$  = נתונים שלא במשטרה ולא בטראומה

תחילה נציג את הנתונים שברשותנו באופן הבא:



דרך הצגה נוספת:

דרישת ההומוגניות הבסיסית היא שההסתברות ללכידה (capture) ע"י מקור מסוים שווה עבור כל פרט באוכלוסיה, ובנוסף פרופורציית הנפגעים (מתוך הקובץ המשולב) המדווחים גם ע"י המשטרה מקרב הנפגעים המדווחים בקובץ טראומה, שווה לפרופורציית הנפגעים המדווחים ע"י המשטרה מסך

$$\frac{PH}{H} = \frac{P}{Total} : \text{כלומר: כל התאונות.}$$

$$\frac{PH}{PH} = \frac{PH}{PH} \text{ כ-מנוסחת } \frac{PH}{PH}$$

כלומר: היחס בין הנפגעים שלא מדווחים הן ע"י המשטרה והן ע"י טראומה ("התא החסר"), לאלה שמדווחים ע"י המשטרה אך לא ע"י טראומה ("משטרה בלבד"), שווה ליחס בין אלה שמדווחים ע"י טראומה אך לא ע"י המשטרה ("טראומה בלבד") לאלה שמדווחים ע"י טראומה ומשטרה (נתוני קובץ משולב). יש לשים לב שבניסוח זה מוגדרים ארבעה מקטעים זרים של הנתונים, והקשרים ביניהם נראה ששתי הנחות אלה אקוויולנטיות.

$$\frac{PH}{H} = \frac{P}{Total}$$

$$Total = \underline{PH} + PH + \underline{PH} + PH$$

$$P = \underline{PH} + PH; H = \underline{PH} + PH \Rightarrow$$

$$PH * (\underline{PH} + PH + \underline{PH} + PH) = (\underline{PH} + PH) * (\underline{PH} + PH) \Rightarrow$$

$$\frac{\underline{PH}}{PH} = \frac{PH}{PH}$$

דרך אחרת להציג ולהבין את הנחת ה-CR היא:

$$\frac{PH}{H} = \frac{P}{Total};$$

$$\frac{PH/Total}{H/Total} = \frac{\hat{Pr}(P \cap H)}{\hat{Pr}(H)} = \hat{Pr}(P/H);$$

$$\frac{P}{Total} = \hat{Pr}(P)$$

$$CR \Rightarrow \hat{Pr}(P/H) = \hat{Pr}(P)$$

הסיכוי לכך שנפגע כלשהו מתוך כל הנפגעים יופיע בקובץ הנפגעים של המשטרה, נקרא גם הסיכוי שנפגע כלשהו מתוך כל הנפגעים ילכד ע"י המשטרה מתוך כל הנפגעים (the police capture probability).

הסיכוי לכך שנפגע כלשהו מתוך כל הנפגעים יופיע בקובץ הנפגעים של הטראומה, נקרא גם הסיכוי שנפגע כלשהו מתוך כל הנפגעים ילכד ע"י הטראומה מתוך כל הנפגעים (the trauma capture probability).

הסיכוי לכך שבהינתן שנפגע כלשהו מופיע בין הנפגעים שנרשמו בקובץ טראומה (= נתון שנתפס תפיסה ראשונה) הוא יופיע גם בקובץ הנפגעים של המשטרה, נקרא גם הסיכוי להילכד מחדש ע"י המשטרה (the police re-capture probability).

הנחת ההומוגניות, המכונה במסמך זה גם כהנחת ה-CR (capture-recapture) אומרת שהסיכוי לכך שנפגע כלשהו מתוך כל הנפגעים ילכד ע"י המשטרה, שווה לסיכוי שבהינתן שנפגע כלשהו נלכד ע"י טראומה הוא ילכד גם ע"י המשטרה.

מהנחת ה-CR נובע גם:

$$\frac{PH}{H} = \frac{P}{Total} \Rightarrow$$

$$Total = P * \frac{1}{\left(\frac{PH}{H}\right)} = P * CF$$

where

$$CF = \frac{1}{\left(\frac{PH}{H}\right)} = \frac{1}{\hat{Pr}(P/H)}$$

כאשר: CF = Correction Factor

כלומר, מקדם התיקון לנתוני המשטרה, אם מתקיימת הנחת ה-CR, הוא 1 חלקי פרופורציית הנפגעים בקובץ המשולב מתוך סך הנפגעים המדווחים ע"י רישום הטראומה, שזה שווה בעצם ל-1 חלקי ההסתברות המותנית של נפגע להיות מדווח ע"י המשטרה בהינתן שהוא דווח ע"י טראומה. כאשר  $\hat{Pr}(P/H)$  מחושב באמצעות מודל, משמש  $1/\hat{Pr}(P/H)$  כגורם תיקון לנתוני המשטרה.

דרך נוספת להבנת הנחת ה-CR:

$$\frac{\hat{Pr}(P \cap H)}{\hat{Pr}(H)} = \hat{Pr}(P/H);$$

$$CR \Rightarrow \hat{Pr}(P/H) = \hat{Pr}(P)$$

$$\Rightarrow \hat{Pr}(P \cap H) = \hat{Pr}(P) * \hat{Pr}(H)$$

כלומר, הנחת ה-CR שקולה להנחת אי-תלות המקורות, המשטרה והטראומה - ראה, לדוגמה, Tilling and Sterne (1999). בשל הנחה זאת ניתן לחשב את  $\hat{Pr}(P)$ , כלומר, את ההסתברות של נפגע להילכד ע"י המשטרה, תוך שימוש בנתוני שלושת סוגי הנתונים (משטרה בלבד, טראומה בלבד וחיתוך המשטרה והטראומה), ואז להשתמש ב-  $1/\hat{Pr}(P)$  כגורם תיקון לנתוני המשטרה.

כל אחת משתי הדרכים להבנת הנחת ה-CR מובילה לחישוב שונה של גורם התיקון לנתוני המשטרה:

דרך א' - מובילה לפיתוח מודל לקביעת  $\hat{Pr}(P/H)$ , תוך שימוש בנתוני הטראומה, ואז ניפוח נתוני המשטרה ע"י  $1/\hat{Pr}(P/H)$ ;

דרך ב' - משתמשת בנתוני משטרה וטראומה לפיתוח מודל לקביעת  $\hat{Pr}(P)$ , ואז ניפוח נתוני משטרה על ידי  $1/\hat{Pr}(P)$ .

בדרך א', משתמשים בנתוני הטראומה, הן מהקובץ המשולב והן מקובץ הטראומה בלבד, לשם פיתוח מודל לקביעת ההסתברות ללכידת נפגע קשה ע"י המשטרה. היתרון בשיטה זאת הוא שניתן להפעיל בתהליך בחירת המשתנים של המודל אלגוריתמים מתוחכמים, לא בהכרח פרמטריים (כגון: CART), כאשר לאפשרות זו ישנה חשיבות במחקר הראשון בנושא. לכן, בעבודה זו ניישם את הדרך א'. יחד עם זאת, החיסרון בשיטה זו הוא הצורך להשתמש רק במשתנים הנמצאים הן בקובץ המשטרה והן בקובץ הטראומה.

בדרך ב', משתמשים הן בנתוני המשטרה והן בנתוני הטראומה לקביעת מודל ל- $Pr(P)$ . דרך זו דורשת חישובים מורכבים יותר, ורק מיקרים מאוד פרטיים שלה מיושמים, משום שבמקרים אלה ניתן להראות אקוויולנטיות של המודל המתקבל, למודל הנמצא בתוכנות סטטיסטיות סטנדרטיות. בחירת משתנים מתאימים למודל אינה ניתנת לביצוע ישיר כמו במקרה הראשון, פרט למקרים מסוימים, ודורשת חישובים מורכבים. אולם, יחד עם זאת, ניתן לראות מהפיתוחים ב-Alho (1990), יתרון חשוב של השיטה הזאת על פני השיטה הקודמת. היתרון הוא שניתן להשתמש במשתנים שונים למידול ההסתברות של כל מקור. כלומר, ניתן למדל את ההסתברות לתפיסה של המשטרה ע"י אוסף אחד של משתנים ואת ההסתברות לתפיסה של טראומה ע"י אוסף אחר של משתנים. זאת תכונה חשובה, לאור העובדה שקיימים משתנים שונים בקבצים השונים. אולם, כאמור, שימוש בשיטה כה מורכבת לבחינה הראשונית של בניית מקדמי הניפוח, והבנת הגורמים המשפיעים על הסתברויות התפיסה, אינה מומלצת, ולכן לא נבחרה במחקר זה.

נעיר, שנושא חישוב הסתברויות תפיסה כפונקציה של משתנים מסבירים הינו נושא מחקר פעיל כעת - ראה, לדוגמה, את Stoklosa et al (in press)<sup>3</sup>, המלווה בספריית PL.popN של תוכנת R שהתפרסמה לראשונה באפריל 2011<sup>4</sup>.

יש לשים לב שהשימוש בשיטת ה-Capture-Recapture במחקר זה שונה מהשימוש הרגיל. בשימוש הרגיל ישנו מידע משני מקורות לגבי חלק מאוכלוסייה בתקופה נתונה, והמטרה היא להשלים את המידע לגבי כל האוכלוסייה, בתקופה בה נאסף המידע החלקי. לעומת זאת, בישום הנוכחי המטרה היא להשתמש במידע לגבי התקופה בה נאסף המידע החלקי, לשם השלמת המידע לגבי כל האוכלוסייה, על מנת להשתמש במידע זה בתקופות מאוחרות יותר וכאשר יהיה נתון רק אחד ממקורות המידע (קובץ המשטרה בלבד).

---

<sup>3</sup> Stoklosa, Hwang, Wu and Huggins. (In Press). Heterogeneous capture-recapture models with covariates: A partial likelihood approach for closed populations. Biometrics. DOI: 10.1111/j.1541-0420.2011.01596.x  
<sup>4</sup> ראה: <http://cran.r-project.org/web/packages/PL.popN/PL.popN.pdf>

### 3. חקר מקדים של הנתונים

בשלב המקדים בוצעה סטטיסטיקה תיאורית של נתוני הקובץ המשולב (A) - נתוני המשטרה וטראומה ביחד. כמו כן, כדי לאמת את נכונות הקלטים נבדקו נתוני קובץ זה בשילוב קבצים אחרים שהתקבלו מהלמ"ס.

#### 3.1. נתוני קובץ A המשולב, כולל יו"ש

קובץ A ("משטרה וטראומה") כלל כ-5,343 רשומות של נפגעים שנמצאו גם בקובץ המשטרה וגם בקובץ הטראומה. טבלה 3.1 מציגה את פילוג הרשומות לפי חומרת הפגיעה - המשתנה HUMRAT\_PGIA\_LMS אשר נוצר בעקבות שילוב הנתונים מטרואומה עם נתוני המשטרה והמשקף את תוצאות השילוב. ניתן לראות ש:

- ב-62% מהמקרים (3317) נמצאה התאמה בין הגדרת הנפגעים במשטרה ובטראומה;
- ב-36% מהמקרים (1938) נפגעים קל לפי המשטרה נמצאו כנפגעים קשה לפי הטראומה,
- בעוד שב-1.35% מהמקרים (72) נפגע שהוגדר קשה ע"י המשטרה נתגלה כנפגע קל לפי הטראומה<sup>5</sup>. (הכול מתוך המקרים בהם יש רישום הן של משטרה והן של טראומה).

טבלה 3.1. פילוג הרשומות בקובץ A, כולל יו"ש, לפי חומרת הפגיעה

של נפגעים שנקבעה לאחר הבדיקה

אחוז	מספר	חומרת פגיעה לפי למ"ס HUMRAT_PGIA_LMS	(ריק)
62.08	3317	במקרים שהחומרה זהה במשטרה ובטראומה השדה כאן יהיה ריק	
36.27	1938	אם במשטרה היה רשום 3 (נפגע קל) ובטראומה היה רשום 2 (נפגע קשה) החומרה הסופית היא 2 (נפגע קשה)	2
1.35	72	אם במשטרה היה רשום 2 (נפגע קשה) ובטראומה היה רשום 3 (נפגע קל) החומרה הסופית היא 3 (נפגע קל)	3
0.26	14	במקרה של פטירה מאוחרת לפי המשטרה וגם טראומה	5
0.04	2	במקרים שבמשטרה היה 3 (נפגע קל) ובטראומה היה 2 (נפגע קשה) וגם היה פטירה מאוחרת, או שבמשטרה היה הרוג ובטראומה היה פטירה מאוחרת, יקבל 7	7
100	5343	סה"כ	

טבלה 3.2 מציגה פילוג של הרשומות לפי הסיווג המקורי של הנפגעים ע"י המשטרה (משתנה HUMRAT\_PGIA) והסיווג הסופי של הנפגעים בעקבות שילוב הנתונים מטרואומה עם נתוני המשטרה (משתנה HUMRAT\_PGIA\_LMS). ניתן לראות כי:

- ב-3317 רשומות קיימת הסכמה בין המשטרה והטראומה לגבי חומרת הפגיעה, כאשר אלה נפגעים ברמות חומרה שונות, לרבות 1607 נפגעים קשה שהוגדרו ע"י המשטרה;

<sup>5</sup> כאן ובהמשך "נפגע קשה" הינו לפי הגדרה אדמיניסטרטיבית - מאושפז ליותר מ-24 שעות. מחושב לפי הפרשי ימים בין תאריך שחרור לתאריך אשפוז.

- 1926 נפגעים קשה הינם תוספת לנפגעים קשה שנתגלתה בעקבות הצלבת הנתונים עם טראומה - מקרים אלה הוגדרו ע"י המשטרה כנפגעים קל. (כאן יהיה צורך לנפח את נתוני המשטרה - המטרה - בניתוח תהיה לאפיין תצפיות כאלה ולנסות על-פי מידע זה לקבוע מקדמי ניפוח לנתוני המשטרה);

- ב-66 מקרים המשטרה קבעה נפגע קשה אשר נתגלה כנפגע קל לפי הנתונים של טראומה, כאשר סה"כ היו  $1673=1607+66$  נפגעים קשה שדווחו ע"י המשטרה. מקרים אלה מעלים בטעות את מספר הנפגעים קשה. (כאן יהיה צורך להפחית מנתוני המשטרה - המטרה בניתוח תהיה לאפיין תצפיות כאלה ולנסות על-פי מידע זה לנכות תצפיות כאלה מקובץ המשטרה);

- לגבי 12 ו-6 מקרים לא היה מידע על חומרת הפגיעה לפי המשטרה והם הוגדרו לפי הטראומה.

טבלה 3.2. פילוג הרשומות בקובץ A, כולל יו"ש, לפי הסיווג המקורי של חומרת הפגיעה ע"י המשטרה ולפי חומרת הפגיעה של הנפגעים שנקבעה לאחר השילוב עם טראומה

HUMRAT_PGIA - חומרת פגיעה לפי המשטרה					HUMRAT_PGIA_LMS - חומרת פגיעה בעקבות הבדיקה
מספר אחוז (ללא שינוי)	1 הרוג	2 נפגע קשה	3 נפגע קל	סה"כ	
1	106	1607	1603	3317	
0.02	1.98	30.08	30.00	62.08	
12	0	0	1926	1938	2 - במשטרה 3 בטרומה
0.22	0.00	0.00	36.05	36.27	(נפגע קשה)
6	0	66	0	72	3 - במשטרה 2 בטרומה
0.11	0.00	1.24	0.00	1.35	(נפגע קל)
0	14	0	0	14	5 - פטירה מאוחרת, לפי המשטרה וגם טראומה
0.00	0.26	0.00	0.00	0.26	
0	2	0	0	2	7 - פטירה מאוחרת (הרוג לפי משטרה ופטירה מאוחרת לפי טראומה)
0.00	0.04	0.00	0.00	0.04	
19	122	1673	3529	5343	סה"כ
0.36	2.28	31.31	66.05	100.00	

### 3.2. נתוני קובץ A המשולב, ללא יו"ש

ללא מחוז יו"ש (יחידות משטרה 41, 43) בקובץ נותרו 5139 רשומות. טבלה 3.3 מציגה פילוג משולב של רשומות אלה לפי הסיווג המקורי של הנפגעים ע"י המשטרה (משתנה HUMRAT\_PGIA) והסיווג הסופי של הנפגעים בעקבות שילוב הנתונים מטרומה עם נתוני המשטרה (משתנה HUMRAT\_PGIA\_LMS). ניתן לראות כי:

- ב-3167 רשומות קיימת הסכמה בין משטרה וטרומה לגבי חומרת הפגיעה, כאשר אלה נפגעים ברמות חומרה שונות, לרבות 1501 נפגעים קשה שהוגדרו ע"י המשטרה;

- 1879 נפגעים קשה הינם תוספת לנפגעים קשה שנתגלתה בעקבות הצלבת הנתונים עם טראומה - מקרים אלה הוגדרו ע"י המשטרה כנפגעים קל. (כאן יהיה צורך לנפח את נתוני המשטרה);

- ב-59 מקרים המשטרה קבעה נפגע קשה שנתגלה כנפגע קל לפי הנתונים של טראומה, כאשר סה"כ היו  $1560=1501+59$  נפגעים קשה שדווחו ע"י המשטרה. מקרים אלה מעלים בטעות את מספר הנפגעים קשה. (כאן יהיה צורך להפחית מנתוני המשטרה - לנכות תצפיות כאלה מקובץ המשטרה);

- בנוסף, בקובץ נמצאים 100 הרוגים שהוגדרו ע"י המשטרה ועוד 16 מקרים שהוגדרו כנפגעים קשה ע"י המשטרה אך מהווים פטירה מאוחרת. (בנוסף, בקובץ D שהתקבל מהלמ"ס ישנן 145 רשומות של הרוגים שנמצאו רק בקובץ המשטרה ולא נמצאו בקובץ הטראומה, מתוכם 144 מקרים בגבולות הקו הירוק).

טבלה 3.3. פילוג הרשומות בקובץ A, ללא יו"ש, לפי הסיווג המקורי של חומרת הפגיעה ע"י המשטרה ולפי חומרת הפגיעה של הנפגעים שנקבעה לאחר השילוב עם טראומה

HUMRAT_PGIA - חומרת פגיעה לפי המשטרה					HUMRAT_PGIA_LMS - חומרת פגיעה בעקבות הבדיקה
סה"כ	3 נפגע קל	2 נפגע קשה	1 הרוג	(ריק)	מספר אחוז (ללא שינוי)
3167	1565	1501	100	1	
61.63	30.45	29.21	1.95	0.02	
1891	1879	0	0	12	2 - במשטרה 3 בטראומה 2 (נפגע קשה)
36.80	36.56	0.00	0.00	0.23	
65	0	59	0	6	3 - במשטרה 2 בטראומה 3 (נפגע קל)
1.26	0.00	1.15	0.00	0.12	
14	0	0	14	0	5 - פטירה מאוחרת, לפי המשטרה וגם טראומה
0.27	0.00	0.00	0.27	0.00	
2	0	0	2	0	7 - פטירה מאוחרת (הרוג לפי משטרה ופטירה מאוחרת לפי טראומה)
0.04	0.00	0.00	0.04	0.00	
5139	3444	1560	116	19	סה"כ
100.00	67.02	30.36	2.26	0.37	

נשתמש במשתנה PazuaKasheTrauma שמסמן "נפגע קשה" ע"פ הגדרה אדמיניסטרטיבית בעקבות העיבודים של הלמ"ס - שילוב המידע מטראומה עם קובץ המשטרה, כאשר PazuaKasheTrauma = 1 מסמן זיהוי נפגע קשה (תאריך שחרור פחות תאריך הגעה לבית חולים גדול מ-1). טבלה 3.4 מציגה פילוג של המקרים שסווגו כנפגע קשה, בעקבות שילוב המידע מטראומה עם המידע מהמשטרה, לפי חומרת הפגיעה המקורית שנקבעה במשטרה (משתנה HUMRAT\_PGIA) והסיווג הסופי של הנפגעים שנקבע בעקבות הבדיקה (משתנה HUMRAT\_PGIA\_LMS).

סה"כ, בקובץ המשולב A, ללא יו"ש, נמצאו 3408 מקרים של נפגעים קשה, מתוכם:

- 1501 מקרים של נפגעים קשה שהוגדרו כנפגעים קשה הן ע"י המשטרה והן ע"י הטראומה;

- 1891 מקרים שהתווספו בעקבות בדיקת המידע מקובץ טראומה, לרבות 1879 מקרים של נפגעים קל לפי המשטרה ועוד 12 מקרים שהיו ללא פגיעה (מעורבים שלא נפגעו) לפי דיווח המשטרה אך נמצאו בקובץ טראומה;

- ועוד 16 מקרים של פטירה מאוחרת.



טבלה 3.4. פילוג הרשומות בקובץ A, ללא יו"ש, שנתגלו כנפגעים קשה לפי הטראומה, לפי הסיווג המקורי של חומרת הפגיעה ע"י המשטרה ולפי מצב חומרת הפגיעה שנקבע לאחר הבדיקה

HUMRAT_PGIA - חומרת פגיעה לפי המשטרה					HUMRAT_PGIA_LMS - חומרת פגיעה בעקבות הבדיקה
סה"כ	3 נפגע קל	2 נפגע קשה	1 הרוג	(ריק)	מספר אחוז (ללא שינוי)
1501	0	1501	0	0	
44.04	0.00	44.04	0.00	0.00	
1891	1879	0	0	12	2 - במשטרה 3 בטרומה 2 (נפגע קשה)
55.49	55.13	0.00	0.00	0.35	
0	0	0	0	0	3 - במשטרה 2 בטרומה 3 (נפגע קל)
0.00	0.00	0.00	0.00	0.00	
14	0	0	14	0	5 - פטירה מאוחרת, לפי המשטרה וגם טראומה
0.41	0.00	0.00	0.41	0.00	
2	0	0	2	0	7 - פטירה מאוחרת (הרוג לפי משטרה ופטירה מאוחרת לפי טראומה)
0.06	0.00	0.00	0.06	0.00	
3408	1879	1501	16	12	סה"כ
100.00	55.13	44.04	0.47	0.35	

### 3.3. הערכות ראשוניות של מספר הנפגעים קשה

על סמך הממצאים של מספרי הנפגעים קשה שנתגלו בקובץ המשולב של המשטרה וטראומה (A) ובקבצים המשלימים (B, C), בוצעו חישובים ראשוניים של סך מספר הנפגעים קשה, בעזרת שיטות שונות שנמצאו בספרות. בשנת 2008, המצב עם הנפגעים קשה, ללא יו"ש, היה כלהלן:

#### הערכה 1:

ב-17 בתי החולים של רישום הטראומה המשטרה *דיווחה* על 1826 נפגעים קשה שהתקבלו כלהלן:  
 1560 (נפגעים קשה לפי הגדרת המשטרה בקובץ A המשולב - ראה טבלה 3.3)  
 + 250 נפגעים קשה שהוגדרו ע"י המשטרה ולא נמצאו בטרומה (קובץ C, ב-17 בתי החולים הנבחרים)  
 + 16 מקרים של פטירה מאוחרת (בקובץ A המשולב)  
**= 1826.**

#### הערכה 2:

ב-17 בתי החולים של רישום הטראומה *בעקבות הצלבת המידע מטרומה* עם המידע של המשטרה נמצאו סה"כ 3658 נפגעים קשה שהתקבלו כלהלן:  
 1560 (נפגעים קשה לפי הגדרת המשטרה שנמצאו גם בטרומה, לפי קובץ A המשולב)  
 - 59 מקרים (שהוגדרו כנפגעים קשה ע"י המשטרה אך נמצאו כנפגעים קל בטרומה, לפי קובץ A)  
 + 1879 מקרים (שהוגדרו כנפגעים קל ע"י המשטרה ונמצאו כנפגעים קשה בטרומה, לפי קובץ A)  
 + 12 מקרים (שדווחו כמעורבים ללא פגיעה ע"י המשטרה ונתגלו כנפגעים קשה לפי הטראומה, לפי קובץ A)  
 + 250 נפגעים קשה שהוגדרו ע"י המשטרה ולא נמצאו בטרומה (קובץ C, ב-17 בתי החולים הנבחרים)  
 + 16 מקרים של פטירה מאוחרת (בקובץ A המשולב)  
**= 3658.**

## הערות להערכה 2:

\* (א) המספר "1879" (תוספת מקרים של נפגעים קשה שבמקור הוגדרו כנפגעים קל ע"י המשטרה) מורכב משני חלקים: 906 מקרים של נפגעים קל שהיו בקובץ "ת"ד" ועוד 973 מקרים של נפגעים קל שהיו בקובץ "כללי עם נפגעים".

