



בריאות

ניתוחים סטטיסטיים של נתוני קורונה: סיכום שולחן עגול

פרופ' רון קנת
פרופ' דוד שטיינברג
פרופ' עדנה שכטמן
ד"ר ראובן גל



הון
אנושי

השכלה
גבוהה

חברה

חינוך

כלכלה

מדע
וטכנולוגיה

סביבה
ואנרגיה

תכנון
ארוך טווח

תעשייה
וחדשנות

תשתיות
פיזיות



ניתוחים סטטיסטיים של נתוני קורונה: סיכום שולחן עגול

מארגנים:

פרופ' רון קנת

פרופ' דוד שטיינברג

פרופ' עדנה שכטמן

מנחה:

ד"ר ראובן גל

אפריל, 2021

אין לשכפל כל חלק מפרסום זה ללא רשות מראש ובכתב ממוסד שמואל נאמן מלבד לצורך ציטוט של קטעים קצרים במאמרי סקירה ופרסומים דומים תוך ציון מפורש של המקור.

הדעות והמסקנות המובאות בפרסום זה הן על דעת המחבר/ים ואינן משקפות בהכרח את דעת

מוסד שמואל נאמן.

תוכן עניינים

2.....	תוכן עניינים
3.....	תקציר מנהלים
4.....	1. ארגון המפגש
6.....	2. סיכום המפגש
9.....	3. ריכוז ממצאים
10.....	4. נספחים: הערות בכתב מחלק מהמשתתפים
10.....	4