(ב) בנוסף, קיים קובץ B המכיל 2,364 רשומות של נפגעים שנמצאו רק בקובץ הטרואמה ולא נמצאו להן רשומות מקבילות בקובץ המשטרה. בקובץ זה יש 1608 מקרים עם משך אשפוז מעל 1 יום, דהיינו המתאימים להגדרת "נפגע קשה". מתוך מספרים אלה יכולה להיות תוספת לסך הנפגעים קשה.

כמו כן, על סמך קובץ C הכולל נפגעים קשה שהוגדרו ע"י המשטרה ולא נמצאו בטרואמה, ניתן להראות הערכה של סה"כ נפגעים קשה (בכל בתי החולים) בשנת 2008, ללא יו"ש (ללא יו"ש בקובץ C נמצאו 487 נפגעים קשה):

## הערכה 3:

סה"כ ארצי, ללא יו"ש, בשנת 2008, המשטרה דיווחה על 2063 נפגעים קשה שהתקבלו כלהלן:  
1560 (נפגעים קשה לפי הגדרת המשטרה בקובץ A המשולב)  
+ 487 נפגעים קשה שהוגדרו ע"י המשטרה ולא נמצאו בטרואמה (קובץ C, ללא יו"ש)  
+ 16 מקרים של פטירה מאוחרת (בקובץ A המשולב)  
= 2063.

## הערכה 4:

סה"כ ארצי, בשנת 2008, בעקבות הצלבת המידע מטרואמה עם המידע של המשטרה ניתן לקבל 3895 נפגעים קשה, לפי החישוב שלהלן:

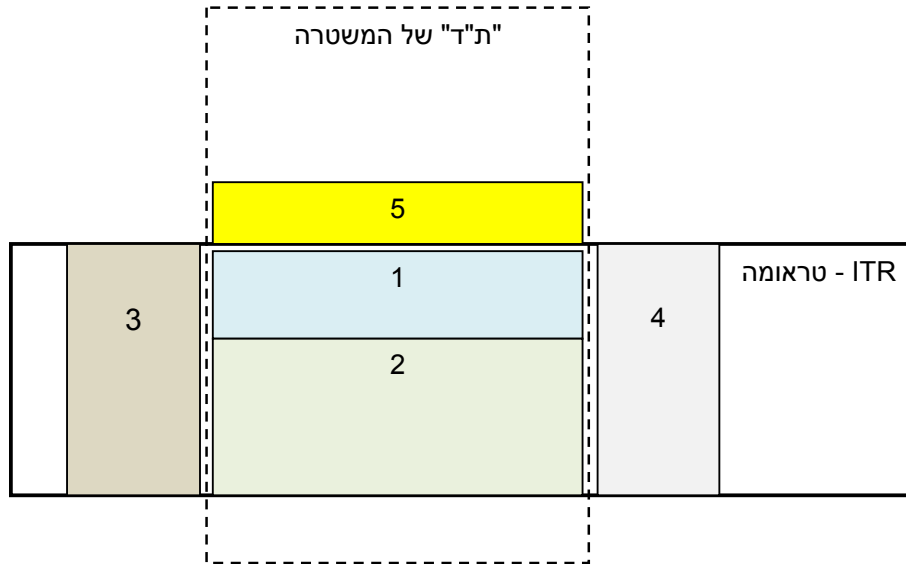
1560 (נפגעים קשה לפי הגדרת המשטרה שנמצאו גם בטרואמה, לפי קובץ A המשולב)  
- 59 מקרים (שהוגדרו כנפגעים קשה ע"י המשטרה אך נמצאו כנפגעים קל בטרואמה, לפי קובץ A)  
+ 1879 מקרים (שהוגדרו כנפגעים קל ע"י המשטרה ונמצאו כנפגעים קשה בטרואמה, לפי קובץ A)  
+ 12 מקרים (שדווחו כמעורבים ללא פציעה ע"י המשטרה ונתגלו כנפגעים קשה לפי הטרואמה, לפי קובץ A)  
+ 487\*\* נפגעים קשה שהוגדרו ע"י המשטרה ולא נמצאו בטרואמה (קובץ C, בכל בתי החולים, ללא יו"ש)  
+ 16 מקרים של פטירה מאוחרת (ראינו בקובץ A המשולב)  
= 3895.

## הערות להערכה (4):

\*\* (א) לגבי המספר "487" (מספר נפגעים קשה שנמצאו במשטרה ולא נמצאו בטרואמה, בייחוד ביתר בתי החולים, מלבד 17 הכלולים בטרואמה) ניתן להפעיל מקדם ניפוח, מתוך הנחה על דיווח חסר של המשטרה לעומת בתי החולים.

(ב) בנוסף, קיים קובץ B המכיל 2,364 רשומות של נפגעים שנמצאו רק בקובץ הטראומה ולא נמצאו בקובץ המשטרה. בקובץ זה יש 1608 מקרים עם משך אשפוז מעל 1 יום, דהיינו המתאימים להגדרת "נפגע קשה". מתוך מספרים אלה יכולה להיות תוספת לסך הנפגעים קשה.

מכאן, את תמונת המצב עם השלמת מספרי הנפגעים קשה בתאונות הדרכים בישראל ניתן להציג בעזרת ציור 3.1.

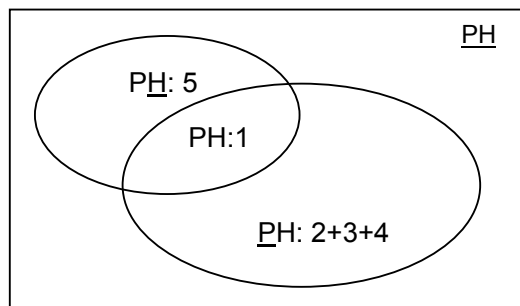


כאשר:

- 1 - נפגעים קשה שהוגדרו כ"קשה" ע"י המשטרה ונמצאו גם בטראומה;
- 2 - נפגעים שהוגדרו כנפגעים "קל" ע"י המשטרה, בקובץ "ת"ד", ונתגלו כנפגעים קשה בטראומה;
- 3 - נפגעים שהוגדרו כנפגעים "קל" ע"י המשטרה, היו בקובץ "כללי עם נפגעים", ונתגלו כנפגעים קשה בטראומה;
- 4 - נפגעים קשה בעקבות תאונות דרכים שנמצאו בטראומה בלבד ולא נמצאו בקבצי המשטרה;
- 5 - נפגעים קשה שהוגדרו ע"י המשטרה בקובץ "ת"ד" ולא נמצאו בטראומה.

ציור 3.1. מרכיבי הנתונים להערכת מספר נפגעים קשה בתאונות הדרכים בארץ.

במונחים של הצלבה בין מאגר המשטרה (P) עם מאגר הטראומה (H) - שיטת ה-CR שתאורה בפרק 2 - ניתן להציג את התמונה באופן הבא - ציור 3.2.



ציור 3.2. הצגת מרכיבי הנתונים בישראל להערכת מספר נפגעים קשה במונחים של שיטת ה-CR.

כלומר, סך הנפגעים קשה שדווחו ע"י המשטרה בקובץ "ת"ד" (P) מורכב מקבוצות 1+5; סך הנפגעים קשה שנמצאו בטרומה (H) מורכב מקבוצות 1+2+3+4; סך כולל של הנפגעים קשה בתאונות הדרכים (Total) מוגדר כסך הקבוצות 1+2+3+4+5 בתוספת חלק שלא נמצא באף מאגר (PH).

#### הערכה 5:

בהסתמך על הגדרת מקדם התיקון שהוצגה בפרק 2, את המספר הכולל של נפגעים קשה בתאונות הדרכים בשנת 2008 ניתן להעריך באופן הבא:

$$\text{Total} = P * CF = P * 1 / (PH/H) = P * H / PH$$

כלומר, את מספר הנפגעים קשה שדווח ע"י המשטרה יש להכפיל ביחס בין מספר הנפגעים קשה שנמצא בטרומה חלקי המספר המשותף של נפגעים קשה שנמצא הן במשטרה והן בטרומה.

במונחים של קבוצות 1-5 שהוגדרו לעיל, הנוסחה תיראה כלהלן:

$$\text{Total} = P * H / PH = (1+5) * (1+2+3+4) / (1)$$

ע"פ הערכות לשנת 2008 שהוצגו לעיל - נפגעים קשה בתאונות הדרכים ללא יו"ש (הערכה 4), המספרים בתוך הקבוצות יהיו כלהלן:

$$\text{קבוצה 1} = 1560 - 59 + 16 = 1517 \text{ מקרים}$$

$$\text{קבוצה 5} = 487 \text{ מקרים}$$

$$\text{קבוצות 2 ו-3 ביחד} = 1879 + 12 = 1891 \text{ מקרים}$$

$$\text{קבוצה 4} = 1608 \text{ מקרים}$$

מכאן, המספר הכולל של נפגעים קשה בתאונות הדרכים בשנת 2008 יהיה כלהלן:

$$\text{Total} = P * H / PH = (1517 + 487) * (1517 + 1891 + 1608) / (1517) = \mathbf{6626}$$

הערכה זו מסתמכת על הנחת ההומוגניות של שיטת CR מבלי לוודא תנאים לקיומה בנתונים הנבחרים ולכן, מתאימה לשמש כאומדן מקורב בלבד. כמו כן, הערכה זו לא התחשבה בתופעות כגון:  
- הפחתה במספר הנפגעים קשה שהוגדרו לכאורה ע"י המשטרה בשל היותם נפגעים קל לפי הטרומה;

- תיקון לפי יתר בתי החולים: בשל העובדה שרישום הטרומה כולל רק 17 בתי החולים, קיים צורך להשלים את תת-הדיווח מיתר בתי החולים;  
- תיקון בשל המקרים שנרשמו ביו"ש.

#### הערכה 6:

על סמך השיטה שפותחה לתיקון מספר הנפגעים קשה במחקר האירופי SafetyNet - Broughton, Amoros et al (2008) - בוצעה הערכה נוספת של סך מספר הנפגעים קשה בשנת 2008, ללא יו"ש. ע"פ המחקר האירופי, נגדיר:

(i) ser - מספר הנפגעים קשה שדווחו ע"י המשטרה עם משך אשפוז i ימים;

(i) sli - מספר הנפגעים קל שדווחו ע"י המשטרה עם משך אשפוז i ימים;

(i) hnp - מספר הנפגעים עם משך אשפוז i ימים שדווחו ע"י המאגר הרפואי ולא דווחו ע"י המשטרה.

אזי המספר הכולל שנפגעים קשה עם משך אשפוז i ימים יוערך באופן הבא:

$$N = \text{estimated total of serious (i)} + \text{estimated total of slight (i)}$$

כאשר

$$\text{estimated total of serious (i)} = \text{ser(i)} * [\text{ser(i)} + \text{sli(i)} + \text{hnp(i)}] / [\text{ser(i)} + \text{sli(i)}]$$

$$\text{estimated total of slight (i)} = \text{sli(i)} * [\text{ser(i)} + \text{sli(i)} + \text{hnp(i)}] / [\text{ser(i)} + \text{sli(i)}]$$

טבלה 3.5 מציגה מספרים של הנפגעים קשה וקל בתאונות הדרכים בישראל, בשנת 2008, ללא יו"ש, לפי קבצי המחקר (A, B, C) ומקורות נוספים<sup>6</sup>, אשר נדרשים לביצוע הערכה לפי שיטת ה-SafetyNet. יש לשים לב כי המספר הכולל של נפגעים קל לפי המשטרה במקרה זה הינו סה"כ משני קבצים: "ת"ד" ו-"כללי עם נפגעים". (זאת, מכיוון שחלק מהנפגעים קשה בקובץ המשולב A של המשטרה והטראומה נתגלו בקובץ "כללי עם נפגעים"). כמו כן, משך האשפוז נבחר מעל 1 יום, לפי ההגדרה של נפגעים קשה בארץ.

על סמך המספרים בטבלה 3.5, ערכי ההגדרות יהיו כלהלן:

$$1608 = \text{hnp (1)}, 1891 = \text{sli (1)}, 1517 = \text{ser (1)}, i=1$$

לכן, לפי הערכה בשיטת ה-SafetyNet, התוצאות יהיו כלהלן:

$$2233 = \text{estimated total of serious (1)}$$

$$2783 = \text{estimated total of slight (1)}$$

מכאן, מספר הנפגעים קשה בשנת 2008, ללא יו"ש, היה **5016**.

מבחינת מרכיבי הערכה זו עולה כי המספר הכולל של הנפגעים קשה, למעשה, מהווה סכום של שלושה מספרים שהם: מספר הנפגעים קשה שדווחו כנפגעים קשה ע"י המשטרה ונמצאו בטראומה; מספר הנפגעים שדווחו כנפגעים קל ע"י המשטרה ונמצאו בטראומה (דהיינו, תוקנו לנפגעים קשה); ומספר הנפגעים קשה שנמצאו בטראומה ולא נמצאו במשטרה. במילים אחרות, מספר זה משקף את סה"כ הנפגעים קשה שנמצאו בקובץ טראומה (מאגר "H" לפי שיטת ה-CR או סכום הקבוצות 1,2,3,4 לפי ציור 3.1). מכאן, באומדן זה חסר מספר הנפגעים קשה שנמצאו במשטרה ולא נמצאו בטראומה.

טבלה 3.5 מספרים של נפגעים קשה וקל בתאונות הדרכים בישראל, בשנת 2008, ללא יו"ש, לביצוע הערכת הנפגעים קשה בשיטת ה-SafetyNet

	Police serious	Police slight	Not police	Estimated total Serious	Estimated total Slight	% serious not reported by police
Hospital <= 1 day	59	1565	757	87	2294	32%
Hospital > 1 day	1517	1891	1608	2233	2783	32%
Not hospital	487	110660				
Grand total	2063	114116*	2365			

\* מקבצי ת"ד ו-"כללי עם נפגעים", ביחד

<sup>6</sup> פרסום למ"ס "תאונות דרכים עם נפגעים 2008 חלק א" - סיכומים לפי קובץ ת"ד ולפי תיקי "כללי עם נפגעים"

כדי לתת אפשרות לביצוע הערכות בשנים הבאות, ללא זמינות הקובץ המשולב של המשטרה והמאגר הרפואי, Broughton, Amoros et al (2008) מציעים לפעול באופן הבא: לחשב את היחס בין מספר הנפגעים המוערך בעזרת המידע המשולב לבין מספר הנפגעים המדווח ע"י המשטרה, ולהכפיל אותו במספר הנפגעים המדווח ע"י המשטרה, כאשר חישוב זה מבוצע הן עבור הנפגעים קשה והן עבור הנפגעים קל שתרמו להערכת הנפגעים קשה (ראה אומדן N לעיל). על סמך הנתונים בטבלה 3.5, ההערכה תהיה כלהלן:

$$N = 2233/2063 * N_1 + 2783/114116 * N_2 = 1.082 * N_1 + 0.024 * N_2$$

כאשר

$N_1$  - מספר הנפגעים קשה המדווח ע"י המשטרה (number of serious casualties reported by the police),

$N_2$  - מספר הנפגעים קל המדווח ע"י המשטרה (number of slight casualties reported by the police).

כאמור, היתרון בהצגה האחרונה של החישוב הינו באפשרות להזין, כל שנה, מספרים מעודכנים של הנפגעים קשה והנפגעים קל שדווחו ע"י המשטרה ולקבל הערכה מתוקנת של סך הנפגעים קשה.

בהערכות השונות שהודגמו לעיל בפרק זה התקבלו אומדנים שונים לסך מספר הנפגעים קשה, כאשר כל הערכה מהווה קירוב מסוים ומלווה בהנחות והסתייגויות מסוימות. ההסתייגות העיקרית לגבי כל ההערכות המקורבות נמצאת בצורך להוכיח קיום של הנחות שיטת ה-CR. בנוסף, בחלק מהשיטות חסרה התחשבות במרכיבים מסוימים של התופעה כגון: הצורך בתיקון (הפחתה) של מספר הנפגעים קשה שהוגדרו ע"י המשטרה בשל היותם נפגעים קל לפי הטראומה; הצורך להתחשב בבתי חולים נוספים, מעבר ל-17 בתי החולים הכלולים ברישום הטראומה.

להערכה מדויקת יותר של מספר הנפגעים קשה נדרש פיתוח מודלים מסבירים עבור ההסתברות להיות מדווח בטעות ע"י המשטרה כנפגע קשה (מודל ה-False Positive) ועבור ההסתברות להיות מדווח כנפגע קשה ע"י המשטרה בהיותו נפגע קשה לפי הטראומה, כאשר בין המשתנים המסבירים להסתברויות אלה יהיו מאפיינים שונים של הנפגע והתאונה.

## 4. מודל ה-False Positive - ניפוי נפגעים קל בקרב נפגעים קשה של

### המשטרה

השיטה להערכת מספר נפגעים קשה בתנאי הארץ כוללת שני כלים:

- (א) מודל False-Positive המאפשר ניפוי של מספר נפגעים קשה שזוהו ע"י המשטרה ולא אומתו ע"י רישום הטראומה (הנפגעים קל שבטעות זוהו כנפגעים קשה ע"י המשטרה);
- (ב) מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא בקובץ טראומה. מודל זה יאפשר תיקון של תת-הדיווח בנתוני המשטרה - ניפוח מספר הנפגעים קשה שדווחו ע"י המשטרה בעזרת הנתונים מרישום הטראומה.

המודל הראשון - False Positive (FP) מוצג בפרק זה; המודל השני - בפרק 5. פיתוח המודלים התבסס על בחינה פרטנית של מאפייני הנפגעים, הן ע"פ המשטרה והן ע"פ הטראומה. (כדי לסייע בהבנת מאפייני הנפגעים וערכיהם ששימשו בפיתוח המודלים, נספח ב' מציג מבנה של קובץ A - הנתונים המשולבים של המשטרה והטראומה שהתקבלו מהלמ"ס).

כפי שהוצג לעיל (ראה סעיף 3.1), מבחינת נתוני הקובץ המשולב של המשטרה והטראומה (קובץ A, כולל יו"ש) עלה שמתוך 1673 הנפגעים קשה שהמשטרה דיווחה עליהם, 66 היו בעצם נפגעים קל. אלו הם המקרים המכונים כ-False Positive של המשטרה. המשטרה זיהתה נכון 1607 נפגעים קשה בלבד. 1673 נתונים אלה (1607+66) ישמשו לקביעת המודל שעל פיו תנופינה תצפיות מנתוני הנפגעים המזוהים כקשה ע"י המשטרה.

### 4.1. בחינת הנתונים

אחד המשתנים המועמדים לקביעת מודל ה-FP הינו גיל. גיל בעת תאונה מחושב כ-2008 פחות שנת לידה. כאשר משתנה שנת לידה היה חסר, השתמשנו במשתנה גיל שסופק ע"י טראומה. משתנה זה לא קיים עבור קובץ נתוני המשטרה, אולם, השתמשנו בו כדי לא להפסיד תצפיות. המשתנה שיצרנו באופן זה מכונה כ-age\_police.

טבלה 4.1 מציגה סטטיסטיקות של רשומות הנפגעים קשה לפי המשטרה לגבי המשתנה התלוי false\_p (השווה ל-1 עבור המקרים FP ואפס אחרת) ולגבי המסבירים האפשריים. משתנה המגדר (GENDER) מקבל את הערך "0" עבור זכר ו-"1" עבור נקבה.

טבלה 4.1. סטטיסטיקות של רשומות הנפגעים קשה לפי המשטרה, לפיתוח מודל FP

משתנה	מספר מקרים חסרים (missing)	N - מספר מקרים נבדקים	Mean	Minimum	Maximum
false_p	0	1673	0.039	0	1.0
gender	3	1670	0.316	0	1.0
AveratNehiga	898	775	2.985	0	9.0
age_police	0	1673	35.597	0	94.0
GOREM_TEUNA	0	1673	2.240	1.0	9.0
KVUTZAT_OHLUSIYA_LMS	0	1673	1.415	1.0	4.0
PEULAT_NIFGA_LMS	0	1673	3.641	1.0	9.0
SUG_NIFGA_LMS	0	1673	2.455	1.0	9.0
SUG_REHEV_LMS	517	1156	6.176	1.0	19.0
SUG_TEUNA	0	1673	4.045	1.0	19.0
YEHIDA	0	1673	27.293	11.0	61.0
YOM_BASHAVUA	0	1673	4.117	1.0	7.0

מטבלה 4.1 ניתן לראות שלמשתנים עבירת נהיגה (AveratNehiga) וסוג רכב (SUG\_REHEV\_LMS) ישנם ערכים חסרים רבים, מה שמגביל את השימושיות שלהם לצורכי פיתוח המודל.

## 4.2. פיתוח המודל

בפיתוח המודל לזיהוי המקרים של FP היו מספר שלבים:

- (1) הרצת רגרסיה לוגיסטית בצעדים עם רשימת המשתנים ההתחלתית (פרט לשני המשתנים שצוינו לעיל עם ערכים חסרים רבים) - בשלב זה לא נמצא אף משתנה מסביר מובהק;
- (2) שימוש בטכניקות CART ו-GBM לגילוי הקשרים בין המשתנים ולזיהוי ערכי המשתנים החשובים יותר לחיזוי המשתנה המוסבר, מה שהביא ליצירת משתנים מסבירים אגרטיביים;
- (3) הרצת רגרסיה לוגיסטית עם המשתנים המסבירים האגרטיביים - יצירת מודל להערכת ההסתברות ל-FP עבור צירופים שונים של המשתנים המסבירים. פירוט לגבי שיטות CART ו-GBM אשר שימשו במחקר לזיהוי המשתנים המסבירים החשובים ביותר (והצירופים שלהם) עבור המודלים המסבירים מוצג בנספח ג'. בהסתמך על המשתנים שיצאו חשובים ב-GBM ונרמזו כחשובים ב-CART (הופיעו בראשי העצים שנבנו), לפיתוח מודל ה-FP בוצעה חלוקה והגדרה חדשה של המשתנים המסבירים כמוצג להלן:

```
I_SUG_REHEV_LMS = SUG_REHEV_LMS *0;  
if SUG_REHEV_LMS in (4 5 6 7 10 12 14 15 17 19) then I_SUG_REHEV_LMS=1;  
if SUG_REHEV_LMS=. and SUG_NIFGA_LMS in (2 3) then I_SUG_REHEV_LMS=0;  
if SUG_REHEV_LMS=. and SUG_NIFGA_LMS in (1 4 5 6 7 8 9) then I_SUG_REHEV_LMS=1;
```

```
I_SUG_TEUNA = SUG_TEUNA *0;  
if SUG_TEUNA in (1 3 6 9 10 12 15 18 19) then I_SUG_TEUNA=1;
```

```
L_age_police=log(age_police+1);
```

```
I_YOM_BASHAVUA = YOM_BASHAVUA *0;  
if YOM_BASHAVUA in (1 5 6 7) then I_YOM_BASHAVUA=1;
```

```
I_age_police=age_police*0;  
if age_police>33 then I_age_police=1;  
if age_police>74 then I_age_police=2;
```

```
I_YEHIDA = YEHIDA *0;  
if YEHIDA in (34 38 43) then I_YEHIDA=1;
```

כלומר, נוצרו משתנים חדשים: I\_SUG\_REHEV\_LMS, I\_SUG\_TEUNA, I\_YOM\_BASHAVUA, I\_YEHIDA, I\_age\_police, אשר משקפים הקבצות מסוימות של ערכי המשתנים: סוג רכב, סוג תאונה, יום בשבוע, קבוצת גיל הנפגע, יחידת המשטרה, בהתאמה. (במשתנה "סוג רכב למ"ס" היו ערכים חסרים רבים, לכן, ביצירת המשתנה הקטגורי שלו - I\_SUG\_REHEV\_LMS - נעשה שימוש במשתנים חלופיים שהוצעו ע"י CART).

משתנים אלה והאינטראקציות ביניהם שימשו לפיתוח מודל מסביר להסתברות ה-FP (להיות מדווח בטעות כנפגע קשה ע"י המשטרה). המודל פותח בעזרת רגרסיה לוגיסטית בצעדים, עבור המשתנה המוסבר: false\_p=1.



המודל התכנס לפתרון; רמת השונות המוסברת ע"י המודל (Max-rescaled R-Square) היה 10.6%. טבלה 4.2 מציגה את המודל: במודל ההסתברות ל-FP נותרו 4 משתנים מסבירים, לפי הקטגוריות של סוג רכב, סוג תאונה, יום בשבוע, יחידת המשטרה (כל המסבירים מובהקים עם  $p < 0.05$ ).

טבלה 4.2. מודל מסביר להסתברות ה-FP (דיווח שגוי על נפגע קשה ע"י המשטרה)

Analysis of Maximum Likelihood Estimates					
משתנה מסביר	DF	Estimate- מקדם המודל	Standard Error	Wald Chi-Square	Pr > ChiSq*
Intercept	1	-1.9273	0.2002	92.7024	<.0001
I_SUG_REHEV_LMS סוג רכב	1	-1.2314	0.3341	13.5812	0.0002
I_SUG_TEUNA סוג תאונה	1	-0.6511	0.3144	4.2875	0.0384
I_YOM_BASHAVUA יום בשבוע	1	-0.8600	0.2590	11.0300	0.0009
I_YEHIDA יחידת המשטרה	1	-1.5244	0.5985	6.4870	0.0109

\*מובהקות המשתנה

### 4.3. ממצאי המודל

על סמך המודל שפותח, טבלה 4.3 מציגה אומדני ההסתברות ל-FP עבור כל צרוף של המשתנים המסבירים (סה"כ יש 16 מקרים של צירופי המשתנים), לרבות שכיחויות המקרים בקובץ המשולב (מתוך 1673 נפגעים קשה בקובץ המשולב שזוהו ע"י המשטרה). ניתן לראות שיש שילובי המשתנים עבורם צפוי ששיעור ה-FP יהיה גבוהי יחסית - 12.7% - ויש כאלה ששיעור ה-FP הינו אפסי.

לדוגמא, מקרה 1 בטבלה 4.3 מציג צירוף המשתנים בו כל ארבעת המשתנים הקטגוריים שווים ל-0. משמעות מקרה זה הינה:

(1) "סוג רכב למ"ס" ידוע והינו רכב פרטי, או משא עד 3.5 טון, או אופנוע עד 250 סמ"ק, או אוטובוס, או רכב עבודה, או רכבת; או כאשר "סוג רכב למ"ס" לא ידוע ו-"סוג נפגע למ"ס" הינו נהג או נוסע ברכב בעל 4 גלגלים או יותר;

ו- (2) "סוג תאונה" הינו התנגשות חזית-צד, או התנגשות צד-צד, או התנגשות חזית-חזית, או התנגשות עם רכב חונה, או התנגשות עם עצם דומם, או החלקה, או נפילה מרכב נע, או שריפה, או התנגשות אחור אל חזית;

ו- (3) "יום בשבוע" הינו בין שני עד רביעי;

ו- (4) "יחידת המשטרה" כל היחידות פרט ל- 34, 38, 43 (הנגב, לכיש, יהודה).

במקרה 1, ההסתברות לדיווח שגוי ע"י המשטרה תהיה 12.7%, עם רווח סמך 9.0%-17.7%.

מקרה 2 מציג את צירוף המשתנים בו שלושת המשתנים הראשונים שווים ל-0, בעוד שמשתנה יחידת המשטרה שווה ל-1 (ראה טבלה 4.3). משמעות המקרה הינה:

(1) "סוג רכב למ"ס" ידוע והינו רכב פרטי, או משא עד 3.5 טון, או אופנוע עד 250 סמ"ק, או אוטובוס, או רכב עבודה, או רכבת; או כאשר "סוג רכב למ"ס" לא ידוע ו-"סוג נפגע למ"ס" הינו נהג או נוסע ברכב בעל 4 גלגלים או יותר;

ו- (2) "סוג תאונה" הינו התנגשות חזית-צד, או התנגשות צד-צד, או התנגשות חזית-חזית, או התנגשות עם רכב חונה, או התנגשות עם עצם דומם, או החלקה, או נפילה מרכב נע, או שריפה, או התנגשות אחור אל חזית;

ו- (3) "יום בשבוע" הינו בין שני עד רביעי;

ו- (4) "יחידת המשטרה" הינה 34, או 38, או 43 (הנגב, לכיש או יהודה).

במקרה 2, ההסתברות לדיווח שגוי ע"י המשטרה תהיה 3.1%, עם רווח סמך 9.3%-1.0%.

טבלה 4.3. אומדני ההסתברות ל-FP (דיווח שגוי ע"י המשטרה) עבור צירופי המשתנים המסבירים

מס' מקרה	I_SUG_REH_EV_LMS	I_SUG_TEU_NA	I_YOM_B_ASHAVUA	I_YEHID_A	שכיחות המקרה**	הסתברות לדיווח שגוי - p	גבול תחתון*	גבול עליון*
1	0	0	0	0	173	0.127	0.090	0.177
2	0	0	0	1	46	0.031	0.010	0.093
3	0	0	1	0	327	0.058	0.039	0.085
4	0	0	1	1	65	0.013	0.004	0.042
5	0	1	0	0	78	0.071	0.039	0.123
6	0	1	0	1	9	0.016	0.005	0.056
7	0	1	1	0	95	0.031	0.016	0.058
8	0	1	1	1	25	0.007	0.002	0.025
9	1	0	0	0	90	0.041	0.021	0.077
10	1	0	0	1	6	0.009	0.002	0.033
11	1	0	1	0	112	0.018	0.009	0.035
12	1	0	1	1	20	0.004	0.001	0.015
13	1	1	0	0	233	0.022	0.012	0.039
14	1	1	0	1	39	0.005	0.001	0.017
15	1	1	1	0	312	0.009	0.005	0.018
16	1	1	1	1	43	0.002	0.001	0.008

\* ברמת סמך של 95% \*\* מבין רשומות הנפגעים קשה של המשטרה בקובץ המשולב

לשם קבלת אמד למספר המקרים שיש לנכות מקובץ המשטרה, יש לחשב עבור כל אחד מששה עשר השילובים את השכיחות שלו בקובץ, להכפיל שיעור זה באמד המתאים (מעמודה "הסתברות לדיווח שגוי - p" בטבלה 4.3) ולחבר. אם, לדוגמה, נפעיל כללים אלה על רשימות הנפגעים קשה של המשטרה בקובץ המשולב לשנת 2008, נקבל כאמד המסכם את הערך 66.0, עם רווח סמך (7.85; 53.5). אופן הערכת רווחי הסמך לאמד ההסתברות ל-FP מתואר בנספח ד'.

### חישוב אמד לסך הנפגעים קשה לניפוי מקובץ המשטרה, כולל יו"ש

אנו נניח שההסתברויות לדיווח שגוי ע"י המשטרה שחושבו לפי הקובץ המשולב תקפות לגבי כל רשימת הנפגעים קשה שזוהו ע"י המשטרה, בשנת 2008. במקרה זה ניתן לחשב את מספר המקרים השגויים שיש לנפות מסך כל הנפגעים קשה שהמשטרה דיווחה עליהם. הערכה כזאת ניתן לבצע, באופן דומה, כולל או לא כולל מחוזות יו"ש. להלן תוצאות הערכה כולל יו"ש.

זכור, בקובץ המשולב A של המשטרה והטראומה יש 1673 נפגעים קשה לפי דיווח המשטרה. בקובץ המשטרה בלבד (קובץ C) יש 528 נפגעים קשה נוספים - כאלה שהמשטרה דיווחה עליהם ושלא נמצאו בקובץ המשולב. סה"כ, בשנת 2008, המשטרה דיווחה על  $2201=1673+528$  נפגעים קשה. טבלה 4.4 מציגה פילוג של הנפגעים קשה שדווחו ע"י המשטרה, לפי יחידות המשטרה.

טבלה 4.4. פילוג נפגעים קשה שדווחו ע"י המשטרה\* בשנת 2008,

לפי יחידות המשטרה

המשטרה - יחידת	מספר	אחוז
11	300	13.63
12	307	13.95
14	377	17.13
20	294	13.36
33	44	2.00
34	115	5.22
38	103	4.68
41	58	2.64
43	96	4.36
51	168	7.63
52	173	7.86
61	166	7.54
סה"כ	2201	100

\*בקבצי A ו-C ביחד

טבלה 4.5 מציגה, עבור כל צרוף של המשתנים המסבירים - 16 צירופי המשתנים שיצאו ממודל ה-FP, את שכיחויות המקרים בקרב סך הנפגעים קשה שדווחו ע"י המשטרה בשנת 2008 (2201 נפגעים קשה). לשם קבלת אמד מסכם למספר המקרים שיש לנפות מתוך המספר הכולל שדווח ע"י המשטרה, בוצעה הכפלה של השכיחויות (מטבלה 4.5) באמד המתאים של ההסתברות לדיווח שגוי ("ק" מטבלה 4.3), עבור כל אחד מששה עשר השילובים, וחיבור התוצאות לאורך הטבלה. האמד המסכם שהתקבל למספר הנפגעים קשה שהיו שגויים בדיווח המשטרה בשנת 2008 הינו כ-91, עם רווח סמך (117.5 ; 73.5).

טבלה 4.5. שכיחויות המקרים לפי צירופים שונים של המשתנים המסבירים, להערכת הסתברויות ה-FP, בקרב סך הנפגעים קשה שדווחו ע"י המשטרה בשנת 2008

מס' מקרה	I_SUG_REHEV_LMS	I_SUG_TEUNA	I_YOM_BASHAVUA	I_YEHIDA	שכיחות המקרה
1	0	0	0	0	254
2	0	0	0	1	57
3	0	0	1	0	435
4	0	0	1	1	73
5	1	1	0	0	117
6	0	1	0	1	17
7	0	1	1	0	120
8	0	1	1	1	28
9	1	0	0	0	114
10	1	0	0	1	12
11	1	0	1	0	147
12	1	0	1	1	30
13	1	1	0	0	301
14	1	1	0	1	46
15	1	1	1	0	399
16	1	1	1	1	51

## 5. מודל לחישוב ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה -

### לניפוח נפגעים קשה של המשטרה

#### 5.1. כללי

הכלי השני להערכת מספר נפגעים קשה בתנאי הארץ הינו מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא בקובץ טראומה. מודל זה יאפשר תיקון של תת-הדיווח בנתוני המשטרה - ניפוח מספר הנפגעים קשה שדווחו ע"י המשטרה בעזרת הנתונים מרישום הטראומה.

המטרה בסעיף זה היא למצוא את הגורמים המשפיעים על ההסתברות של נפגע קשה הנמצא בקובץ טראומה, להיות מזוהה כנפגע קשה ע"י המשטרה<sup>7</sup>. זאת הסתברות מותנית של תפיסה ע"י המשטרה בתנאי שהייתה תפיסה ע"י טראומה. הסתברות זאת מכונה כ"הסתברות תפיסה מחדש" ע"י המשטרה (recapture probability), ואת הגורמים המשפיעים עליה נמצא באמצעות מודל הקושר בין הגורמים להסתברות. המודל שיבנה יחשב את הסתברות התפיסה מחדש (recapture) של נתוני הטראומה (הנתונים ש"נתפסו בשלב הראשון") ע"י המשטרה ("הלוכד בשלב השני").

מכיוון שהמודל כולל תצפיות הן מקובץ טראומה בלבד (קובץ B) והן מהקובץ המשולב (קובץ A), כאשר בסוף התהליך המודל יופעל על נתוני המשטרה, היה צריך להשתמש במודל רק במשתנים המופיעים הן בקובץ הטראומה והן בקובץ המשטרה. מאידך, מספרם של משתנים אלה מועט ביותר, ויש חשש שההנחה החשובה ביותר - שההסתברות להילכד ע"י המשטרה שווה עבור כל הנפגעים בתוך שכבה - לא תתקיים (ההשלכות של הפרת הנחה זו נדונות בהרחבה בספרות).

הפתרון שהוצע על ידינו למצב זה - חוסר המשתנים להתאמת המודל - היה אלגוריתם דו-שלבי. בשלב הראשון, ניסינו לבדוק אם ניתן לנבא משתנים שעשויים להיות משמעותיים בקביעת ההסתברות ללכידה (חוזרת) ע"י המשטרה, ושנמצאים רק בקובץ המשטרה ולא בקובץ הטראומה, באמצעות המשתנים מקובץ הטראומה. משתנים מלאכותיים אלה שנמצאו בשלב הראשון שימשו לצורך בחירת המודל שעל-פיו נאמוד את הסתברויות הלכידה (החוזרת) ע"י המשטרה. הניבויים בשלב הראשון היו ניתנים לביצוע בעזרת הקובץ המשולב שהיה לנו (קובץ A).

#### 5.2. הנתונים ששימשו לבניית המודל

הנתונים לבניית המודל - ההסתברות של נפגע קשה הנמצא בקובץ טראומה, להיות מזוהה כנפגע קשה ע"י המשטרה - באו משני קבצים: הקובץ המשולב A וקובץ הטראומה בלבד (B).

הבעיה הראשונה שהתעוררה בתהליך המידול היא, שנתון המחוז מופיע בנתוני המשטרה ולא בנתוני הטראומה, ולכן, לדוגמה, לא ניתן לנכות נתונים אלה באופן ישיר מקובץ הטראומה. מכיוון שמספרם של נתונים אלה נמוך, אזי אם הפילוג שלהם ביו"ש אינו שונה מהותית מזה שמחוץ ליו"ש, יהיה ניתן להשתמש בנתוני הטראומה, כולל יו"ש, לצורך בניית המודל בסעיף זה.

<sup>7</sup> הניתוח בשלב זה התייחס לנתונים לאמר ניכוי המקרים של ה- False Positive - ראה פרק 4. כלומר, הנתונים בניתוח לא כללו נפגעים קשה לפי המשטרה שלא נמצאו כנפגעים קשה בטראומה.

טבלה 5.1 מציגה השוואה של התפלגות חומרת הפגיעה בין הנתונים כולל יו"ש להתפלגות שבנתונים ללא יו"ש וביו"ש בלבד. במחזור יו"ש היו סה"כ 204 תצפיות. ניתן לראות שהתפלגות ערכי המשתנה HUMRAT\_PGIA\_LMS דומה בין נתונים עם ובלי יו"ש. נתוני יו"ש בלבד (ההפרש בין הנתונים עם ובלי יו"ש) מתפלגים באופן דומה לנתונים ללא יו"ש. מבחנים פורמאליים מצביעים על שוני מובהק בין ההתפלגות ביו"ש לעומת ההתפלגות מחוץ ליו"ש, אך במספר הנתונים הנבדק אין משמעות למובהקות, מאחר וכל סטייה קטנה בהתפלגות הנתונים בין שתי הקבוצות (יו"ש מול ללא יו"ש) תצא בדרך כלל מובהקת. לכן, לצורך בניית המודל הוחלט להשתמש בכל נתוני הנפגעים קשה שדווחו ע"י רישום הטרומה, **כולל יו"ש**.

טבלה 5.1. השוואה של התפלגות חומרת הפגיעה בין הנתונים כולל יו"ש, ללא יו"ש ו-יו"ש בלבד (בקובץ המשולב A)

נתוני יו"ש בלבד	הנתונים ללא יו"ש	הנתונים כולל יו"ש	מספר אחוז התפלגות HUMRAT_PGIA_LMS:
150 73.5	3167 61.63	3317 62.08	(ריק) במקרים שהחומרה זהה במשטרה ובטראומה השדה כאן יהיה ריק
47 23.0	1891 36.80	1938 36.27	2 - אם במשטרה היה רשום 3 (נפגע קל) ובטראומה היה רשום 2 (נפגע קשה) החומרה הסופית היא 2 (נפגע קשה)
7 3.4	65 1.26	72 1.35	3 - אם במשטרה היה רשום 2 (נפגע קשה) ובטראומה היה רשום 3 (נפגע קל) החומרה הסופית היא 3 (נפגע קל)
0	14 0.27	14 0.26	5 - במקרה של פטירה מאוחרת לפי המשטרה וגם טראומה
0	2 0.04	2 0.04	7 - במקרים שבמשטרה היה 3 (נפגע קל) ובטראומה היה 2 (נפגע קשה) וגם היה פטירה מאוחרת, או שבמשטרה היה הרוג ובטראומה היה פטירה מאוחרת, יקבל 7
204 100	5139 100	5343 100	סה"כ

כאמור, הנתונים לבניית מודל לחישוב ההסתברות להיות מזוהה כנפגע קשה, בהינתן שהנפגע קשה נמצא בקובץ טראומה, באו משני קבצים: הקובץ המשולב A וקובץ טראומה בלבד B.

טבלה 5.2 מציגה פילוג של נתוני הקובץ המשולב A לפי חומרת הפגיעה לפי הטרומה (PazuaKasheTrauma) ומצב הסיווג בעקבות הצלבת הנתונים של המשטרה והטרומה (HUMRAT\_PGIA\_LMS). נפגעים קשה לפי הטרומה (PazuaKasheTrauma=1) הינם האוכלוסייה הרלוונטית לפיתוח המודל אשר כוללת:

- 1607 נפגעים קשה שהמשטרה זיהתה נכון;
- 1938 נפגעים קשה שלא היו במשטרה אך הופיעו בטראומה;
- בנוסף, קיימים 14+2 מקרים של הנפגעים שדווחו ע"י המשטרה כהרוגים, אולם הוגדרו בסופו של דבר כנפגעים קשה בשל פטירה מאוחרת. מקרים אלה הוצאו מהניתוח מתוך הנחה שתקונים אלה (של מספר הרוגים) יבוצעו לאחר הפעלת האלגוריתמים המתוארים בדו"ח זה.

סה"כ מהקובץ המשולב A נשתמש לפיתוח המודל ב-**3545** (1607+1938) רשומות. מבין 1938 תצפיות של נפגעים קשה שלא היו במשטרה ונמצאו בטראומה, 933 דווחו כנפגעים קל בתיקי "ת"ד",

993 דווחו בתיקי "כללי עם נפגעים", לגבי 12 הנוספים במשטרה לא הייתה הערכה של חומרת הפגיעה.

טבלה 5.2. פילוג של נתוני הקובץ המשולב A לפי חומרת הפגיעה לפי הטראומה ומצב הסיווג בעקבות בדיקת הנתונים

PazuaKasheTrauma - חומרת הפגיעה לפי הטראומה			HUMRAT_PGIA_LMS - סיווג הנתונים בעקבות הבדיקה מספר אחוז
סה"כ	1 – נפגע קשה	0 - אחר	
3317	*1607	1710	(ריק) במקרים שהחומרה זהה במשטרה ובטראומה השדה כאן יהיה ריק
62.08	30.08	32.00	
1938	**1938	0	2 - אם במשטרה היה רשום 3 (נפגע קל) ובטראומה היה רשום 2 (נפגע קשה) החומרה הסופית היא 2 (נפגע קשה)
36.27	36.27	0.00	
72	0	72	3 - אם במשטרה היה רשום 2 (נפגע קשה) ובטראומה היה רשום 3 (נפגע קל) החומרה הסופית היא 3 (נפגע קל)
1.35	0.00	1.35	
14	***14	0	5 - במקרה של פטירה מאוחרת לפי המשטרה וגם טראומה
0.26	0.26	0.00	
2	**2	0	7 - במקרים שבמשטרה היה 3 (נפגע קל) ובטראומה היה 2 (נפגע קשה) וגם היה פטירה מאוחרת, או שבמשטרה היה הרוג ובטראומה היה פטירה מאוחרת, יקבל 7
0.04	0.04	0.00	
5343	3561	1782	
100.00	66.65	33.35	סה"כ

\*נפגעים קשה שהמשטרה זיהתה נכון

\*\*נפגעים קשה ע"פ הטראומה שהמשטרה לא זיהתה

\*\*\* הנפגעים שדווחו ע"י המשטרה כהרוגים, אולם הוגדרו בסופו של דבר כנפגעים קשה בשל פטירה מאוחרת (לא ייכללו בנייתוח)

בקובץ טראומה בלבד (B) היו 1608 תצפיות שהתאימו להגדרה האדמיניסטרטיבית של "נפגע קשה" - מאושפז ליותר מ-24 שעות (לפי הפרש ימים בין תאריך שחרור לתאריך אשפוז). מכאן, לבניית המודל ישמשו סה"כ 5153 תצפיות (1608+3545) משני הקבצים.

### 5.3. בחינת משתנים מסבירים

כאמור, בשלב הראשון של פיתוח המודל, נבחנו קשרים של מאפייני הנפגעים קשה בטראומה לעומת מאפייני הנפגעים במשטרה, במטרה לאתר משתנים שעשויים להיות משמעותיים בקביעת ההסתברות ללכידה ע"י המשטרה, באמצעות המשתנים מקובץ הטראומה, כאשר המשתנים באים במקור מקובץ המשטרה. לצורך כך, נבחנו קשרים בין מאפייני הנפגעים קשה, הבאים משני המקורות בקובץ המשולב. סה"כ נבחנו 9 קבוצות של משתנים המוצגות בטבלה 5.3, וביניהן:

גיל הנפגע; מגדר הנפגע; מיקום האשפוז - בית חולים (לעומת יחידת המשטרה); חודש פגיעה; סוג אתר תאונה (סוג דרך); מיקום התאונה - ישוב; סוג נפגע; סוג תאונה; סוג רכב בו נסע הנפגע.

מבין המשתנים הנ"ל, יש משתנים משותפים לשני המקורות (המשטרה והטראומה) כגון: גיל ומגדר של הנפגע; חודש תאונה; בית חולים. ביתר המשתנים קיים שוני ברישום מאפייני הנפגעים, בשני המקורות.

טבלה 5.3. משתנים מקובץ טראומה ומקובץ המשטרה לבחינה הקשרים ביניהם לצורך ניבוי ההסתברות להיות מדווח כנפגע קשה ע"י המשטרה

מס'	שם משתנה בקובץ טראומה	קטגוריות*	משמעות המשתנה	משתנה למ"ס (המשטרה)
<b>(א) משתנים ישירים מקובץ טראומה</b>				
1	age	0-14,15-24,25-54,55-64, 65+	קבוצת גיל	SHNAT_LEDA
2	Sex_cod	1-זכר, 2-נקבה, 9-לא ידוע	מין	MIN
3	Hospital_cod	17 שמות + אחר (לא ברישום טראומה)	בית חולים	YEHIDA, KOD BET_HOLIM_LMS
4	injury_month	1-12	חודש פציעה	HODESH_TEUNA
<b>(ב) משתני טראומה שדרשו ניבוי על סמך משתני המשטרה</b>				
5	Site_sub_type + Site_sec_sub_type	1-בינעירוני, 2-עירוני, 3-עפר, 4-מדרכה, 5-מסילה, 6-תחנה, 88-אחר 1-כביש, 2-צומת, 3-מחלף, 88-אחר	תת סוג אתר + תת סוג אתר משני	SUG_DEREH 4 מצבים: צומת עירוני, קטע עירוני, צומת לא עירוני, קטע לא עירוני
6	Injured_city_cod	--	מיקום תאונה-ישוב	SEMEL_YISHUV
7	Ecode_sec_sub_type	סוגים אפשריים לבדיקה: 1,2,3,4,5,7, אחר	סוג נפגע	SUG_NIFGA_LMS
8	Car_accident_type1- Car_accident_type3	סוגים אפשריים לבדיקה: 1, 2-4 ביחד, 5, 8-11 ביחד, אחר	סוג תאונה	SUG_TEUNA
9	Injury_car_type_cod	סוגים אפשריים לבדיקה: פרטי- מסחרי (11-13 ביחד, 19), משא (14), אופנוע (15), אוטובוס (16, 18), אופניים אחר (22)	סוג רכב נפגע	SUG_REHEV_LMS

\* ראה משמעותות לקטגוריות בנספח ב'

#### א. מגדר הנפגע

למגדר הנפגע קיים קידוד שונה בשני המקורות: "1" ו-"2" במשתנה Sex\_cod בקובץ טראומה; "ז" ו-"נ" במשתנה MIN בקובץ המשטרה. מבין 3545 מקרים של הקובץ המשולב, ב-3513 (2290 גברים, 1223 נשים) הייתה הסכמה בין שני המקורות לגבי מגדר הנפגע.

#### ב. גיל הנפגע

עבור נתוני המשטרה גיל הנפגע חושב על סמך המשתנה של שנת לידה (SHNAT\_LEDA) ונבנו חמש קטגוריות של קבוצות גיל: 0-14,15-24,25-54,55-64, 65+. במקביל, על סמך המשתנה age של טראומה, נבנו קטגוריות גיל דומות. מהשוואת פילוגי הנפגעים בקובץ המשולב לפי הקטגוריות של קבוצות הגיל משני המקורות, עלה שברוב המקרים (3445 מתוך 3545) הייתה התאמה בין קבוצות הגילים של טראומה למשטרה. 62%-63% מהתצפיות היו מרוכזות בקטגוריות גיל 2-3: בני 15-54. כמו כן, התפלגויות דומות של הנפגעים לפי קבוצות הגיל נצפו בין סוגי הקבצים השונים: הקובץ המשולב לעומת טראומה בלבד והמשטרה בלבד.

#### ג. חודש תאונה

נבדקה עקביות המשתנים "חודש תאונה": injury\_month של טראומה לעומת HODESH\_TEUNA של המשטרה. נמצא שפרט למקרה אחד, הייתה הסכמה בין חודש תאונה לפי הטראומה ולפי המשטרה. כמו כן, נמצאה התפלגות דומה של הנפגעים על פני החודשים בקבצים השונים.

## ד. בית חולים

כשמשמשים במשתנה זה יש לשים לב לנקודות הבאות:

- סימון בתי החולים בקובץ נתוני המשטרה בלבד שונה מסימון בתי החולים בטראומה;

- מכיוון שלא כל בתי החולים שייכים לרישום הטראומה, ישנם בתי חולים המופיעים רק בקובץ המשטרה אך לא בקובץ הטראומה. נקודה זאת הינה בעייתית ביותר. הסיבה לכך היא, שיש בתי חולים המיצגים תצפיות עם הסתברות לכידה שונה מהשאר. כאשר בית חולים נמנה על אחד מ-17 בתי החולים ברישום הטראומה המוצגים לקובץ המשולב<sup>8</sup>, ניתן לנסות ולהעריך הסתברות זאת במהלך בניית המודל. כאשר בית חולים, המופיע בקובץ "משטרה בלבד" אינו נמנה על אחד מבתי חולים אלה, לא ניתן באופן ישיר להעריך הסתברות זאת. (מבין 528 מקרים בקובץ המשטרה בלבד, 261 או 49% היו שייכים ליתר בתי החולים). בעיה זו ניתן לפתור במספר דרכים:

\* אפשר לא לנפח תצפיות אלה, אם כי משמעות הדבר היא שאנו מניחים שהסתברות הלכידה היא 100%;

\* אפשר, לדוגמה, להניח שהסתברות הלכידה עבור תצפיות אלה הינה הסתברות הלכידה הממוצעת על פני קטגוריות בתי החולים שיווצרו;

\* אנו ננסה לבדוק אפשרות שימוש במשתנה חלופי למשתנה בית חולים - יחידת המשטרה. הבעיה היא שמשתנה זה חסר עבור הרשומות בקובץ טראומה בלבד. נבדוק אפשרות ניבוי של יחידת המשטרה על-פי בית חולים, וכן את הקשר ההפוך, ניבוי בית חולים על-פי יחידת המשטרה.

- בנוסף, נתגלו טעויות בודדות שתוקנו בשלב הכנת הנתונים. לדוגמה, בקובץ המשולב שהתקבל מהלמ"ס בקידוד ב"ח לפי סמל טראומה: הקידוד "1602" שאינו קיים ברשימת הטראומה, שונה ל-"1401", שהוא הקידוד המתאים לסמל המשטרה "1412".

בקובץ המשולב מופיע מידע לגבי בית חולים הן מקובץ המשטרה (KOD\_BET\_HOLIM\_LMS) והן מקובץ טראומה (Hospital\_cod). נבדקה מהימנות סימון בית חולים על-פי שני המקורות. ע"פ התוצאות שהתקבלו:

- היו 198 תצפיות עבורן לא ידוע קוד בית חולים על-פי המשטרה; תצפיות אלה היוו 5% מסך התצפיות בקובץ המשולב. אולם, עבור תצפיות אלה קיים זיהוי בית חולים על-פי טראומה;

- לגבי התצפיות שעבורן קיים קוד בית חולים על-פי המשטרה, נמצאה הסכמה בזיהוי בית חולים בין דיווח המשטרה והטראומה ב-93% מהמקרים, וחוסר הסכמה ב-238 מיקרים (7% הנותרים).

המסקנה היא שיש התאמה די טובה, אם כי לא מושלמת, בין נתוני המשטרה והטראומה בנושא זה. ממקור הטראומה המידע לגבי בתי החולים היה שלם. בנתוני המשטרה ישנו חוסר במידע, ואף סתירות לנתוני הטראומה, אם כי בהיקף קטן.

<sup>8</sup> ראה את רשימת בתי החולים בנספח ב': משתנה "סמל בית חולים"



בהמשך, נערכה בחינה של "מיקום" - יחידת המשטרה - עבור 198 התצפיות ללא סימון בית חולים לפי דיווח המשטרה. נמצא כי מרבית הבעיות (65% מהנפגעים ללא רישום בית חולים ע"י המשטרה) הינן מדיווחי יחידה 12 (מרחב גליל), כאשר בתי החולים באזור שלה הם נהרייה (קוד 1107), ורבקה זיו (קוד 1105). מאידך, מבחינת נתוני המשטרה בלבד - קובץ C - עלה כי ליחידה 12 אין ייצוג בולט מבין הרשומות ללא קוד בית חולים.

כמו כן, נערכה בחינת המשך של 238 התצפיות שעבורן לא הייתה התאמה בין קוד בית חולים על-פי המשטרה והטראומה. בוצעה השוואה של התפלגות יחידות המשטרה ב-238 מקרים אלה להתפלגות יחידות המשטרה בקרב המקרים בהם שני המקורות מסכימים (3109 מקרים<sup>9</sup>). ההתפלגויות היו דומות, אולם, בשל גודל המדגם, התקבל הבדל מובהק. לכן, ניתן לראות שכאשר ידוע בית החולים למשטרה, נכונות הדיווח אמנם קשורה סטטיסטית ליחידה המדווחת, אבל פרקטית אין קשר, כלומר, היחידות טועות במספר מקרים שבקירוב פרופורציונאלי למספר המקרים בו הן טיפלו. סה"כ, מבין רישומי המשטרה בקבצים C-A ביחד, קוד בית החולים היה ידוע ב-95% מהמקרים.

כדי להתמודד עם בעיית הדיווח על הנפגעים שמחוץ לרישום הטראומה הלאומי, נבחנה אפשרות ניבוי של יחידת המשטרה על-פי בית חולים, ולהפך. בנתוני המשטרה המשתנה "יחידה" קיים עבור כל התצפיות, כאשר עבור נתוני טראומה בלבד משנתה זה חסר. על סמך הקובץ המשולב נערכה הצלבה בין יחידות המשטרה ובתי החולים בהם אושפזו הנפגעים אשר אפשרה, עבור כל בית חולים, להצביע על יחידה שאליה קשורות מרבית ההפניות לבית החולים.

טבלה 5.4 מציגה את רשימת בתי החולים והיחידות הקשורות אליהם. ניתן לראות שיש יחידות הקשורות למספר בתי חולים.

טבלה 5.4. רשימת בתי החולים ויחידות המשטרה הקשורות אליהם,

ע"פ קובץ משולב של נפגעים קשה בשנת 2008

YEHIDA יחידת המשטרה	Hospital_cod קוד בית חולים	מס'
20	1101	1
11	1102	2
52	1103	3
20	1104	4
12	1105	5
11	1106	6
12	1107	7
38	1108	8
14	1109	9
20	1201	10
51	1301	11
34	1302	12
52	1304	13
61	1401	14
14	1501	15
14	1503	16
61	1601	17

<sup>9</sup> 3109 מקרים "עם הסכמה" + 238 מקרים "ללא הסכמה" + 198 מקרים "ללא סימון בית חולים לפי דיווח המשטרה" = 3545 רשומות בקובץ A המשולב.

על סמך הקשרים המוצגים בטבלה 5.4 נוצר משתנה YEHIDA עבור קובץ הטראומה (דהיינו, נרשמה YEHIDA על סמך Hospital\_cod) ובהמשך, נבדקה יעילות משתנה זה בקביעת מודל להערכת הסתברות הלכידה ע"י המשטרה.

### ה. סוג אתר תאונה - סוג דרך

בהמשך, ניסינו לבדוק האם ניתן לנבא משתנים שעשויים להיות משמעותיים בקביעת ההסתברות ללכידה ע"י המשטרה, ושנמצאים רק בקובץ המשטרה ולא בקובץ טראומה, באמצעות המשתנים מקובץ טראומה. משתנים אלה הם: סוג אתר תאונה (סוג דרך); מיקום התאונה - ישוב; סוג נפגע; סוג תאונה; סוג רכב בו היה הנפגע.

לזיהוי אתר התאונה נבחנה אפשרות בניית משנתה SUG\_DEREH מקובץ המשטרה ע"י שני משתנים מקובץ טראומה: Site\_sub\_type ו-Site\_sec\_sub\_type<sup>10</sup>. כצפוי, השדה "סוג דרך" היה מלא בכל נתוני המשטרה. מבחינת התפלגות משתני האתר בקבצי הטראומה עלה כי:

- השדה Site\_sub\_type סובל מחוסרים בשני הקבצים - נמצאו כ-17% ערכים חסרים או לא ידועים בקובץ המשולב (קובץ A) ו-27% בנתוני טראומה בלבד (קובץ B);

- גם השדה Site\_sec\_sub\_type סובל מחוסרים בשני הקבצים - כ-28% בקובץ המשולב ו-51% בקובץ טראומה בלבד;

- כמו כן, בבחינת שכיחות משותפת של שני המשתנים Site\_sub\_type ו-Site\_sec\_sub\_type, בשני קבצי הטראומה, נמצא כי אין אף מקרה בו Site\_sub\_type היה חסר ו-Site\_sec\_sub\_type לא היה חסר.

לניבוי המשתנה SUG\_DEREH של המשטרה, באמצעות משתנה Site\_sub\_type של טראומה, שימשה תוכנת CART. נמצא שעל סמך הנתונים בקובץ המשולב ניתן ליצור הבחנה בין "סוג דרך" 1+2 ל-3+4, אך לא בין 1 ל-2 או בין 3 ל-4. לכן, נבנה משתנה חדש של "סוג דרך" אשר קיבל את הערך "0", כאשר סוג הדרך הוא "1" או "2", שמשמעותו דרך עירונית, ואת הערך "1", כאשר סוג הדרך הוא "3" או "4", שמשמעותו דרך לא-עירונית. בהרצה נוספת של CART התקבלה חלוקה ברורה של הנתונים לפי סוג דרך חדש (עירוני, לא-עירוני) בהתאם לערכים של המשתנה Site\_sub\_type. בין היתר, התקבל שכאשר Site\_sub\_type הינו כביש עירוני, כביש אחר/דרך עפר, מדרכה ועוד, יסווג סוג הדרך כ"דרך עירונית" (0). לעומת זאת, כאשר Site\_sub\_type=1 (כביש מהיר/ בין עירוני) יסווג סוג הדרך כ"דרך בין-עירונית" (1).

לכן, עבור קובץ טראומה בלבד נבנה משתנה sug\_dereh\_trauma, שיקבל את הערך "0" עבור דרך עירונית ו-"1" עבור דרך בינעירונית, בהתאם לכללים אלה:

```
sug_dereh_trauma=0*Site_sub_type;  
if Site_sub_type=1 then sug_dereh_trauma=1;  
if Site_sub_type in (2 3 4 5 6 7 88 99) then sug_dereh_trauma=0;
```

לבחינת דיוק הניבוי של המשתנה החדש, על סמך הקובץ המשולב, בוצעה השוואה בין זיהוי סוג דרך בעזרת המשתנה sug\_dereh\_trauma לעומת המשתנה "סוג דרך" שבא מהמשטרה, ונמצאה

<sup>10</sup> ראה ערכי המשתנים בנספח ב'

התאמה טובה. עם זאת, בקובץ המשולב נעדיף להשתמש ב"סוג דרך" כפי שהוא מוגדר ע"י המשטרה, כמועמד למודל ההסתברות, בשל אמינותו. לכן, כאשר נמדל את הסתברות הלכידה, נשתמש במשתנה "סוג דרך" כפי שדווח ע"י המשטרה עבור הקובץ המשולב, ובמשתנה sug\_dereh\_trauma עבור נתונים מקובץ טראומה בלבד. שני המשתנים יכונה sug\_derehA.

#### ו. מיקום תאונה - ישוב

נבדקה אפשרות של בנית משתנה semel\_yishuv של המשטרה על סמך המשתנה Injured\_city\_cod של הטראומה. נמצא כי משימה זו אינה אפשרית עקב חוסר עקביות של משתנה הטראומה. בין היתר, נמצא כי:

- בבדיקת סוג דרך (עירוני/ בינעירוני) כנגד סמל ישוב ע"פ נתוני הטראומה עלה שב-90% מהמקרים של דרך עירונית סמל ישוב היה ידוע, אך גם ב-95% מהמקרים בהם סוג הדרך היה בינעירוני, סמל ישוב היה ידוע, ולא כפי שצריך להיות בדרך בינעירונית - סמל לא ידוע.

- גם כאשר סוג דרך נקבע ע"פ נתוני המשטרה (האמינים לפי השערותנו) כנגד סמל ישוב ע"פ נתוני הטראומה, עלה שב-92% מהמקרים שמסומן סוג דרך עירוני סמל הישוב היה ידוע, אך גם ב-88% מהמקרים בהם סוג הדרך היה בינעירוני, סמל ישוב היה ידוע. ברור מכאן ששדה סמל ישוב ע"פ טראומה לא מהימן.

#### ז. סוג תאונה

נבחנו אפשרות בניית משתנה SUG\_TEUNA של המשטרה על סמך המשתנים מקובץ טראומה: Car\_accident\_type1-Car\_accident\_type3. טבלה 5.5 מציגה קידוד של סוגי התאונות על-פי המשטרה והטראומה. בוצעה חלוקה של סוגי התאונות למספר קבוצות מוכללות כפי שמקובל בניתוחי תאונות (פגיעה בהולך רגל; התנגשויות בין כלי רכב; התנגשות חזיתית; תאונות רכב יחיד; אחר) - ראה עמודה "קידוד משותף אגרטיבי" בטבלה 5.5. עבור קידוד המשטרה, המשתנה האגרטיבי נקרא Sug\_p, ואילו עבור קידוד הטראומה המשתנה האגרטיבי של סוג תאונה נקרא Sug\_t1-Sug\_t3.

בבחינת התפלגות הרשומות בקובץ המשולב לפי המשתנה Sug\_p של המשטרה לעומת המשתנה Sug\_t1 (סיווג ראשון) של טראומה, נמצאה הסכמה גבוהה בין שני הסיווגים לגבי קטגוריה 1 "הולכי רגל" (93%-97%) אך הסכמה בינונית עד נמוכה לגבי שאר הקטגוריות. (כמו כן, בדיקת הנתונים העלתה שלגבי רוב התצפיות, סוג תאונה שני ושלישי על-פי טראומה היה חסר - 94% ו-99%, בהתאמה).

בעזרת תוכנת RPART נבנה מודל לחיזוי סוג תאונה על-פי המשטרה (Sug\_p) בעזרת סוגי תאונה על-פי טראומה (Sug\_t1-Sug\_t3). נמצא שקיימת אפשרות ניבוי של סוג תאונה על-פי המשטרה בעזרת משתנה Sug\_t1 של טראומה, כאשר מבוצע איחוד של סוגי תאונות אגרטיביים 2, 3, 5 לסוג מוכלל אחד. כלומר, לפיתוח מודל הסתברות הלכידה ע"י המשטרה ניתן להיעזר בשלוש קטגוריות מוכללות של סוגי התאונות שהם:

"1" - פגיעה בהולך רגל; "235" - התנגשויות בין כלי רכב ואחר; "4" - תאונות רכב יחיד.

טבלה 5.5. קידוד של סוגי התאונות על-פי המשטרה והטראומה, לרבות משתנים אגרטיביים שנבנו למחקר

קידוד טראומה: Car_accident_type1-Car_accident_type3		קידוד משטרה: SUG_TEUNA		קידוד משותף אגרטיבי: Sug_p, Sug_t1-Sug_t3	
פגיעה בהולך רגל	1	פגיעה בהולך רגל	1		1
התנגשות חזית אל צד	2	התנגשות חזית אל צד	2		2
התנגשות חזית אל אחור	3	התנגשות חזית באחור	3		2
התנגשות צד אל צד	4	התנגשות צד בצד	4		2
התנגשות חזית אל חזית	5	התנגשות חזית אל חזית	5		3
התנגשות עם רכב שעצר	6	התנגשות עם רכב שנעצר ללא חניה	6		4
התנגשות עם רכב חונה	7	התנגשות עם רכב חונה	7		4
התנגשות בעצם קבוע	8	התנגשות עם עצם דומם	8		4
ירידה מהדרך	9	ירידה מהכביש או עלייה למדרכה	9		4
התהפכות	10	התהפכות	10		4
החלקה	11	החלקה	11		4
נפילה מרכב נוסע	12	פגיעה בנוסע בתוך כלי רכב	12		5
נהג נרדם	13	נפילה ברכב נע	13		5
דיבר בטלפון נייד	14	שריפה	14		5
עצירת פתע	15	אחר	15		5
אחר	98	התנגשות אחור אל חזית	17		5
לא ידוע	99	התנגשות אחור אל צד	18		5
		התנגשות עם בעל חיים	19		5
		פגיעה ממתען של רכב	20		5

שלוש הקטגוריות המוכללות של סוגי התאונות על סמך קידוד המשטרה נוצר משתנה Sug\_p3 ועל סמך קידוד הטראומה נוצר משתנה Sug\_t1\_3. טבלה 5.6 מציגה התפלגות של רשומות הקובץ המשולב לפי משתנה Sug\_p3 של המשטרה לעומת משתנה Sug\_t1\_3 של הטראומה. ניתן לראות שבעקבות הגדרת שלוש הקטגוריות המוכללות של סוגי התאונות, התקבלה התאמה די טובה בין סיווג התאונות לפי הטראומה והמשטרה.

טבלה 5.6. התפלגות של רשומות הקובץ המשולב לפי משתנה Sug\_p3 של המשטרה לעומת משתנה Sug\_t1\_3 של הטראומה - הקטגוריות המוכללות של סוגי התאונות

מספר אחוז מהסה"כ אחוז בשורה אחוז בעמודה	Sug_t1_3 - קטגוריות מוכללות לפי הטראומה				
	(חסר)	1	4	235	סה"כ
1	0 0.00 0.00 0.00	1011 28.52 <b>96.93</b> 93.35	5 0.14 0.48 0.70	27 0.76 2.59 1.55	1043 29.42
4	1 0.03 0.13 25.00	14 0.39 1.77 1.29	538 15.18 <b>67.84</b> 75.35	240 6.77 30.26 13.76	793 22.37
235	3 0.08 0.18 75.00	58 1.64 3.39 5.36	171 4.82 10.01 23.95	1477 41.66 <b>86.42</b> 84.69	1709 48.21
סה"כ	4 0.11	1083 30.55	714 20.14	1744 49.20	3545 100.0

## ח. סוג נפגע

נבחנה אפשרות בניית משתנה `SUG_NIFGA_LMS` (סוג נפגע) של המשטרה ע"פ המשתנה `Ecode_sec_sub_type` של טראומה. טבלה 5.7 מציגה קידוד של סוגי הנפגעים על-פי המשטרה והטראומה. כשלב ראשון, בוצע איחוד של הסוגים ה"אחרים" של הנפגעים: 7-9 עבור `SUG_NIFGA_LMS`; 6 ו-8-11 עבור `Ecode_sec_sub_type`. לפי השכיחויות בקובץ המשולב אלה הם מקרים נדירים וחלקם אף לא מופיעים בקובץ המשולב (כגון: 8-10 של הטראומה). לאחר איחוד הקטגוריות, על סמך המשתנה מקובץ טראומה נוצר משתנה `SUG_NIFGA_T`, בעוד שעל סמך המשתנה מקובץ המשטרה נוצר משתנה `SUG_NIFGA_P` (לכל אחד מהמשתנים היו שבעה ערכים).

טבלה 5.7. קידוד של סוגי הנפגעים על-פי המשטרה והטראומה\*

סוגי הנפגעים לפי טראומה	Ecode_sec_sub_type	סוגי הנפגעים לפי למ"ס (המשטרה)	SUG_NIFGA_LMS
נהג רכב	1	הולך רגל	1
נוסע רכב	2	נהג-רכב בעל 4+ גלגלים	2
נהג אופנוע	3	נוסע-רכב בעל 4+ גלגלים	3
נוסע אופנוע	4	נהג-אופנוע	4
רוכב אופניים	5	נוסע-אופנוע (לא נהג)	5
רוכב/נוסע חיות	6	נהג-אופניים	6
הולך רגל	7	נוסע-אופניים (לא נהג)	7
רכבת- ללא רכב ממנוע	8	נהג-רכב לא ידוע	8
ים	9	נוסע-רכב לא ידוע	9
אוויר	10		
תחבורה-אחר/לא ידוע	11		

\* התאמה בין סוגי הנפגעים משני המקורות מודגשת בצבעים

בעזרת תוכנת `RPART` נבנה מודל לחיזוי סוג נפגע על-פי המשטרה בעזרת סוג נפגע על-פי טראומה. בבחינת פילוג הרשומות בקובץ המשולב לפי שני המשתנים של סוג נפגע נתגלה ערבוב בין סוגי הנפגעים 4 ו-5 (נהג אופנוע ונוסע אופנוע). לאחר איחוד קטגוריות 4 ו-5 לקטגוריה אחת "45" התקבלה התאמה טובה בין פילוגי הנפגעים בקובץ המשולב לפי שני המשתנים של סוג נפגע: סוג נפגע לפי המשטרה וסוג נפגע חזוי לפי הטראומה, כמוצג בטבלה 5.8.

כתוצאה מתהליך זה, הכללים לבניית משתנה "סוג נפגע" של המשטרה ע"פ סוג נפגע של טראומה נקבעו כלהלן:

```
Sug_nifga_byT_F=.;*final variable for predict police sug nifga by trauma;
if Ecode_sec_sub_type=1 then Sug_nifga_byT_F= 2;
if Ecode_sec_sub_type=2 then Sug_nifga_byT_F= 3;
if Ecode_sec_sub_type in (3 4) then Sug_nifga_byT_F= 45;
if Ecode_sec_sub_type=5 then Sug_nifga_byT_F= 6;
if Ecode_sec_sub_type=7 then Sug_nifga_byT_F= 1;
if Ecode_sec_sub_type in (6 8 9 10 11) then Sug_nifga_byT_F = 2;
```

## ט. סוג רכב בו נסע הנפגע

נבחנה אפשרות בניית משתנה `SUG_REHEV_LMS` (סוג רכב בו נסע הנפגע) של המשטרה ע"פ המשתנה `Injury_car_type_cod` של טראומה. טבלה 5.9 מציגה קידוד של סוגי רכב על-פי המשטרה והטראומה.

טבלה 5.8. פילוג הנפגעים בקובץ המשולב לפי "סוג נפגע" של המשטרה וסוג נפגע חזוי לפי הטראומה

- SUG_NIFGA_P45 סוג נפגע לפי המשטרה	Sug_nifga_byT45 - סוג נפגע חזוי לפי הטראומה					
מספר אחוז מהסה"כ אחוז בשורה אחוז בעמודה	1	2	3	6	45	Total
(חסר)	0 0.00 0.00 0.00	9 0.25 75.00 0.93	1 0.03 8.33 0.14	1 0.03 8.33 0.62	1 0.03 8.33 0.16	12 0.34
1	1021 28.80 <b>97.24</b> 95.07	16 0.45 1.52 1.65	6 0.17 0.57 0.84	5 0.14 0.48 3.09	2 0.06 0.19 0.32	1050 29.62
2	13 0.37 1.52 1.21	800 22.57 <b>93.24</b> 82.73	38 1.07 4.43 5.33	4 0.11 0.47 2.47	3 0.08 0.35 0.48	858 24.20
3	22 0.62 2.70 2.05	126 3.55 15.48 13.03	664 18.73 <b>81.57</b> 93.13	2 0.06 0.25 1.23	0 0.00 0.00 0.00	814 22.96
6	13 0.37 8.18 1.21	5 0.14 3.14 0.52	1 0.03 0.63 0.14	138 3.89 <b>86.79</b> 85.19	2 0.06 1.26 0.32	159 4.49
45	2 0.06 0.32 0.19	5 0.14 0.79 0.52	0 0.00 0.00 0.00	7 0.20 1.11 4.32	619 17.46 <b>97.79</b> 98.41	633 17.86
789	3 0.08 15.79 0.28	6 0.17 31.58 0.62	3 0.08 15.79 0.42	5 0.14 26.32 3.09	2 0.06 10.53 0.32	19 0.54
סה"כ	1074 30.30	967 27.28	713 20.11	162 4.57	629 17.74	3545 100.0

יש לשים לב שבין ערכי המשתנה "סוג רכב" של טראומה ישנו קוד "26" - הולך רגל, בעוד שבנתוני המשטרה שדה זה נשאר ריק במקרה של נפגע הולך רגל; לכן, בוצעה השלמה של שדה זה בנתוני המשטרה (עבור המקרים של sug\_nifga\_lms=1 נרשם SUG\_REHEV\_LMS=26).

כשלב ראשון בבחינה הנוכחית, נערכה חלוקה אגרגטיבית של סוגי הרכב לשמונה קטגוריות, בהתאם למקובל בניתוחי בטיחות, כאשר הקטגוריות המוכללות סומנו בעזרת אותיות כלהלן:

A - רכב פרטי-מסחרי קל;

B - רכב משא;

C - אופנועים;

D - אוטובוסים;

E - אופניים;

F - הולך רגל (עקב הימצאות קוד זה בקובצי הטראומה);

G - מונית;

H - יתר סוגי הרכב (נדירים למדי).

טבלה 5.9. קידוד סוגי הרכב על-פי המשטרה והטראומה, לרבות הקטגוריות המוכללות שנקבעו ותוצאות חיזוי משתנה המשטרה ע"פ משתנה הטראומה (בקובץ המשולב)

תוצאות חיזוי קטגוריות סוג רכב של המשטרה ע"פ משתנה הטראומה	הגדרת קטגוריות מוכללות של רכב לפי הטראומה	Injury_car_type_cod - ערכי משתנה סוג רכב לפי הטראומה	SUG_REHEV_LMS - ערכי משתנה סוג רכב לפי המשטרה
A	A	11 מכונית פרטית	1 רכב נוסעים פרטי
A	A	12 ג'יפ ורכב שטח 4 X4	2 משא עד 3.5 טון אחד
A	A	13 מיני ואן פרטי	3 משא עד 3.5 טון לא אחד
B	B	14 משאית	4 משא 1.4 עד 9.9 טון משקל כולל
C	C	15 אופנוע	5 משא 10 עד 15.9 טון משקל כולל
D	D	16 אוטובוס	6 משא 16 עד 33.9 טון משקל כולל
G	G	17 מונית	7 משא +34.0 טון משקל כולל
D	D	18 מיניבוס מסחרי	8 אופנוע עד 50 סמ"ק
A	A	19 טרנזיט מסחרי	9 אופנוע 51 עד 250 סמ"ק
H	H	20 טרקטור	10 אופנוע 251 עד 500 סמ"ק
H	H	21 טרקטורון	11 אוטובוס
E	E	22 אופניים	12 מונית
H	H	23 רכבת	13 רכב עבודה
H	H	24 בעל חיים	14 טרקטור
H	H	25 הולך בליידיס, קורקינט	15 אופניים
F	F	26 הולך רגל	16 רכבת
H	H	27 ללא מעורבות של רכב נוסף	17 אחר
H	H	98 אחר	18 אוטובוס זעיר
A	I	99 לא ידוע	19 אופנוע +501 סמ"ק
			*26 הולך רגל

\*השלמת נתון עבור המשטרה

כמו כן, בקובץ טראומה סומנה קטגוריה I - "לא ידוע". חלוקה אגרגטיבית זו בוצעה בקרב ערכי המשתנים הן של המשטרה והן של טראומה - אותיות הקידוד שהתקבלו מוצגות בטבלה 5.9.

בעזרת תוכנת RPART נבנה חיזוי סוג רכב (מוכלל) לפי המשטרה בעזרת סוג רכב (מוכלל) על-פי טראומה, ולהפך. נמצאה התאמה בין הקטגוריות המוכללות של שני המשתנים - ראה עמודה "תוצאות חיזוי קטגוריות סוג רכב של המשטרה ע"פ משתנה הטראומה" בטבלה 5.9. כלומר, נמצאה התאמה ישירה בין שמונה הקטגוריות המוכללות שהוגדרו בשני המשתנים, כאשר קטגוריה I של טראומה נכללה בקטגוריה A של המשטרה.

טבלה 5.10 מציגה, לדוגמא, את פילוגי הנפגעים בקובץ המשולב לפי שני המשתנים של סוג רכב: סוג רכב לפי המשטרה (SUG\_REHEV\_p8) לעומת סוג רכב חזוי לפי הטראומה (SUG\_REHEV\_pred\_8). ניתן לראות שלגבי הקטגוריות העיקריות הניבוי היה מאוד טוב - מראה בין 86% עד 98% הצלחה. הניבוי פחות טוב עבור קטגוריות G, H, אך קטגוריות אלה מהוות פחות מ-4% מהאוכלוסייה, כך שהניבוי לגבי קטגוריות אלה פחות מדאיג.

טבלה 5.10. פילוגי הנפגעים בקובץ המשולב לפי שני המשתנים של סוג רכב: סוג רכב לפי המשטרה (SUG\_REHEV\_p8) לעומת סוג רכב חזוי לפי הטראומה (SUG\_REHEV\_pred\_8)

SUG_REHEV_p8 - סוג רכב לפי המשטרה	SUG_REHEV_pred_8 - סוג רכב חזוי לפי הטראומה								סה"כ
מספר אחוז בשורה אחוז בעמודה	A	B	C	D	E	F	G	H	סה"כ
A	1400 95.04 97.02	10 0.68 35.71	2 0.14 0.32	13 0.88 13.27	6 0.41 3.70	32 2.17 2.98	0 0.00 0.00	10 0.68 11.90	1473
B	2 10.53 0.14	17 89.47 60.71	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	19
C	2 0.32 0.14	0 0.00 0.00	620 97.79 98.57	0 0.00 0.00	7 1.10 4.32	2 0.32 0.19	0 0.00 0.00	3 0.47 3.57	634
D	4 4.40 0.28	0 0.00 0.00	0 0.00 0.00	82 90.11 83.67	1 1.10 0.62	2 2.20 0.19	0 0.00 0.00	2 2.20 2.38	91
E	1 0.60 0.07	0 0.00 0.00	2 1.20 0.32	0 0.00 0.00	143 86.14 88.27	14 8.43 1.30	0 0.00 0.00	6 3.61 7.14	166
F	11 1.08 0.76	0 0.00 0.00	2 0.20 0.32	1 0.10 1.02	4 0.39 2.47	995 97.26 92.64	0 0.00 0.00	10 0.98 11.90	1023
G	19 42.22 1.32	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	0 0.00 0.00	2 4.44 0.19	24 53.33 96.00	0 0.00 0.00	45
H	4 4.35 0.28	1 1.09 3.57	3 3.26 0.48	2 2.17 2.04	1 1.09 0.62	27 29.35 2.51	1 1.09 4.00	53 57.61 63.10	92
סה"כ	1443	28	629	98	162	1074	25	84	*3543

\*ערכים חסרים: 2

## י. סיכום

לסיכום בחינת הקשרים בין מאפייני הנפגעים קשה בטראומה ובמשטרה שנערכה בפרק זה, נספח ה' מציג ריכוז המשתנים ששימשו לבניית המודל להסתברות לכידה ע"י המשטרה.

לבניית מודל הסתברות הלכידה ע"י המשטרה שימשו שמונה משתנים שהם: גיל הנפגע; מגדר הנפגע; יחידת המשטרה (במקום בית חולים); חודש תאונה; סוג אתר תאונה - סוג דרך; סוג תאונה; סוג נפגע, וסוג רכב בו נסע הנפגע. ערכים של משתנים אלה, ע"פ הקטגוריות המוכללות שנקבעו לעיל, ייוחסו לכל אחת מרשומות הנפגעים קשה מקבצי הטראומה - קבצי A ו-B.

## 5.4. בנית מודל להסתברות לכידה ע"י המשטרה

הכלי השני להערכת מספר נפגעים קשה בתנאי הארץ שנבנה במחקר הינו מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא בקובץ טראומה. מודל זה יאפשר תיקון של תת-הדיווח בנתוני המשטרה - ניפוח מספר הנפגעים קשה שדווחו ע"י המשטרה בעזרת הנתונים מרישום הטראומה.

המודל נבנה על סמך הנתונים משני קבצים: הקובץ המשולב (קובץ A) וקובץ טראומה בלבד (קובץ B), אשר כללו ביחד 5153 רשומות של נפגעים קשה (לאחר ניכוי ה-False Positive - נפגעים קל שבטעות זוהו כנפגעים קשה ע"י המשטרה).



כשלב ראשון בפיתוח המודל, נערך זיהוי המשתנים החשובים להערכת הסתברות הלכידה, בעזרת מודלים CART - בנית עצי סיווג, בשלוש רמות ובשתי רמות. לסיווג הרשומות ע"פ ערכי המשתנים שימשו אותן הקטגוריות המוכללות של המשתנים שנקבעו בשלב המקדים של חקר הנתונים - ראה סעיף 5.3.

בין המשתנים החשובים להערכת הסתברות הלכידה ע"י המשטרה נמצאו:

במודל CART 3 רמות - סוג דרך, סוג נפגע, סוג תאונה, סוג רכב, יחידת המשטרה;  
במודל CART 2 רמות - סוג דרך, סוג רכב, יחידת המשטרה.

כמו כן, בעזרת תוכנת RPART הודגמה אפשרות ליצירת 4 קטגוריות של נפגעים קשה, בהתאם להסתברות הלכידה ע"י המשטרה, בעזרת אותם שלושת המאפיינים: סוג דרך, סוג רכב, יחידת המשטרה.

במקביל, תוצאות הסיווג בעזרת מודל GBM בשתי רמות הצביעו על חשיבות מאפיינים אלה: יחידת משטרה, סוג רכב, סוג נפגע, סוג דרך, קבוצת גיל, מגדר הנפגע, סוג תאונה. באף בדיקה מקדימה לא נמצאה חשיבות ל-"חודש תאונה" בהערכת הסתברות הלכידה ע"י המשטרה.

בעזרת רגרסיה לוגיסטית, נבנה מודל לחיזוי הסתברות הלכידה ע"י המשטרה, על סמך 4 קטגוריות של נפגעים קשה, שהוגדרו קודם בעזרת שלושת המאפיינים שנמצאו במודל CART בשתי רמות - סוג דרך, סוג רכב, יחידת המשטרה. המודל ענה לקריטריון התכנסות (convergence criterion) והתאפיין ב-13.9% של שונות מוסברת (Max-rescaled R-Square).

עם זאת, מבחינת תוצאות המודלים שנבנו עד כה היה ניכר שלא נמצאה רשימה סגורה של המשתנים המסבירים להסתברות לכידה ע"י המשטרה שעליה היו מסכימים כל המודלים. לכן, הוחלט לבצע פיתוח נוסף של מודל מסביר, באמצעות רגרסיה לוגיסטית בצעדים ותוך כדי שימוש במאפייני הנפגעים שנמצאו כחשובים ע"י רוב המודלים וכמו כן, תוך כדי שימוש באינטראקציות בין המשתנים שנמצאו קודם, דהיינו: 4 קטגוריות של נפגעים קשה שהוגדרו בעזרת ערכים מסוימים של מאפייני סוג דרך, סוג רכב ויחידת המשטרה.

להלן הגדרות לקטגוריות של נפגעים קשה (I\_cat) ששימשו לפיתוח המודל:

```
if SUG_REHEV_A="" then I_cat=1;
else if SUG_REHEV_A in ("B" "D" "E" "H" ) then I_cat=1;
else if SUG_REHEV_A in ("A" "C" "F" "G" ) and YEHIDA in (12 20 34 38 51 52 61) then
I_cat=2;
else if sug_derehA=0 or (sug_derehA=. and SUG_REHEV_A in ("C" "F" ) ) then I_cat=3;
else I_cat=4;
```

בנוסף, בטרם התאמת המודל המסכם, בוצעה השלמה של ערכים חסרים במאפייני הנפגעים (ראה פירוט בנספח ו'), כאשר במקביל לשמות של כל המשתנים המסבירים נוספה בהתחלה אות "T\_".

המודל המסכם שהותאם לנתונים ענה לקריטריון התכנסות (convergence criterion) והתאפיין ב-18.4% של שונות מוסברת (Max-rescaled R-Square), מה שהיווה שיפור לעומת המודל שנבנה קודם על סמך ארבע הקטגוריות של נפגעים קשה. במודל המסכם נמצאה השפעה מובהקת על הסתברות הלכידה ע"י המשטרה של מאפיינים אלה: מגדר הנפגע וסוג תאונה (עם  $p < 0.05$ ); קבוצת

גיל הנפגע, יחידת המשטרה, סוג דרך, סוג נפגע, קטגוריות נפגע (עם  $p < 0.001$ ). יצוין כי הקטגוריות של נפגעים קשה (I\_cat) משקפות בעיקר את סוג הרכב בו נסע הנפגע.

המודל המסכם שהותאם לנתונים מוצג בטבלה 5.11. מודל זה מומלץ לשימוש להערכת ההסתברות להיות מדווח ע"י המשטרה כנפגע קשה.

טבלה 5.11. מודל מסביר להערכת הסתברות הלכידה ע"י המשטרה

Analysis of Maximum Likelihood Estimates						
Parameter - משתנה מסביר	ערך המשתנה - משמעותו	DF	- Estimate מקדם המודל	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.9661	0.5357	13.4690	0.0002
T_gender - מגדר הנפגע	0 - זכר	1	0.1578	0.0753	4.3838	0.0363
T_gender - מגדר הנפגע	1 - נקבה	0	0	.	.	.
T_age_c - קבוצת גיל הנפגע	1 - בני 0-14	1	-0.4906	0.1266	15.0214	0.0001
T_age_c - קבוצת גיל הנפגע	2 - בני 15-24	1	-0.0554	0.1209	0.2099	0.6468
T_age_c - קבוצת גיל הנפגע	3 - בני 25-54	1	-0.2444	0.1141	4.5922	0.0321
T_age_c - קבוצת גיל הנפגע	4 - בני 55-64	1	0.0154	0.1437	0.0115	0.9148
T_age_c - קבוצת גיל הנפגע	5 - בני +65	0	0	.	.	.
YEHIDA_2 - יחידת משטרה	0 - יתר היחידות	1	0.8212	0.1852	19.6640	<.0001
YEHIDA_2 - יחידת משטרה	1 - אם יחידה 12 20 34 38 51 52 61	0	0	.	.	.
T_sug_derehA - סוג דרך	0 - עירונית	1	-0.8120	0.0872	86.6408	<.0001
T_sug_derehA - סוג דרך	1 - לא עירונית	0	0	.	.	.
T_Sug_p3 - סוג תאונה	1 - פגיעה בהולך רגל	1	0.6508	0.2625	6.1477	0.0132
T_Sug_p3 - סוג תאונה	4 - רכב יחיד	1	0.1025	0.0837	1.5004	0.2206
T_Sug_p3 - סוג תאונה	235 - התנגשויות בין כלי רכב או אחר	0	0	.	.	.
T_Sug_nifga_A - סוג נפגע	1 - הולך רגל	1	-1.9145	0.5582	11.7623	0.0006
T_Sug_nifga_A - סוג נפגע	2 - נהג רכב	1	-2.0417	0.4956	16.9692	<.0001
T_Sug_nifga_A - סוג נפגע	3 - נוסע ברכב	1	-2.0884	0.4976	17.6148	<.0001
T_Sug_nifga_A - סוג נפגע	6 - נהג אופניים	1	-2.1735	0.4933	19.4110	<.0001
T_Sug_nifga_A - סוג נפגע	45 - נהג/נוסע אופנוע	1	-1.6978	0.5035	11.3691	0.0007
T_Sug_nifga_A - סוג נפגע	789 - אחר	0	0	.	.	.
I_cat - קטגורית נפגע*		1	-1.3953	0.2158	41.8088	<.0001
I_cat - קטגורית נפגע		2	-0.4457	0.2242	3.9537	0.0468
I_cat - קטגורית נפגע		3	-0.4934	0.1557	10.0421	0.0015
I_cat - קטגורית נפגע		4	0	.	.	.

\* משמעותיות לקטגוריות נפגע (I\_cat):

- 1 - כאשר סוג רכב B (רכב משא מעל 10 טון), D (אוטובוס), E (אופניים) או H (רכב עבודה, טרקטור, רכבת או אחר);
- 2 - כאשר סוג רכב A (רכב פרטי, או מסחרי, או משא עד 10 טון, או אוטובוס זעיר), C (אופנוע, לסוגיו), F (פגיעה בהולך רגל, אין סוג רכב בו היה הנפגע) או G (מונית) ויחידת המשטרה: 12, 20, 34, 38, 51, 52 או 61;
- 3 - כאשר סוג דרך עירונית, או סוג דרך לא ידוע וסוג רכב C או F;
- 4 - ביתר המקרים.

בעזרת המודל שמוצג בטבלה 5.11, נמצא כי הסתברות הלכידה ע"י המשטרה (להיות מדווח כנפגע קשה ע"י המשטרה, בהינתן דיווח נפגע קשה בטרומה) בשנת 2008 הייתה: **0.31** בממוצע, עם סטיית תקן **0.17**, ערך מינימאלי - 0.052, ערך מרבי - 0.845.

להלן מספר דוגמאות להערכת הסתברות הלכידה ע"י המשטרה, בעזרת המודל.

דוגמא 1: עבור נפגע קשה הולך רגל בן 10, שנפגע בדרך עירונית באזור חיפה (מרחב חוף), בתאונת פגיעה בהולך רגל - הסתברות הדיווח כנפגע קשה ע"י המשטרה תהיה 0.471.

דוגמא 2: עבור נפגע קשה נהג בן 30, שהיה ברכב נוסעים פרטי ונפגע בתאונת התנגשות חזית-צד בצומת לא עירוני במרחב הנגב (יחידת משטרה 34) - הסתברות הדיווח כנפגע קשה ע"י המשטרה תהיה 0.356.

דוגמא 3: עבור נפגעת קשה בת 48, שהייתה נוסעת ברכב נוסעים פרטי ונפגעה בתאונת התנגשות חזית-חזית בקטע דרך עירונית במרחב גליל (יחידת משטרה 12) - הסתברות הדיווח כנפגע קשה ע"י המשטרה תהיה 0.165.

דוגמא 4: עבור נפגע קשה רוכב אופנוע בן 20, שנפגע בתאונת החלקה בקטע דרך לא עירונית במרחב יהודה (יחידת משטרה 43) - הסתברות הדיווח כנפגע קשה ע"י המשטרה תהיה 0.785.

## 6. סיכום: יצירת אמד למספר נפגעים קשה בתאונות הדרכים על סמך דיווח

### המשטרה

#### 6.1. כלים סטטיסטיים שפותחו במחקר

מטרת מחקר זה הייתה בפיתוח כלים סטטיסטיים להערכת מספר כולל (אמיתי) של נפגעים קשה בתאונות הדרכים בישראל. הנחת הבסיס בהערכה זו הינה שאת היקף התופעה - המספר הכולל של הנפגעים קשה בתאונות בשנה מסוימת - יש לאמוד על סמך מספר הנפגעים קשה שנרשמו, בתקופה זו, בבתי החולים. לכן, הכלים הסטטיסטיים שפותחו במחקר אמורים לאפשר חישוב מקדמי תיקון/ניפוח למספרי הנפגעים קשה המדווחים, בשנה מסוימת, על-ידי המשטרה.

קבצי הנתונים למחקר הוכנו ע"י הלשכה המרכזית לסטטיסטיקה (למ"ס) והם התבססו על תוצרי עבודה מקדימה של הלמ"ס - שילוב מידע מבתי החולים (רישום הטראומה הלאומי) עם קובץ נתוני תאונות הדרכים של המשטרה, בשנת 2008. פיתוח הכלים במחקר התבסס על שלשה סוגי נתונים שהם: נתוני הקובץ המשולב "משטרה וטראומה", נתוני קובץ "טראומה בלבד", ונתוני קובץ "משטרה בלבד" (כאשר שני הקבצים האחרונים כללו רשומות שלא נכנסו לקובץ המשולב).

נפגע קשה במחקר הנוכחי מוגדר לפי "הגדרה אדמיניסטרטיבית" שהיא: "אדם שאושפז בעקבות תאונה בבית חולים לתקופה של 24 שעות ויותר", ללא קשר לחומרת הפציעה לפי הסקאלה הרפואית<sup>11</sup>. לכל הנפגעים מרישום הטראומה, ע"י הלמ"ס חושב מאפיין על סמך הגדרה זו, דהיינו הנפגע מסומן כ-"נפגע קשה" כאשר "תאריך שחרור פחות תאריך הגעה לבית חולים" גדול מ-1.

במחקר פותחו כלים לתיקון מספרי הנפגעים קשה שמדווחים ע"י המשטרה בכל הארץ, לרבות בתאונות שהתרחשו במחוזות יהודה ושומרון. אם זאת, הכלים שפותחו במחקר ישימים גם לתיקון מספרי הנפגעים קשה בגבולות הקו הירוק, ללא יו"ש, או בחיתוכי מידע אחרים.

להערכת מספר נפגעים קשה בתנאי הארץ פותחו שני כלים:

(א) מודל False-Positiven המחשב הסתברות לדיווח שגוי על נפגע קשה ע"י המשטרה והמאפשר ניכוי של מספר נפגעים קשה שזוהו ע"י המשטרה ולא אומתו ע"י רישום הטראומה (הנפגעים קל שבטעות זוהו כנפגעים קשה ע"י המשטרה);

(ב) מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא ברישום הטראומה.

מודל א': מודל ההסתברות לדיווח שגוי על נפגע קשה ע"י המשטרה מוצג בטבלה 6.1. להערכת ההסתברות לדיווח שגוי המודל נעזר בארבעה משתנים מסבירים, לפי הקטגוריות של סוג רכב, סוג תאונה, יום בשבוע ו-יחידת המשטרה. ליצירת המודל נבנו משתנים חדשים: I\_SUG\_REHEV\_LMS, I\_SUG\_TEUNA, I\_YEHIDA, I\_YOM\_BASHAVUA, אשר משקפים הקבוצות מסוימות של ערכי המשתנים: סוג רכב, סוג תאונה, יום בשבוע, יחידת המשטרה, בהתאמה.

<sup>11</sup> כגון: ISS - Injury Severity Score

טבלה 6.1. מודל מסביר להסתברות דיווח שגוי על נפגע קשה ע"י המשטרה

משתנה מסביר	מקדם המודל
Intercept	-1.9273
I_SUG_REHEV_LMS סוג רכב	-1.2314
I_SUG_TEUNA סוג תאונה	-0.6511
I_YOM_BASHAVUA יום בשבוע	-0.8600
I_YEHIDA יחידת המשטרה	-1.5244

להלן פירוט להגדרות המשתנים במודל שבטבלה 6.1, על סמך המשתנים מקובץ המשטרה:

I\_SUG\_REHEV\_LMS = SUG\_REHEV\_LMS \*0;  
 if SUG\_REHEV\_LMS in (4 5 6 7 10 12 14 15 17 19) then I\_SUG\_REHEV\_LMS=1;  
 if SUG\_REHEV\_LMS=. and SUG\_NIFGA\_LMS in (2 3) then I\_SUG\_REHEV\_LMS=0;  
 if SUG\_REHEV\_LMS=. and SUG\_NIFGA\_LMS in (1 4 5 6 7 8 9) then I\_SUG\_REHEV\_LMS=1;

I\_SUG\_TEUNA = SUG\_TEUNA \*0;  
 if SUG\_TEUNA in (1 3 6 9 10 12 15 18 19) then I\_SUG\_TEUNA=1;

I\_YOM\_BASHAVUA = YOM\_BASHAVUA \*0;  
 if YOM\_BASHAVUA in (1 5 6 7) then I\_YOM\_BASHAVUA=1;

I\_YEHIDA = YEHIDA \*0;  
 if YEHIDA in (34 38 43) then I\_YEHIDA=1;

מודל ב': הכלי השני להערכת מספר נפגעים קשה בתנאי הארץ הינו מודל להערכת ההסתברות להיות מזוהה כנפגע קשה ע"י המשטרה, בהינתן שנפגע קשה נמצא בקבצי הטראומה. מודל זה מאפשר תיקון של תת-הדיווח בנתוני המשטרה - ניפוח מספר הנפגעים קשה שדווחו ע"י המשטרה בעזרת הנתונים מרישום הטראומה.

לפיתוח מודל זה נדרש לאתר גורמים המשפיעים על ההסתברות של נפגע קשה הנמצא בקובץ טראומה, להיות מזוהה כנפגע קשה ע"י המשטרה<sup>12</sup>. זאת הסתברות מותנית של תפיסה ע"י המשטרה בתנאי שהייתה תפיסה ע"י טראומה. הסתברות זו מכונה כ"הסתברות תפיסה מחדש" ע"י המשטרה (recapture probability), ואת הגורמים המשפיעים עליה יש למצוא באמצעות מודל הקושר בין הגורמים להסתברות.

מכיוון שהמודל מתבסס על תצפיות מקבצי הטראומה, כאשר בסוף התהליך המודל צריך להיות מופעל על נתוני המשטרה, היה צורך באיתור מאפייני הנפגעים המשותפים לשני סוגי הקבצים. פיתוח המודל היה דו-שלבי. בשלב הראשון, נבחנו קשרים בין מאפייני הנפגעים קשה בטראומה ובמשטרה ונמצאו משתנים כאלה שעשויים להיות משמעותיים בקביעת ההסתברות ללכידה (חוזרת) ע"י המשטרה, כאשר הם ניתנים לשחזור מלא בנתוני הטראומה על סמך הנתונים בקובץ המשטרה. בשלב השני, משתנים מלאכותיים כאלה (הקבצות קטגוריות של מאפייני הנפגעים) שנמצאו בשלב הראשון שימשו לצורך בחירת המודל שעל-פיו נאמדו הסתברויות הלכידה ע"י המשטרה. לבחינת הקשרים בין מאפייני הנפגעים בטראומה ובמשטרה שימש הקובץ המשולב; לפיתוח מודל מסביר

<sup>12</sup> לאחר ניכוי המקרים של ה- False Positive, בעזרת מודל א'

להסתברות להיות מדווח כנפגע קשה ע"י המשטרה שימשו נתונים משני קבצים: הקובץ המשולב וקובץ טראומה בלבד (סה"כ 5153 רשומות על נפגעים קשה לפי הטראומה).

המודל להערכת הסתברות הלכידה ע"י המשטרה מוצג בטבלה 6.2. להערכת הסתברות זו משמשים מאפיינים אלה: מגדר הנפגע, קבוצת גיל הנפגע, יחידת המשטרה, סוג דרך, סוג תאונה, סוג נפגע, וקטגוריות נפגע l\_cat המשקפות בעיקר את סוג הרכב בו נסע הנפגע.

טבלה 6.2. מודל מסביר להערכת הסתברות הלכידה ע"י המשטרה

משתנה מסביר	ערך המשתנה - משמעותו	מקדם המודל
Intercept		1.9661
T_gender - מגדר הנפגע	0 - זכר	0.1578
T_gender - מגדר הנפגע	1 - נקבה	0
T_age_c - קבוצת גיל הנפגע	1 - בני 0-14	-0.4906
T_age_c - קבוצת גיל הנפגע	2 - בני 15-24	-0.0554
T_age_c - קבוצת גיל הנפגע	3 - בני 25-54	-0.2444
T_age_c - קבוצת גיל הנפגע	4 - בני 55-64	0.0154
T_age_c - קבוצת גיל הנפגע	5 - בני 65+	0
YEHIDA_2 - יחידת משטרה	0 - יתר היחידות	0.8212
YEHIDA_2 - יחידת משטרה	1 - אם יחידה 12, 20, 34, 38, 51, 52, 61	0
T_sug_derehA - סוג דרך	0 - עירונית	-0.8120
T_sug_derehA - סוג דרך	1 - לא עירונית	0
T_Sug_p3 - סוג תאונה	1 - פגיעה בהולך רגל	0.6508
T_Sug_p3 - סוג תאונה	4 - רכב יחיד	0.1025
T_Sug_p3 - סוג תאונה	235 - התנגשויות בין כלי רכב או אחר	0
T_Sug_nifga_A - סוג נפגע	1 - הולך רגל	-1.9145
T_Sug_nifga_A - סוג נפגע	2 - נהג רכב	-2.0417
T_Sug_nifga_A - סוג נפגע	3 - נוסע ברכב	-2.0884
T_Sug_nifga_A - סוג נפגע	6 - נהג אופניים	-2.1735
T_Sug_nifga_A - סוג נפגע	45 - נהג/נוסע אופנוע	-1.6978
T_Sug_nifga_A - סוג נפגע	789 - אחר	0
l_cat - קטגורית נפגע*	1	-1.3953
l_cat - קטגורית נפגע	2	-0.4457
l_cat - קטגורית נפגע	3	-0.4934
l_cat - קטגורית נפגע	4	0

\* המשמעותיות לקטגוריות נפגע (l\_cat) הן:

- 1 - כאשר סוג רכב B (רכב משא מעל 10 טון), D (אוטובוס), E (אופניים) או H (רכב עבודה, טרקטור, רכבת או אחר);
- 2 - כאשר סוג רכב A (רכב פרטי, או מסחרי, או משא עד 10 טון, או אוטובוס זעיר), C (אופנוע, לסוגיו), F (פגיעה בהולך רגל, אין סוג רכב בו היה הנפגע) או G (מונית) ויחידת המשטרה: 12, 20, 34, 38, 51, 52 או 61;
- 3 - כאשר סוג דרך "עירונית", או סוג דרך לא ידוע וסוג רכב C או F;
- 4 - ביתר המקרים.

להלן פירוט הגדרות למשתני המודל שבטבלה 6.2, על סמך המשתנים מקובץ המשטרה:

```
*T_gender;
if min='r' then gender=0;
if min='n' then gender=1;
T_gender=gender;
if gender=. then T_gender=0;
```

```
*T_age_c;
old_age=age;
age=2008-shnat_leda;
```

```

if age=. then age=old_age;
police_age_c=.;*categorical age;
if age>=0 and age<=14 then police_age_c=1;
if age>=15 and age<=24 then police_age_c=2;
if age>=25 and age<=54 then police_age_c=3;
if age>=55 and age<=64 then police_age_c=4;
if age>=65 then police_age_c=5;
age_c= police_age_c;*categorical age;

```

```

T_age_c=age_c;
if age_c=. then T_age_c=3;

```

```

YEHIDA_2=0;
if YEHIDA in (12 20 34 38 51 52 61) then YEHIDA_2=1;

```

```

*sug_derehA;
sug_dereh2=0*sug_dereh ;
if sug_dereh>2 then sug_dereh2=1;
sug_derehA=sug_dereh2;
T_sug_derehA=sug_derehA;
if sug_derehA=. then T_sug_derehA=0;

```

```

*sug teuna - Sug_p3;
Sug_p = SUG_TEUNA*0;
if SUG_TEUNA=1 then Sug_p = 1;
if SUG_TEUNA=2 then Sug_p = 2;
if SUG_TEUNA=3 then Sug_p = 2;
if SUG_TEUNA=4 then Sug_p = 2;
if SUG_TEUNA=5 then Sug_p = 3;
if SUG_TEUNA=6 then Sug_p = 4;
if SUG_TEUNA=7 then Sug_p = 4;
if SUG_TEUNA=8 then Sug_p = 4;
if SUG_TEUNA=9 then Sug_p = 4;
if SUG_TEUNA=10 then Sug_p = 4;
if SUG_TEUNA=11 then Sug_p = 4;
if SUG_TEUNA>=12 then Sug_p =5;
Sug_p3=Sug_p;
if Sug_p in (2 3 5) then Sug_p3=235;
T_Sug_p3 =Sug_p3 ;
if Sug_p3 =. then T_Sug_p3 =235;

```

```

*Sug_nifga_A;
SUG_NIFGA_P=SUG_NIFGA_LMS;
if SUG_NIFGA_LMS in (7 8 9) then SUG_NIFGA_P=789;
if SUG_NIFGA_LMS in (4 5 ) then SUG_NIFGA_P=45;
Sug_nifga_A= SUG_NIFGA_P;
T_Sug_nifga_A=Sug_nifga_A;
if Sug_nifga_A=. then T_Sug_nifga_A=1;

```

```

*sug rehev - SUG_REHEV_A;
if SUG_REHEV_LMS=. and sug_nifga_lms=1 then SUG_REHEV_LMS=26;
if SUG_REHEV_LMS=1 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=2 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=3 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=4 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=5 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=6 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=7 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=8 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=9 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=10 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=11 then SUG_REHEV_p8="D";
if SUG_REHEV_LMS=12 then SUG_REHEV_p8="G";
if SUG_REHEV_LMS=13 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=14 then SUG_REHEV_p8="H";

```

```

if SUG_REHEV_LMS=15 then SUG_REHEV_p8="E";
if SUG_REHEV_LMS=16 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=17 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=18 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=19 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=26 then SUG_REHEV_p8="F";
SUG_REHEV_A= SUG_REHEV_p8;

*define I_cat;
if SUG_REHEV_A="" then I_cat=1;
else if SUG_REHEV_A in ("B" "D" "E" "H") then I_cat=1;
else if SUG_REHEV_A in ("A" "C" "F" "G") and YEHIDA in (12 20 34 38 51 52 61) then I_cat=2;
else if sug_derehA=0 or (sug_derehA=. and SUG_REHEV_A in ("C" "F")) then I_cat=3;
else I_cat=4;

```

לבסוף נציין ששני המודלים שנבנו במחקר הם מודלי רגרסיה לוגיסטית. במודל כזה, ההסתברות  $p$  (המשתנה המוסבר) מחושבת לפי הנוסחה הבאה:

$$p = \frac{1}{1 + e^{-z}}$$

כאשר:

$$z = b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

$x_i$  - המשתנים המסבירים במודל,

$b_i$  - מקדמי המודל.

## 6.2. יצירת אמד למספר כולל של נפגעים קשה על סמך דיווח המטרה

כדי לקבל אמד למספר הכולל של נפגעים קשה, על סמך קובץ המטרה, יש לבצע שתי פעולות:

(א) לתקן את מקרי ה-False Positive - מקרים של דיווח שגוי ע"י המטרה, בעזרת מודל א' שהוצג בסעיף 6.1;

(ב) לנפח את מספר המקרים שנותר על סמך הסתברויות הלכידה ע"י המטרה, בעזרת מודל ב' שהוצג בסעיף 6.1.

למעשה, לכל רשומה של נפגע קשה שדווח ע"י המטרה, מחושב ערך  $p_{final}$  המהווה "מספר כולל" (או אמיתי) של נפגעים קשה עם מאפיינים אלה שאמור להיות בקובץ המטרה. ערך ה- $p_{final}$  מוערך באופן הבא:

$$p_{final} = p_{TP} * (1/p_{catch})$$

$$p_{TP} = 1 - p_{FP}$$

כאשר

$p_{TP}$  - הסתברות לדיווח נכון ע"י המטרה (TP - true-positive);

$p_{FP}$  - הסתברות לדיווח שגוי ע"י המטרה (FP - False Positive), לפי מודל א' שהוצג לעיל;

$p_{catch}$  - הסתברות לכידה ע"י המטרה, לפי מודל ב' שהוצג לעיל.

הגדלים הללו ( $p_{final}$ ) מסוכמים על פני כל הנפגעים קשה שבקובץ המטרה וכך מתקבל אמד למספר הכולל של נפגעים קשה, בשנה מסוימת. (שיטה ליצירת רווח סמך למספר הנפגעים הכולל מוצגת בנספח ז').



את ההיגיון שבשיטה הנ"ל נדגים במקרה פשוט בו לכל הנפגעים יש אותה ההסתברות להיות מדווח נכון ולכולם אותה ההסתברות הלכידה. לדוגמה, נניח שבקובץ נפגעים קשה של המשטרה מדווח על 100 נפגעים. נניח שהסיכוי ל- true-positive לכל נפגע הוא 0.9, אזי נצפה ל-90 נפגעים קשה שדווחו נכון. אם ההסתברות הלכידה של נפגע קשה אמיתי היא  $1/3$ , הרי שכל אחד מ-90 הנפגעים של המשטרה מייצג 3 נפגעים אמיתיים. מכאן נבא שסה"כ יש 270 נפגעים קשה. לגבי אותו התהליך ניתן לומר שכל נפגע הוא בעצם 0.9 לפי ההסתברות ה-TP, ושה-0.9 נפגע מייצג  $2.7 = 0.9 * 1 / (1/3)$  נפגעים; אם נחבר גודל זה על פני כל 100 הנפגעים שהיו בקובץ המשטרה, נקבל את האומדן של 270 נפגעים כמספר כולל.

למעשה, ערך ה- $p_{TP}$  מהווה מקדם תיקון, בעוד שערך ה- $(1/p_{catch})$  מהווה מקדם ניפוח - לכל רשומה של נפגעים קשה בקובץ המשטרה.

בשיטה זו בוצעה הערכה של המספר הכולל של נפגעים קשה בשנת 2008 והתקבלו תוצאות כלהלן:

מספר הנפגעים קשה שדווח ע"י המשטרה היה  $N = 2201$ ;

בעקבות הפעלת מודל א' והסתברויות ה-TP, מספר הנפגעים קשה תוקן ל-  $N_{TP} = 2111$ ;

בעקבות הפעלת מודל ב' של ההסתברות הלכידה ע"י המשטרה המספר הכולל של נפגעים קשה עלה ל-  $N_{final} = 6709$ .

מאפייני ההסתברויות והערכים ששימשו לאמידת המספר הכולל של נפגעים קשה בשנת 2008 מוצגים בטבלה 6.3. ניתן לראות שבממוצע, 96% מהנפגעים קשה של המשטרה היו מדווחים נכון, כאשר המספר האמיתי של נפגעים קשה בתאונות היה גבוה פי 3.05 לעומת דיווח המשטרה.

טבלה 6.3. מאפייני ההסתברויות והערכים ששימשו לאמידת המספר הכולל של נפגעים קשה בשנת 2008

הסתברות/ערך מוערך	ממוצע	סטית תקן	ערך מינימאלי	ערך מרבי
$p_{catch}$	0.404	0.171	0.052	0.845
$p_{TP}$	0.959	0.037	0.873	0.998
$p_{final}$	3.048	2.163	1.160	19.196

לביצוע הערכה זו (של המספר הכולל של נפגעים קשה בשנת 2008) לכל רשומה של נפגעים קשה של המשטרה התקבלו ערכים של  $p_{catch}$ ,  $p_{TP}$ ,  $p_{final}$ , דהיינו מקדמי התיקון והניפוח המאפשרים לבצע הערכות של המספר הכולל של נפגעים קשה עבור חיתוכי מידע שונים.

טבלה 6.4 מציגה, לדוגמה, את תוצאות ההערכה בעזרת הכלים הסטטיסטיים שפותחו במחקר:

א - המספר הכולל של נפגעים קשה בשנת 2008, לפי סוגי הנפגעים בקובץ המשטרה;

ב - המספר הכולל של נפגעים הולכי רגל בשנת 2008, לפי סוג דרך;

ג - המספר הכולל של נפגעים הולכי רגל בשנת 2008, לפי קבוצות גיל הנפגעים.

מטבלה 6.4 ניתן לראות שכצפוי, החסרים בדיווחי המשטרה אינם אחידים. אם המספר הכולל של נפגעים קשה גבוה, בממוצע, פי 3 לעומת דיווח המשטרה, הרי, לדוגמה, המספר האמיתי של נפגעים קשה הולכי רגל גבוה פי 2.5 לעומת דיווח המשטרה, בעוד שהמספר הכולל של נפגעים קשה רוכבי אופניים גבוה פי 9.5.

בין נפגעים קשה הולכי רגל, תת-דיווח מרבי נמצא בקרב הנפגעים בצומת עירוני (המספר האמיתי של נפגעים קשה גבוה פי 2.8 לעומת המספר המדווח ע"י המשטרה) ובקרב ילדים עד גיל 14 (המספר האמיתי של נפגעים קשה גבוה פי 3 לעומת המספר המדווח ע"י המשטרה).

טבלה 6.4. תוצאות הערכה בעזרת הכלים הסטטיסטיים שפותחו במחקר

א - המספר הכולל של נפגעים קשה בשנת 2008, לפי סוגי הנפגעים בקובץ המשטרה

סוג נפגע	מספר מדווח ע"י המשטרה (N)	מספר המדווחים נכון (N <sub>TP</sub> )	המספר הכולל המוערך (N <sub>final</sub> )	רמת תת-דיווח ע"י המשטרה: יחס מוערך/ מדווח
הולך רגל	661	652	1685	2.5
נהג רכב	574	539	1595	2.8
נוסע ברכב	485	456	1432	3.0
נהג אופנוע	334	320	897	2.7
נוסע אופנוע	31	30	93	3.0
אופניים - נהג	103	101	983	9.5
אחר	13	13	24	1.9
סה"כ	2201	2111	6709	3.0

ב - המספר הכולל של נפגעים הולכי רגל בשנת 2008, לפי סוג דרך

סוג דרך - מיקום תאונה	מספר מדווח ע"י המשטרה (N)	מספר המדווחים נכון (N <sub>TP</sub> )	המספר הכולל המוערך (N <sub>final</sub> )	רמת תת-דיווח ע"י המשטרה: יחס מוערך/ מדווח
1 עירוני בצומת	154	156	433	2.8
2 עירוני לא בצומת	417	423	1120	2.6
3 לא עירוני בצומת	21	21	35	1.7
4 לא עירוני לא בצומת	60	61	97	1.6
סה"כ	652	661	1685	2.5

ג - המספר הכולל של נפגעים הולכי רגל בשנת 2008, לפי קבוצות גיל הנפגעים

קבוצת גיל	מספר מדווח ע"י המשטרה (N)	מספר המדווחים נכון (N <sub>TP</sub> )	המספר הכולל המוערך (N <sub>final</sub> )	רמת תת-דיווח ע"י המשטרה: יחס מוערך/ מדווח
0-14	183	181	557	3.0
15-24	76	75	173	2.3
25-54	141	139	357	2.5
55-64	77	76	165	2.1
65+	178	175	419	2.4
לא ידוע	6	6	14	2.4
סה"כ	661	652	1685	2.5

**לסיכום**, הכלים הסטטיסטיים שפותחו במחקר מאפשרים לבצע הערכות של המספר הכולל של נפגעים קשה בתאונות הדרכים בישראל על סמך נתוני המשטרה. אומדנים אלה ניתן לקבל עבור אוכלוסיות שונות של נפגעים קשה וכמו כן, לבצע הערכות דומות של המספר הכולל של נפגעים קשה עבור שנים אחרות. עם זאת, שימוש בכלים ובמקדמי תיקון/ניפוח שפותחו במחקר לתיקון מספרי הנפגעים קשה בשנים האחרות מתבסס על הנחת דמיון בין מצב הדיווח שהיה בשנת 2008 ובשנים האחרות.

שימוש באומדנים של המספר הכולל של נפגעים קשה בתאונות הדרכים בישראל עשוי לתרום להגברת הדיוקים בהערכות בטיחות שונות ובתכנון פעילויות התערבות. בין היתר, בדומה למחקר

שנעשה בהולנד (Reurings & Bos, 2009), ניתן לבצע הערכות רטרואקטיביות של המספרים הכוללים של נפגעים קשה בתאונות ולבחון את מגמות השינוי לאורך זמן על סמך בסיס מתוקן זה.

## מראי מקום

להלן רשימת מראי מקום לדו"ח המחקר ולנספח א'.

מתת (2004). תאונות דרכים בישראל: היקפן, מאפייניהן ואומדן נזקן למשק הלאומי. משרד התחבורה, מתת - מרכז תכנון תחבורה בע"מ.

פלג ק., אהרונוסון-דניאל ל. (2004). תאונות דרכים - פצועים במצב קשה. קבלת החלטות על בסיס נתונים חלקיים. הרפואה, כרך 143, חוב' ב', עמ' 111-115.

Alho, J. (1990). Logistic regression in capture-recapture models. *Biometrics* 46, 623-635.

Alsop, J., Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis and Prevention*, 33, 353-359.

Amoros, E., Martin, J.L., Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident Analysis and Prevention*, 38, 627-635.

Amoros, E., Martin, J.L., Laumon B. (2007). Estimating non-fatal road casualties in a large French county, using capture-recapture method. *Accident Analysis and Prevention*, 39, 483-490.

Aptel, I., Salmi, L.R., Masson, F., Bourdé, A., Henrion, G., Emy, P. (1999). Road accident statistics: discrepancies between police and hospital data in a French island. *Accident Analysis and Prevention*, 31, 101-108.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, Massachusetts: M.I.T. Press.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval Estimation for a Binomial Proportion, *Statistical Science* 16, 101-133.

Broughton J., Amoros E., Bos N., Evgenikos P., Hoeglinger S., Holló P., Pérez C., Tecl J. (2008). D.1.15. Final Report on Task 1.5. SafetyNet - Building the European Road Safety Observatory. Contract No: TREN-04-FP6TR-SI2.395465/506723 "SafetyNet".

CBS (2005). Comparison of road accident injury records between the trauma registry and the police registration. Central Bureau of Statistics, the National Road Safety Authority, September 2005 (in Hebrew).

Cohen, J. (2004). Road accidents in Israel: scope, characteristics and estimate of costs for national economy. MATAT, Ministry of Transport – Economic and Planning department (in Hebrew).

Derriks, H.M. and Mak, P.M. (2007). Underreporting of road traffic casualties. IRTAD special report. The International Transport Forum, OECD.

Elvik, R., Mysen, A.B. (1999). Incomplete accident reporting: meta-analysis of studies made in 13 countries. *Transportation Research Record*, 1665, 133-140.

Elvik, R., Vaa, T. (2004). *The handbook of road safety measures*. Elsevier.

ETSC (1999). *Reducing the Severity of Road Injuries through Post Impact Care*. European Transport Safety Council, Brussels.

ETSC (2007). *Social and economic consequences of road traffic injury in Europe*. European Transport Safety Council, Brussels.

- Fleiss, J.,L., Levin, B.,Paik, M.,C . (2003). *Statistical Methods for Rates and Proportions*, Third Edition, John Wiley & Sons, Inc.
- Gutoskie, P. (2003). The availability of hospitalized road user data in OECD member countries (2001). IRTAD Operational Committee.
- Hauer, E., Hakkert, A.S. (1988). Extent and some implications of incomplete accident reporting. *Transportation Research Record*, 1185, 1-10.
- Jarvis, S.N., Lowe, P.J., Avery, A., Levene, S., Cormack, R. (2000). Children are not goldfish – mark/recapture techniques and their application to injury data. *Injury Prevention*, 6, 46-50.
- Lopez, D.G., Rosman, D.L., Jelinek, G.A., Wilkes, G.J., Sprivulis, P.C. (2000). Complementing police road-crash records with trauma registry data - an initial evaluation. *Accident Analysis and Prevention*, 32, 771-777.
- Morrison, A., Stone, D.H. (2000). Capture-recapture: a useful methodological tool for counting traffic related injuries? *Injury Prevention*, 6, 299-304.
- Peleg, K., Aharonson-Daniel, L. (2004). Road traffic accidents - severe injuries. How missing data can impair decision making? *Harefuah*, 143 (2), 111-115 (in Hebrew).
- Reurings, M.C.B., Bos, N.M., van Kampen, L.T.B. (2007). Calculation of the real number of traffic in-patients, in 1997-2003; methods and results. SWOV report R-2007-8.
- Reurings, M.C.B. and Bos, N.M. (2009). Seriously injured road crash casualties in the Netherlands in the period 1993-2008; The real number of in-patients with a minimum MAIS of 2. SWOV report R-2009-12.
- Rosman, D.L., (1995). The feasibility of linking hospital and police road crash casualty records without names. *Accident Analysis and Prevention*, 28, 271-274.
- Simpson, H.F. (1996). Comparison of hospital and police casualty data: a national study. TRL Report 173.
- Stone, R.D. (1984). Computer linkage of transport and health data. TRRL laboratory Report 1130.
- Tersero, F., Andersson, R. (2004). Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accident Analysis and Prevention*, 36, 13-20.
- Tilling, K., Sterne, J. (1999). Capture-recapture models including covariate effects. *Amer. J. Epidemiology* 149, 2, 392-400.
- Tsui, K.L., Soa, F.L., Szeb, N.N., Wong, S.C., Leung, T.F. (2009). Misclassification of injury severity among road casualties in police reports. *Accident Analysis and Prevention*, 41, 84-89.

## נספח א': ממצאים מסקר ספרות

### a. Studies on under-reporting of road accident casualties

Similarly to other countries, the official statistics on road accidents and the majority of road safety analyses in Israel are based on data collected by the police. While fatal casualties are mostly clearly defined and well reported, non-fatal casualties are usually underreported, as frequently discussed in the literature (e.g., Hauer and Hakkert, 1988; Elvik and Mysen, 1999; Amoros et al., 2006; Derriks and Mak, 2007). The level of under-reporting is typically identified by comparing accidents and casualties recorded at hospitals with those recorded by the police in the areas serviced by the hospitals.

Elvik and Mysen (1999) summarized the reporting level of hospital-treated traffic injuries in thirteen countries (based on 49 evaluation studies) and found that it varies vastly from 21% to 88%, with a weighted mean equal to 39%. The reporting level was higher for injuries that resulted in admission to hospital as in-patients, namely 72% to 88%. Note that the "reporting level" in this analysis was defined as the number of police reported cases related to the number of cases in other data sources. However, the definition of the denominator was not consistent throughout all the studies considered.

Some studies (e.g., Elvik and Vaa, 1999; Amoros et al., 2006) defined the level of accident reporting as dependent on road user groups and accident types. In general, the reporting level is highest for car occupants, slightly lower for pedestrians, lower for motorcycle riders and lowest for cyclists. Moreover, the reporting level is typically lower for single-vehicle rather than for multi-vehicle accidents, especially when two-wheelers are involved.

Incomplete road accident reporting means that not all reportable accidents and injuries are recorded in official statistics, and therefore is part of a general problem of the availability of accurate information on road accidents. Hauer and Hakkert (1988) examined the implications of incomplete accident reporting and stated that it increased the uncertainty of the estimated effects of road safety measures. The lower and more variable the level of accident reporting, the more uncertain are estimates which are supposed to provide a basis for the evaluation of road safety policy and interventions. Ideally, the results of studies based on official accident statistics should be adjusted to account for the level of accident reporting prevailing at the time of analysis (Hauer and Hakkert, 1988). A need for national linked datasets of road accidents was recently strengthened in the European Union (ETSC, 2007).

In Israel, the awareness of the problem of accident under-reporting has increased recently because of the possibilities of comparison of various data sources. A study by Cohen (2004) compared accident data from the Central Bureau of Statistics (CBS) files with accident files from the vehicle insurance company Avner (which until 2002 kept the data on all vehicles having the obligatory component of insurance). The first source includes so-called "tav dalet" accident files that are opened by the police for accidents with injuries only, where the case satisfies a number of conditions concerning the type of driver's violation involved, a time gap

between the accident occurrence and the reporting to the police, the number of injuries occurred, the number of vehicles involved, etc. The second source is based on so-called "general" accident files which are opened by the police for the rest of reported injury accidents. Detailed comparisons of both sources revealed multiple discrepancies between the numbers reported, inconsistent long-time trends, unequal reporting levels by different police units, etc. All these discrepancies between the data sources raised many questions concerning the quality of the official road accident statistics, namely the CBS files (Cohen, 2004).

A study by Peleg and Aharonson-Daniel (2004) compared the trends in the numbers of road accident injuries based on the Israeli National Trauma Registry (ITR) versus the officially published accident statistics (CBS files). The ITR includes data of all those patients that were injured and hospitalized as well as those that were injured and died in the emergency rooms. It is maintained by the Israel Center for Trauma and Emergency Medicine Research of the Gertner Institute, and as of 2006, the ITR contained data on hospitalized patients from ten major trauma centers that, according to estimates, covered over 85% of all trauma hospitalizations in the country.

In the study by Peleg and Aharonson-Daniel (2004), significant discrepancies were identified between the official (police) numbers and the ITR numbers of injured people who are hospitalized due to road accidents. Specifically, the ITR was found to be a more comprehensive source of data. For example, over the three years between 1998 and 2000, the ITR numbers were higher than the police numbers despite the fact that the latter covers the whole country, while the ITR included only eight hospitals out of 24 total hospitals in Israel. Another important finding was that while the police figures showed a reduction in the number of injured, the ITR data indicated an increase. One of the conclusions of the study was that road safety decision-making should be based on data gathered from multiple data sources.

In absolute figures, as of year 2002, the police reported on 2419 serious casualties (those hospitalized due to accidents for more than 24 hours), while the ITR reported on 5171 patients due to road accidents of which 79% were hospitalized for more than 24 hours (Cohen, 2004). In other words, at least 1,500 serious injuries seemed lacking in the official accident statistics.

A consequent study by the CBS (2005) carried out a detailed comparison of the ITR data on hospitalized in 2002 with the combined police accident file ("tav dalet" and "general") for the same year. As the ITR data do not include personal identifications, these were reconstructed from the original hospital files. Then, the extended ITR file was matched with the police files, using casualty identifications, which enabled to produce three working files: "ITR only" records (1,341), "ITR and police" records (4,023), and the "Police only" records (177,621). The study provided detailed descriptive statistics of each working file, which contributed to the understanding of the phenomenon. Among other things, the analysis revealed that over 800

casualties which were reported as “slight injuries” by the police were actually hospitalized for more than 3 days. Accounting for additional hospitals throughout the country (besides the nine surveyed in the ITR), the CBS (2005) estimated that the total figure of serious injuries should be between 3,700 and 4,600 instead of the officially reported 2,576. Note that all the comparisons discussed above referred to the “administrative” definition of serious injury (the one introduced by the police - “a person hospitalized due to accident for more than 24 hours”). A “medical” definition of injury severity (using the Injury Severity Score) was applied for in-depth considerations of the phenomenon, but not for comparing the availability of police versus hospital records.

Summing up the findings, the under-reporting of serious injuries based on police accident files only, is obvious.

#### **b. Methods for correcting the under-reporting**

A literature survey demonstrates that usually more road accident casualties are reported by hospitals than by the police (ETSC, 1999; Elvik & Vaa, 2004), i.e. the phenomenon is not unique for Israel. To improve the reported figures, in some countries (e.g., the U.K., the Netherlands, France, Australia) probability-based statistical techniques were developed for matching the hospital and police data. In the U.K., the original matching algorithm was developed by Stone (1984), which matched the Stats19 police data for Scotland with the Scottish Hospital In-patient Statistics (SHIPS), both collected for year 1980. The matching variables involved were geographic location, sex and age of casualty, user class (driver, etc), injury severity, hour and day of hospital admission, where for each variable, tolerance values were defined. Stone (1984) reported that 70 per cent of SHIPS records were matched using this procedure.

Simpson (1996) used the linkage developed by Stone to match casualties, who attended a sample of 16 hospitals in the U.K. following a road accident, with police casualty data (Stats19). About half of the casualties attending hospitals were matched with police casualty records. The police and clinical assessments (using MAIS score) agreed in 76 per cent of the matched cases. Following multiple comparisons of age, road user, injury severity and other casualty characteristics, it was estimated that the number of seriously injured casualties in national data should be increased by a factor of 2.76 and the number of slight casualties – by a factor of 1.70.

According to Gutoskie (2003), in order to estimate the number of hospitalized road users the Netherlands links data elements common to police reported accident records and hospital medical records (the national patient register – LMR), for example, accident date, time, injury date of birth and gender of the casualties. Due to a privacy policy being in place, casualty names are not used to match records. The data linking procedures are carried out every five years, and the results are used to derive annual estimates in the intervening years. Based on the data linked for the period between 1997 and 2003, a SWOV study by Reurings et al.

(2007) developed a correction method to calculate the annual number of in-patients if no linking has taken place for that year.

In Canada, a pilot study carried out in one province showed that about 70% of police-reported and hospital medical records were matched using a probabilistic record matching approach, e.g. casualty age, month, day and time of accident (Gutoskie, 2003).

According to Derriks and Mak (2007), who analyzed the results of a survey of the OECD countries concerning under-reporting of road accident casualties, matching procedures between the police and medical databases are applied in Sweden, Great Britain, Austria and the Netherlands. The method of comparison of several databases within a country is highly recommended for the estimation of real numbers of casualties.

Rosman (1995) studied matches of hospital admissions and police records in Western Australia. The process showed that 90% of links could be identified when a phonetic code of the casualty's family name is used alongside age, gender, road user type and accident date. When names were not applied, only about 50% of the original links were found.

Lopez et al. (2000) examined the consistency of hospital and police reporting of outcomes of road traffic crashes using a database of the linked police crash reports and trauma registry records in the same area. Criteria for inclusion into the trauma registry included trauma-related causes with subsequent stay of more than 24 h or death due to injuries. Out of 497 cases of road-related injuries within the combined trauma registry in two hospitals, only 82% had matching police records. Linkage rates were associated with gender, injury severity and the number of vehicles involved. Moreover, Lopez et al. (2000) confirmed slightly higher under-reporting for pedestrians, whose police classification of injury severity was correct in 78% of cases.

Alsop and Langley (2001) evaluated the extent of under-reporting of seriously injured motor vehicle traffic crash victims, as recorded by police in New Zealand, and to what extent this coverage was biased by crash, injury, demographic, and geographic factors. Results showed that in 1995 less than two-thirds of all hospitalized vehicle occupant traffic crash victims were recorded by the police. Reporting rates varied significantly by age, injury severity, length of stay in hospital, month of crash, number of vehicles involved, whether or not a collision occurred, and geographic region, but not by gender, ethnicity or day of the week of the crash.

Amoros et al. (2006) and Amoros et al. (2007) matched the police data with the road trauma registry in the Rhone county in France, using a semi-automated record-linkage procedure. The linking variables were accident date and time, accident location, type of road user, date of birth and gender of the casualty.

Tsui et al. (2009) evaluated the discordance between police reports of injury severity among road casualties and the length of hospital stay and the Injury Severity Scale (ISS) by linking information from the crash records of the Hong Kong Police with the trauma records of a regional hospital. Sensitivity and specificity analyses suggest that police injury grading



diverges noticeably from the definition of a 12-h hospital stay, and that police reports overestimate injury severity remarkably. Specifically, the likelihood of police mis-classification is significantly related to age, ISS and position of the victim.

To sum up, in the literature, there are examples of matching the police and hospital (or trauma registry) data, where this is frequently done by using a probabilistic based matching procedure. Results of such studies are sometimes suggested for the correction of national casualty figures.

Concerning the statistical techniques that can be applied for the development of correction factors, two approaches are suggested in the literature:

- capture-recapture (CR) method;
- multivariate analysis.

The CR method is applied when two or more data sources (samples) refer to the same population and, after being matched, produce the number of cases that appear in all the sources together and the number of cases that are left unmatched in each of the sources. Moreover, it is known that by merging different lists and eliminating duplicate cases the numbers are likely to be biased downwards. Specifically, the final merged list misses those cases that were in the population, but were not ascertained in any of the lists. Using all the figures, an estimate of the total population can be provided. The CR method is widely used in biology, demography, ecology, epidemiology and recently was applied to road injury estimation (e.g., Aptel et al., 1999; Jarvis et al., 2000; Tersero and Andersson, 2004; Amoros et al., 2007). The applicability of CR methods is based on a number of assumptions that are not always realized for road accident data: equal probability for each case within the population of being captured; independence of sources; closed population; appropriateness of the “matching”.

Jarvis et al (2000) analyzed data on children injured in motor vehicle accidents, from police reported and hospital recorded files, and demonstrated examples of heterogeneity in the data (e.g., unequal chance of being reported for different ages, road user groups, injury severity), which violate the CR method's assumptions. Similarly, Morrison and Stone (2000) advocated the necessity of caution when applying corrected rates for traffic related injuries. However, Tersero and Andersson (2004) and Amoros et al. (2007), who recently applied the CR method on police-hospital matched databases for road traffic injuries, discussed each assumption explicitly and demonstrated practical ways of dealing with the data “imperfectness” in order to make them applicable for the CR method.

For example, to satisfy the homogeneous probability assumption, Amoros et al. (2007) conducted a CR analysis stratified by injury severity, road user type, third party's involvement. In order to verify the quality of “matching” this study conducted a sensitivity analysis by using three scenarios of matching: the standard linkage procedure; the standard procedure updated with the additional linkage using first name and family name of the casualty; the standard

procedure updated with an estimation of false negatives and false positives. As expected, the ascertain rates estimated by the study differed according to injury severity levels and road user types. At the same time, similarity could be noted in the estimates received by different scenarios.

Tersero and Andersson (2004), who analyzed the incidence of transport-related injuries in the Leon municipality (Nicaragua), suggested to match police records with both out-patient and in-patient records of the hospital when determining injury morbidity.

The application of the alternative method, namely the multivariate analysis, was illustrated by Amoros et al. (2006), who modeled the probability among casualties in the trauma registry of being reported also by the police as a function of accident and casualty characteristics. The study analyzed the police and trauma registry data for the period between 1997 and 2001 in the Rhone county (France), and estimated that the general police reporting rate was equal to 37.7%. Moreover, significant variations in the reporting rate were found according to injury severity, road user type, third party involvement, road type and road environment. The majority of findings about under-reporting were consistent with previous studies (e.g., Hauer and Hakkert, 1988; Elvik and Mysen, 1999). Besides, a questionnaire was sent to a sample of casualties (using postal addresses from the hospital records) in order to explore the reasons for not calling the police.

Summarizing, both statistical techniques are worth to be examined for the development of correction factors in Israel.

### **c. The SafetyNet study**

Recently, an international study was carried out in Europe, aiming to estimate the *actual* numbers of casualties in Europe from the CARE database by addressing the issue of under-reporting and the differences in national systems for injury classification (Broughton, Amoros et al, 2008). In order to overcome the inconsistencies in the reporting of non-fatal casualties, the study has:

1. estimated the under-reporting level for non-fatal casualties by developing a uniform methodology and applying it in several EU countries,
2. estimated the number of serious casualties per country according to a new common measurement unit.

Using a common methodology agreed, case-studies were carried out in 8 countries (Austria, Czech Republic, France, Greece, Hungary, The Netherlands, Spain, United Kingdom, where for two countries (Austria, the Netherlands) national data were available for the analysis.

In each case-study, files of police and hospital records were assembled for the road accidents that occurred in a common area. These files were compared to identify matching records, i.e. those casualties who were present in both files. For these matching records, certain medical details were added to the police records: length of stay in hospital and injury severity

(specifically the Maximum Abbreviated Injury Score - MAIS, an internationally accepted summary measure of injury severity).

Two matrices were then prepared to summarise the outcome of each study, one based on injury severity and the other on length of stay. These matrices were brought together for analysis, and conversion factors for each study were estimated in a consistent way. These factors allowed the actual number of serious casualties in each country to be estimated consistently from police accident statistics.

It was indicated that the new common measurement unit for counting serious casualties could be based on either injury severity or length of stay, and was concluded that the most robust definition was of a non-fatal casualty with  $MAIS \geq 3$  (inclusive).

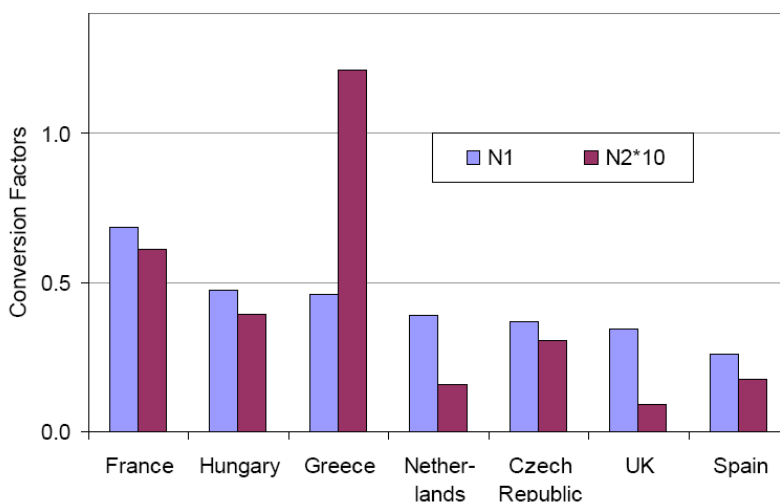
The comparisons have been made of casualty data adjusted by the conversion factors estimated by the national studies. The new common measurement unit applied was a non-fatal casualty with  $MAIS \geq 3$ . It was found that most of these were recorded by the police as seriously injured, but the studies showed that the police recorded some as slightly injured. Consequently, according to this definition, the number of casualties C in a particular country was estimated as:

$$C = N1 * \text{police reported serious casualties} + N2 * \text{police reported slight casualties},$$

where N1 and N2 varied from country to country.

The overall factors from 7 studies are shown below (they could not be estimated in Austria because of data limitations). N2 is considerably smaller than N1 and hence is multiplied by 10 in this figure.

**Conversion Factors for  $MAIS \geq 3$ , all road users**



The method of estimating the conversion factors suggested by Broughton, Amoros et al (2008) was tried for estimating the total number of serious injuries, on the initial step of the current study - see Chapter 3 of the report.

## נספח ב': מבנה קובץ A - נתונים משולבים

שם שדה	שם משתנה	קוד	מאפיינים	מקור
חומרת פגיעה למ"ס	HUMRAT_PGIA_LMS	ריק	במקרים שהחומרה זהה במשטרה ובטראומה השדה כאן יהיה ריק	למ"ס
		2	אם במשטרה היה רשום 3 (פצוע קל) ובטראומה היה רשום 2 (פצוע קשה) החומרה הסופית היא 2 (פצוע קשה)	
		3	אם במשטרה היה רשום 2 (פצוע קשה) ובטראומה היה רשום 3 (פצוע קל) החומרה הסופית היא 3 (פצוע קל)	
		5	במקרה של פטירה מאוחרת לפי המשטרה וגם טראומה	
		7	במקרים שבמשטרה היה 3 (פצוע קל) ובטראומה היה 2 (פצוע קשה) וגם היה פטירה מאוחרת, או שבמשטרה היה הרוג ובטראומה היה פטירה מאוחרת, יקבל 7	
יחידה משטרתית	YEHIDA	11	מרחב חוף (חיפה)	למ"ס
		12	מרחב גליל	
		14	מרחב עמקים	
		20	מרחב ת"א	
		33	מרחב אילת	
		34	מרחב הנגב	
		36	מרחב שמשון (עד 1999)	
		37	מרחב שמשון (החל ב-2004)	
		38	מרחב לכיש	
		41	מרחב שומרון	
		43	מרחב יהודה	
		51	מרחב השרון	
		52	מרחב השפלה	
		53	מרחב נתב"ג	
		61	מחוז ירושלים	
חודש תאונה	HODESH_TEUNA	1	ינואר	למ"ס
		2	פברואר	
		3	מרץ	
		4	אפריל	
		5	מאי	
		6	יוני	
		7	יולי	
		8	אוגוסט	
		9	ספטמבר	
		10	אוקטובר	
		11	נובמבר	
		12	דצמבר	
יום בשבוע	YOM_BASHAVUA	1	ראשון	למ"ס
		2	שני	
		3	שלישי	
		4	רביעי	
		5	חמישי	
		6	שישי	
		7	שבת	
סוג דרך	SUG_DEREH	1	עירוני בצומת	למ"ס
		2	עירוני לא בצומת	
		3	לא עירוני בצומת	
		4	לא עירוני לא בצומת	
סמל ישוב	SEMEL_YISHUV			למ"ס - ראה גיליון "סמלי יישובים"
סוג תאונה	SUG_TEUNA	1	פגיעה בהולך רגל	למ"ס
		2	התנגשות חזית אל צד	
		3	התנגשות חזית באחור	
		4	התנגשות צד בצד	
		5	התנגשות חזית אל חזית	
		6	התנגשות עם רכב שנעצר ללא חניה	

שם שדה	שם משתנה	קוד	מאפיינים	מקור
		7	התנגשות עם רכב חונה	
		8	התנגשות עם עצם דומם	
		9	ירידה מהכביש או עלייה למדרכה	
		10	התהפכות	
		11	החלקה	
		12	פגיעה בנוסע בתוך כלי רכב	
		13	נפילה ברכב נע	
		14	שריפה	
		15	אחר	
		17	התנגשות אחור אל חזית	
		18	התנגשות אחור אל צד	
		19	התנגשות עם בעל חיים	
		20	פגיעה ממשטח של רכב	
גורם עיקרי לתאונה	GOREM_TEUNA	1	עבירה של הנהג	למ"ס
		2	פעולת הולך רגל	
		3	התנהגות הנוסעים	
		4	התנהגות רוכב אופניים	
		5	אי תקינות נרכב	
		6	אחר	
		9	לא ידוע	
סוג נפגע למ"ס	SUG_NIFGA_LMS	1	הולך רגל	למ"ס
		2	נהג-רכב בעל +4 גלגלים	
		3	נוסע-רכב בעל +4 גלגלים	
		4	נהג-אופנוע	
		5	נוסע-אופנוע (לא נהג)	
		6	נהג-אופניים	
		7	נוסע-אופניים (לא נהג)	
		8	נהג-רכב לא ידוע	
		9	נוסע-רכב לא ידוע	
שנת לידה	SHNAT_LEDA			למ"ס
חומרת פגיעה	HUMRAT_PGIA	1	הרוג	למ"ס
		2	פצוע קשה	
		3	פצוע קל	
פעולת נפגע למס	PEULAT_NIFGA_LMS	1	ישב ברכב במושב קדמי	למ"ס
		2	ישב ברכב במושב אחורי	
		3	ישב ברכב בארגז מאחור	
		4	עמד ברכב	
		5	עלה לרכב	
		6	ירד מהרכב	
		7	פעולה אחרת	
		8	לא ידוע	
		9	הולך רגל	
מין	MIN	ז	זכר	
		ג	נקבה	
קוד בית חולים	KOD_BET_HOLIM_LMS			למ"ס-ראה קודים בגיליון "קודים בתי חולים-משטרה"
סמל יישוב מגורים	SEMEL_YISHUV_MEGURIM			למ"ס-ראה קובץ 2008 "רשימת יישובים ואוכלוסייתם"
קבוצת אוכלוסייה למס	KVUTZAT_OHLUSIYA_LMS	1	יהודים	למ"ס
		2	ערבים	
		3	אחרים	
		4	זרים	
סוג רכב בו נסע	SUG_REHEV_LMS	1	רכב נוסעים פרטי	למ"ס
		2	משא עד 3.5 טון אחוד	
		3	משא עד 3.5 טון לא אחוד	
		4	משא 1.4 עד 9.9 טון משקל כולל	
		5	משא 10 עד 15.9 טון משקל כולל	
		6	משא 16 עד 33.9 טון משקל כולל	

שם שדה	שם משתנה	קוד	מאפיינים	מקור
		7	משא +34.0 טון משקל כולל	
		8	אופנוע עד 50 סמ"ק	
		9	אופנוע 51 עד 250 סמ"ק	
		10	אופנוע 251 עד 500 סמ"ק	
		11	אוטובוס	
		12	מונית	
		13	רכב עבודה	
		14	טרקטור	
		15	אופניים	
		16	רכבת	
		17	אחר	
		18	אוטובוס זעיר	
		19	אופנוע +501 סמ"ק	
עבירת נהיגה של הנהג	AveratNehiga	0	אין עבירה	למ"ס
		1	מהירות מופרזת בניסיונות המקרה	
			מהירות מופרזת לגבי החוק	
		3	אי מתן זכות קדימה במעבר חצייה	
		4	אי מתן זכות קדימה בצומת להולכי רגל	
		5	סטייה מהנתיב	
			אי ציות לרמזור	
			עקיפה שלא כחוק	
			אי שמירה על הימין	
			פניה שלא כחוק	
			אי מתן זכות קדימה לרכב	
			אי ציות לתמרור - עצור	
			אי ציות לתמרור - זכות קדימה	
			נסיעה לאחור	
			אי ציות לתמרור הוריי אחר	
			אי שמירה על רווח בין מכוניות	
		6	נסיעה נגד הכיוון	
		7	לא ידוע	
			איתות	
			סינוור	
			חוסר תאורה	
			מטען שלא כחוק	
			בלמים בלתי תקינים	
			הגה בלתי תקין	
			לקויים מכניים אחרים	
			העמדת רכב בצורה בלתי חוקית	
			חוסר מחזירור	
			פתיחת דלת לצד הכביש	
			בלימת פתע	
		8	אי הבטחת שלום הנוסעים בעלייתם ובירידתם מהרכב	
			שאר הסיבות	
			שיכרות/סמים	
			הנהג נרדם	
		9	סיבות אישיות אחרות/ מגבלות רפואיות/גופניות	
סוג תיק	SUG_TIK	1	ת"ד	למ"ס
		3	כללי עם נפגעים	
גיל	age			טראומה
מין	Sex_cod	1	זכר	טראומה
		2	נקבה	
		9	לא ידוע	
סמל בית חולים	Hospital_cod	1101	שיבא- תל השומר	טראומה
		1102	רמב"ם	
		1103	אסף הרופא	
		1104	וולפסון	
		1105	רבקה זיו	
		1106	הלל יפה	
		1107	נהריה	
		1108	ברזילי	

שם שדה	שם משתנה	קוד	מאפיינים	מקור
		1109	פוריה	
		1201	איכילוב	
		1301	בלינסון	
		1302	סורוקה	
		1304	קפלן	
		1401	הדסה עין כרם	
		1501	הסקוטי	
		1503	המשפחה הקדושה	
		1601	שערי צדק	
שנה של פציעה	injury_year			טראומה
חודש של פציעה	injury_month			טראומה
הפרש בימים של "תאריך הגעה לבית חולים" לעומת "תאריך פציעה"	Injury_arrival_diff		תאריך הגעה- תאריך פציעה	טראומה
הפרש בימים של "תאריך שחרור מבית חולים" לעומת "תאריך הגעה לבית חולים"	Hospital_stay		תאריך שחרור- תאריך הגעה	טראומה
ISS	ISS	1	פצוע קל	טראומה
		2	פצוע בינוני	
		3	פצוע קשה	
		4	פצוע קשה מאוד	
		9	לא ידוע	
יעד בשחרור	Release_destination	1	בית	טראומה
		2	מת	
		3	בי"ח כללי אחר	
		4	מוסד/מחלקה שיקומית	
		5	מוסד/מחלקה סיעודית	
		6	מוסד הכולל מגורים	
		7	מחלקה שלא לצורך טיפול בטראומה	
		8	אחר	
		9	לא ידוע	
יחידה מאשפזת שהייה בטיפול נמרץ	hospitalization_unit		מחלקה מאשפזת ראשונה	קודים עפ"י מילון הנתונים של משרד הבריאות. מצורפת טבלת יחידות בתקן
	ICU	1	שהה במחלקת ט.נ.	טראומה
		2	לא שהה במחלקת ט.נ.	
סוג כביש (סוג אתר- סוג כביש מורכב משילוב של שלושת השדות: סוג אתר, תת סוג אתר, תת סוג אתר משני)	Site_type		כביש/רחוב	טראומה
		2	בית	
		3	מוסד מגורים	
		4	שטח צבאי	
		5	משק חקלאי	
		6	אתר/מבנה תעשייה	
		7	מקום ספורט/פנאי	
		8	מוסד לימודים	
		9	מבנה ציבור/מסחרי	
		88	אחר	
		99	לא ידוע	
תת סוג אתר- רק כאשר סוג אתר= 1 (כביש/רחוב)	Site_sub_type		כביש מהיר/בין עירוני	טראומה
		2	כביש עירוני	
		3	כביש אחר / דרך עפר	
		4	מדרכה/מדרכוב	
		5	מסילת רכבת	

שם שדה	שם משתנה	קוד	מאפיינים	מקור
		6	תחנת אוטובוס / רכבת	
		88	אחר	
		99	לא ידוע	
תת סוג אתר משני	Site_sec_sub_type	1	בכביש	טראומה
		2	בצומת	
		3	במחלף	
		88	אחר	
		99	לא ידוע	
מקום תאונה/ישוב	Injured_city_cod	נומרי		ראה קובץ 2008 "רשימת יישובים ואוכלוסייתם"
סוג נפגע	Ecode_sec_sub_type	1	נהג רכב	טראומה
		2	נוסע רכב	
		3	נהג אופנוע	
		4	נוסע אופנוע	
		5	רוכב אופניים	
		6	רוכב/נוסע חיות	
		7	הולך רגל	
		8	רכבת- ללא רכב ממנוע	
		9	ים	
		10	אוויר	
		11	תחבורה-אחר/לא ידוע	
סוג תאונה	Car_accident_type1- Car_accident_type3	1	פגיעה בהולך רגל	טראומה
		2	התנגשות חזית אל צד	
		3	התנגשות חזית אל אחור	
		4	התנגשות צד אל צד	
		5	התנגשות חזית אל חזית	
		6	התנגשות עם רכב שעצר	
		7	התנגשות עם רכב חונה	
		8	התנגשות בעצם קבוע	
		9	ירידה מהדרך	
		10	התהפכות	
		11	החלקה	
		12	נפילה מרכב נוסע	
		13	נהג נרדם	
		14	דיבר בטלפון נייד	
		15	עצירת פתע	
		98	אחר	
		99	לא ידוע	
סוג רכב נפגע	Injury_car_type_cod	11	מכונית פרטית	טראומה
		12	ג'יפ ורכב שטח X4	
		13	מיני ואן פרטי	
		14	משאית	
		15	אופנוע	
		16	אוטובוס	
		17	מונית	
		18	מיניבוס מסחרי	
		19	טרנזיט מסחרי	
		20	טרקטור	
		21	טרקטורון	
		22	אופניים	
		23	רכבת	
		24	בעל חיים	
		25	רולר בליידס,קורקינט	
		26	הולך רגל	
		27	ללא מעורבות של רכב נוסף	
		98	אחר	
		99	לא ידוע	
מיקום הולך רגל	Pedestrian_location_cod	1	חצה-במעבר חציה	טראומה
		2	חצה-לא במעבר חציה	
		3	חצה-לא ידוע היכן	
		4	לא חצה-אך היה על הכביש	
		5	לא חצה-לא היה על הכביש	



שם שדה	שם משתנה	קוד	מאפיינים	מקור
		99	לא ידוע אם חצה או לא	
Ecode	ECODE1-ECODE2	Ennn.n	קודים לפי icd-9. שדה מרובה= 2 מופעים	טראומה
האם מסומן כ"פצוע קשה" ע"פ הגדרה אדמיניסטרטיבית בעקבות העיבודים של הלמ"ס	PazuaKasheTrauma		פצוע קשה (תאריך שחרור- תאריך הגעה לבית חולים גדול מ-1)	
		1	לא	
		0		

## נספח ג': שיטות ה-CART וה-GBM

### שיטת ה-CART

Recursive Partitioning and Regression Trees - ראה

Breiman, Friedman, Olshen, and Stone. (1984) Classification and Regression Trees.

Wadsworth.

התוכנה:

Terry M Therneau and Beth Atkinson. R port by Brian Ripley. (2009).

rpart: Recursive Partitioning. R package version 3.1-45.

<http://CRAN.R-project.org/package=rpart>

יתרונות השיטה:

- מתקבל עץ ברור שבאמצעותו ניתן לעמוד על הקשרים בין המשתנים המסבירים למשתנה התלוי;
- מתקבלות הצעות למשתנים חלופיים במקרה של ערכים חסרים;
- ניתן לקבל רשימת משתנים חלופיים בכל שלב, יחד עם ציון מידת הריחוק שלהם ממשתנים אופטימאליים.

חסרונות השיטה:

- לא ניתן לקבל באמצעותה את חשיבות המשתנים (Variable importance) – ביישום הספרייה הנמצאת ב-R;
- ידוע מהניסיון שהחסרה של תצפית בודדת או מספר תצפיות עלולה לשנות את העץ. המודלים ב-GBM יציבים יותר.

### שיטת ה-GBM

Generalized Boosted Regression Modeling

התוכנה:

Greg Ridgeway <greg@rand.org> (2010). gbm: Generalized Boosted Regression Models. R

package version 1.6-3.1. <http://CRAN.R-project.org/package=gbm>

This package implements the generalized boosted modeling framework. Boosting is the process of iteratively adding basis functions in a greedy fashion so that each additional basis function further reduces the selected loss function. This implementation closely follows Friedman's Gradient Boosting Machine (Friedman, 2001).

J.H. Friedman (2001). Greedy Function Approximation: A Gradient Boosting Machine, Annals of Statistics 29(5):1189-1232.

יתרונות השיטה:

- ניתן לקבל באמצעותה את חשיבות המשתנים (Variable importance);
- שיטה מתקדמת יותר משתי השיטות האחרות. ניתן להסתכל על CART כמקרה פרטי של שיטה זאת.

חסרונות השיטה:

- לא מתקבל מודל ברור הקושר בין המסבירים לתוצאה. יחד עם זאת, ניתן ליצר גרפים הרומזים לפחות לגבי צורת הקשר החד-מימדי בין כל משתנה לתוצאה;
- לא מתקבלות הצעות למשתנים חלופיים כמו ב-CART, במקרה והמשתנים חסרים;
- לא מתקבל מידע לגבי החלפת המשתנים, כאשר כאן הכוונה לכך שלמרות שערכי המשתנים שנבחרו ידועים (לא חסרים) עדיין נרצה להשתמש במשתנים אחרים מאלה שנבחרו.

## נספח ד': הערכת רווחי הסמך לאמד ההסתברות ל-FP

לשם קבלת רווחי סמך לאמד ההסתברות ל-FP (דיווח שגוי ע"י המשטרה), יש לבצע bootstrap פרמטרי, ולהשתמש באמדי הפרמטרים שבטבלה 4.3 ובאמד למטריצת הקווריאנס של האמידים ( *Estimated Covariance Matrix* ) הנתון בטבלה הבאה:

Parameter	Intercept	I_SUG_REH EV_LMS	I_SUG_TEUNA	I_YOM_BASH AVUA	I_YEHIDA
Intercept	0.040069	-0.01628	-0.01968	-0.03169	-0.02109
I_SUG_REHEV_LMS	-0.01628	0.11165	-0.04066	0.004279	0.0073
I_SUG_TEUNA	-0.01968	-0.04066	0.098873	0.004925	0.003462
I_YOM_BASHAVUA	-0.03169	0.004279	0.004925	0.06706	0.003059
I_YEHIDA	-0.02109	0.0073	0.003462	0.003059	0.358207

נסמן:

X0 = 1  
 X1=I\_SUG\_REHEV\_LMS  
 x2=I\_SUG\_TEUNA  
 x3=I\_YOM\_BASHAVUA  
 x4=I\_YEHIDA  
 X={X0, X1, X2, X3, X4}

יצירת רווח סמך למספר ה-FP בקובץ נתון תהיה כלהלן:

- נחשב שכיחות של כל אחד מ-16 המקרים שבטבלה 4.3 בקובץ הנפגעים קשה על-פי המשטרה. נסמן שכיחויות אלה כ- f1- f16, בהתאמה.

- נגדיל 10000 בטות (בטא אפס עד בטא ארבע) מהתפלגות רב-נורמלית, עם וקטור תוחלות השווה לחמשת אמדי מקדמי הפרמטרים ( $\hat{\beta}_i, i = 0, \dots, 4$ ) שערכינן נתונים על-ידי העמודה *Estimate* בטבלה 4.2 ומטריצת שונות השווה ל- *Estimated Covariance Matrix*.

- עבור כל הגרלה k-ית כזאת (k=1,2,...,10000) נחשב את הביטוי הבא:

$$n_{(k)} = \sum_{i=1}^{16} f_i * \left( \frac{e^{\sum_{k=0}^4 \beta_i^{(k)} X_i}}{1 + e^{\sum_{k=0}^4 \beta_i^{(k)} X_i}} \right) = \sum_{i=1}^{16} f_i * \left( \frac{1}{1 + e^{-\sum_{k=0}^4 \beta_i^{(k)} X_i}} \right)$$

- נסדר את ה- $n_{(k)}$  ימים שהתקבלו בסדר עולה. נעיר שהאינדקס  $k$  מובא בסוגריים כדי להדגיש שהוא אינדקס

ולא חזקה. לשם קבלת רווח סמך, ברמת סמך של 95%:

- הערך ה- $10,000 * 2.5\%$  (250) ישמש כגבול סמך תחתון,

- הערך ה- $10,000 * 97.5\%$  (9750) ישמש כגבול סמך עליון.

יש לשים לב ש- $n_{(k)}$  אינו התמרה לינארית של מקדמי המודל, ולכן התפלגות הערכים הנוצרת כבר אינה

נורמלית ואף מוטיית. לכן, לא נצפה להיווצרות התפלגות סימטרית סביב ערך מרכזי כלשהו. מכאן, גם רווח הסמך לא יהיה בהכרח סימטרי ביחס לאמד הנקודתי שנבחר.

כאמד נקודתי למספר התצפיות - נפגעים המיועדים לניפוי - ניקח את:

$$\hat{n} = \sum_{i=1}^{16} f_i * \left( \frac{e^{\sum_{k=0}^4 \hat{\beta}_k X_i}}{1 + e^{\sum_{k=0}^4 \hat{\beta}_k X_i}} \right) = \sum_{i=1}^{16} f_i * \left( \frac{1}{1 + e^{-\sum_{k=0}^4 \hat{\beta}_k X_i}} \right)$$

הגרלת המספרים האקראיים לצורך בניית רווחי הסמך בוצעה תוך שימוש בספריית `mvtnorm` של R, לפי:

Alan Genz, Frank Bretz, Tetsuhisa Miwa, Xuefei Mi, Friedrich Leisch, Fabian Scheipl, Torsten Hothorn (2011). `mvtnorm`: Multivariate Normal and t Distributions. R package version 0.9-96. URL <http://CRAN.R-project.org/package=mvtnorm>

Alan Genz, Frank Bretz (2009), Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195., Springer-Verlage, Heidelberg. ISBN 978-3-642-01688-2

יישום החישוב יכול להתבצע כדלקמן:

ניתן להגריל פעם אחת את הבטות, לשמור אותן בקובץ, ולהשתמש בקובץ זה הן עבור הנתונים כולל יו"ש והן לא כולל יו"ש, הן עבור שנה זאת והן עבור שנים בהן לא ייווצרו קבצים משולבים חדשים (בהנחה שהמודלים שנבנו עדיין תקפים). למשתמש נותר רק לבצע סכום משוקלל של ההסתברויות, כאשר השקלול הוא השכיחויות של כל תא, ולקחת שברונים של ה- $n$  ימים שנוצרו על ידי השקלול.

## נספח ה': ריכוז משתנים לבניית המודל להסתברות לכידה ע"י המשטרה

להלן פירוט ההגדרות לפי משתני המודל.

### א. מגדר

בקובץ טראומה: 1=Sex\_cod זכר, 2=Sex\_cod נקבה

```
if Sex_cod=1 then gender=0;  
if Sex_cod=2 then gender=1;
```

בקובץ משטרה: MIN זכר, MIN נקבה

```
if min='ז' then gender=0;  
if min='נ' then gender=1;
```

### ב. קבוצות גיל

בקובץ משטרה:

```
age=2008-shnat_leda;  
police_age_c=.;*categorical age;  
if age>=0 and age<=14 then police_age_c=1;  
if age>=15 and age<=24 then police_age_c=2;  
if age>=25 and age<=54 then police_age_c=3;  
if age>=55 and age<=64 then police_age_c=4;  
if age>=65 then police_age_c=5;  
age_c= police_age_c;*categorical age;
```

בקובץ טראומה: באופן דומה נבנו קטגוריות הגיל של טראומה trauma\_age\_c

### ג. חודש תאונה

injury\_month של טראומה, ו-HODESH\_TEUNA

injury\_month: טראומה:

injury\_month: טראומה: HODESH\_TEUNA יושם בתוך משתנה חדש - injury\_month

```
Injury_Month=HodeshTeuna;
```

### ד. משתנה בית חולים - YEHIDA

במקום משתנה "בית חולים" נשתמש במשתנה "יחידה".

בקובץ המשטרה והמשולב המשתנה YEHIDA קיים.

בקובץ טראומה ניצור משתנה זה באופן הבא:

```
*create yehida definition;  
if Hospital_cod=1102 then YEHIDA=11;  
if Hospital_cod=1106 then YEHIDA=11;  
if Hospital_cod=1105 then YEHIDA=12;  
if Hospital_cod=1107 then YEHIDA=12;  
if Hospital_cod=1109 then YEHIDA=14;  
if Hospital_cod=1501 then YEHIDA=14;  
if Hospital_cod=1503 then YEHIDA=14;  
if Hospital_cod=1101 then YEHIDA=20;  
if Hospital_cod=1104 then YEHIDA=20;  
if Hospital_cod=1201 then YEHIDA=20;  
if Hospital_cod=1302 then YEHIDA=34;
```

```

if Hospital_cod=1108 then YEHIDA=38;
if Hospital_cod=1301 then YEHIDA=51;
if Hospital_cod=1103 then YEHIDA=52;
if Hospital_cod=1304 then YEHIDA=52;
if Hospital_cod=1401 then YEHIDA=61;
if Hospital_cod=1601 then YEHIDA=61;

```

## ה. סוג דרך - sug\_derehA

בקובץ טראומה:

```

sug_dereh_trauma=0*Site_sub_type;
if Site_sub_type=1 then sug_dereh_trauma=1;
if Site_sub_type in (2 3 4 5 6 7 88 99) then sug_dereh_trauma=0;
sug_derehA=sug_dereh_trauma;

```

בקובץ משטרה:

```

sug_dereh2=0*sug_dereh ;
if sug_dereh>2 then sug_dereh2=1;*ironi vs non ironi;
sug_derehA=sug_dereh2;

```

משתנה משותף:

```

sug_dereh2=0*sug_dereh ;
if sug_dereh>2 then sug_dereh2=1;*ironi vs non ironi;
sug_derehA=sug_dereh2;

```

## ו. סוג תאונה (Sug\_p3)

בקובץ טראומה:

```

Sug_t1 = Car_accident_type1*0;
if Car_accident_type1=1 then Sug_t1 = 1;
if Car_accident_type1=2 then Sug_t1 = 2;
if Car_accident_type1=3 then Sug_t1 = 2;
if Car_accident_type1=4 then Sug_t1 = 2;
if Car_accident_type1=5 then Sug_t1 = 3;
if Car_accident_type1=6 then Sug_t1 = 4;
if Car_accident_type1=7 then Sug_t1 = 4;
if Car_accident_type1=8 then Sug_t1 = 4;
if Car_accident_type1=9 then Sug_t1 = 4;
if Car_accident_type1=10 then Sug_t1 = 4;
if Car_accident_type1=11 then Sug_t1 = 4;
if Car_accident_type1>=12 then Sug_t1 = 5;
Sug_t1_3=Sug_t1;
if Sug_t1 in (2 3 5) then Sug_t1_3=235;
Sug_p3= Sug_t1_3;

```

משתנה משותף ובקובץ המשטרה:

```

Sug_p = SUG_TEUNA*0;
if SUG_TEUNA=1 then Sug_p = 1;
if SUG_TEUNA=2 then Sug_p = 2;
if SUG_TEUNA=3 then Sug_p = 2;

```

```

if SUG_TEUNA=4 then Sug_p = 2;
if SUG_TEUNA=5 then Sug_p = 3;
if SUG_TEUNA=6 then Sug_p = 4;
if SUG_TEUNA=7 then Sug_p = 4;
if SUG_TEUNA=8 then Sug_p = 4;
if SUG_TEUNA=9 then Sug_p = 4;
if SUG_TEUNA=10 then Sug_p = 4;
if SUG_TEUNA=11 then Sug_p = 4;
if SUG_TEUNA>=12 then Sug_p =5;
Sug_p3=Sug_p;
if Sug_p in (2 3 5) then Sug_p3=235;

```

### ז. סוג נפגע - Sug\_nifga\_A

בקובץ טראומה:

```

Sug_nifga_byT_F=.;*final variable for predict police sug nifga by trauma;
if Ecode_sec_sub_type=1 then Sug_nifga_byT_F= 2;
if Ecode_sec_sub_type=2 then Sug_nifga_byT_F= 3;
if Ecode_sec_sub_type in (3 4) then Sug_nifga_byT_F= 45;
if Ecode_sec_sub_type=5 then Sug_nifga_byT_F= 6;
if Ecode_sec_sub_type=7 then Sug_nifga_byT_F= 1;
if Ecode_sec_sub_type in (6 8 9 10 11) then Sug_nifga_byT_F = 2;
Sug_nifga_A= Sug_nifga_byT_F;

```

בקובץ משטרה ומשותף:

```

*define sug_nifga with combined categories;
SUG_NIFGA_P=SUG_NIFGA_LMS;
if SUG_NIFGA_LMS in (7 8 9) then SUG_NIFGA_P=789;
if SUG_NIFGA_LMS in (4 5 ) then SUG_NIFGA_P=45;
Sug_nifga_A= SUG_NIFGA_P;

```

### ח. סוג רכב בו נסע הנפגע - SUG\_REHEV\_A

בקובץ טראומה:

```

if Injury_car_type_cod in (11 12 13) then Injury_car_8= "A";
if Injury_car_type_cod= 14 then Injury_car_8= "B";
if Injury_car_type_cod= 15 then Injury_car_8= "C";
if Injury_car_type_cod= 16 then Injury_car_8= "D";
if Injury_car_type_cod= 17 then Injury_car_8= "G";
if Injury_car_type_cod= 18 then Injury_car_8= "D";
if Injury_car_type_cod= 19 then Injury_car_8= "A";
if Injury_car_type_cod= 20 then Injury_car_8= "H";
if Injury_car_type_cod= 21 then Injury_car_8= "H";
if Injury_car_type_cod= 22 then Injury_car_8= "E";
if Injury_car_type_cod= 23 then Injury_car_8= "H";
if Injury_car_type_cod= 24 then Injury_car_8= "H";
if Injury_car_type_cod= 25 then Injury_car_8= "H";
if Injury_car_type_cod= 26 then Injury_car_8= "F";

```

```
if Injury_car_type_cod= 27      then Injury_car_8= "H";
if Injury_car_type_cod= 98      then Injury_car_8= "H";
if Injury_car_type_cod= 99      then Injury_car_8= "I";
```

\*predict categorised police by categorised trauma;

```
SUG_REHEV_pred_8=Injury_car_8;
if Injury_car_8="I" then SUG_REHEV_pred_8="A";
```

```
SUG_REHEV_A= SUG_REHEV_pred_8;
```

בקובץ משטרה ומשותף:

```
if SUG_REHEV_LMS=. and sug_nifga_lms=1 then SUG_REHEV_LMS=26;
      *complete pedestrians into SUG_REHEV_LMS;
```

\*categorise police;

```
if SUG_REHEV_LMS=1 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=2 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=3 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=4 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=5 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=6 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=7 then SUG_REHEV_p8="B";
if SUG_REHEV_LMS=8 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=9 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=10 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=11 then SUG_REHEV_p8="D";
if SUG_REHEV_LMS=12 then SUG_REHEV_p8="G";
if SUG_REHEV_LMS=13 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=14 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=15 then SUG_REHEV_p8="E";
if SUG_REHEV_LMS=16 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=17 then SUG_REHEV_p8="H";
if SUG_REHEV_LMS=18 then SUG_REHEV_p8="A";
if SUG_REHEV_LMS=19 then SUG_REHEV_p8="C";
if SUG_REHEV_LMS=26 then SUG_REHEV_p8="F";
```

```
SUG_REHEV_A= SUG_REHEV_p8;
```



## נספח ו': כללים להשלמת ערכים חסרים במאפייני הנפגעים

בטרם התאמת המודל המסכם להסתברות לכידה ע"י המשטרה, בוצעו השלמות של ערכים חסרים במאפייני הנפגעים קשה, כמפורט להלן:

*\*complete missing values according to most probable category;*

```
T_SUG_REHEV_A=SUG_REHEV_A;  
if SUG_REHEV_A="" then T_SUG_REHEV_A="A";  
T_sug_derehA=sug_derehA;  
if sug_derehA=. then T_sug_derehA=0;  
T_gender=gender;  
if gender=. then T_gender=0;  
T_age_c=age_c;  
if age_c=. then T_age_c=3;  
T_Injury_Month=Injury_Month;  
if Injury_Month=. then T_Injury_Month=5;  
  
T_month_4=ceil(T_Injury_Month/3);  
  
T_Sug_p3=Sug_p3;  
if Sug_p3=. then T_Sug_p3=235;  
T_Sug_nifga_A=Sug_nifga_A;  
if Sug_nifga_A=. then T_Sug_nifga_A=1;
```

*\*categories according to lsmeans;*

```
age_4=age_c;  
if age_c in ( 4 5) then age_4=45;  
YEHIDA_2=0;  
if YEHIDA in (12 20 34 38 51 52 61) then YEHIDA_2=1;
```

## נספח ז': שיטה ליצירת רווח סמך למספר הנפגעים הכולל

יצירת רווח סמך למספר הנפגעים הכולל בקובץ נתון תהיה כלהלן:

- נגריל 10000 בטות (בטא אפס עד בטא ארבע) מהתפלגות רב נורמלית, עם וקטור תוחלות השווה לחמשת אמדי מקדמי הפרמטרים  $(\hat{\beta}_i, i = 0, \dots, 4)$  במודל ה-FP, שערכיו נתונים על ידי העמודה "מקדם המודל" בטבלה 6.1) ומטריצת שונות השווה ל- *Estimated Covariance Matrix* של מודל זה.

- עבור כל הגרלה k-ית כזאת ( $k=1,2,\dots,10000$ ) נחשב את ההסתברות ל-TP.

- נגריל 10000 בטות גל (בטא גל אפס עד בטא גל 25) מהתפלגות רב נורמלית, עם וקטור תוחלות השווה לעשרים וחמש אמדי מקדמי הפרמטרים במודל הלכידה ע"י המשטרה  $(\hat{\beta}_i, i = 0, \dots, 24)$  שערכיו נתונים על ידי העמודה "מקדם המודל" בטבלה 6.2) ומטריצת שונות השווה ל- *Estimated Covariance Matrix* של מודל זה.

- עבור כל הגרלה k-ית כזאת ( $k=1,2,\dots,10000$ ) נחשב את ההסתברות ללכידה.

- נחשב בעזרת בטות אלה אמד למספר הנפגעים קשה.

- נסדר את ה-  $n(k)$  ים שהתקבלו בסדר עולה. נעיר שהאינדקס k מובא בסוגריים כדי להדגיש שהוא אינדקס ולא חזקה.

לשם קבלת רווח סמך, ברמת סמך של 95%:

- הערך ה- $2.5\% * 10,000$  (250) ישמש כגבול סמך תחתון,

- הערך ה- $97.5\% * 10,000$  (9750) ישמש כגבול סמך עליון.

יש לשים לב ש-  $n(k)$  אינו התמרה לינארית של מקדמי המודל, ולכן התפלגות הערכים הנוצרת כבר אינה

נורמלית ואף מוטיית. לכן, לא נצפה להיווצרות התפלגות סימטרית סביב ערך מרכזי כלשהוא וגם רווח הסמך לא יהיה בהכרח סימטרי ביחס לאמד נקודתי שנבחר.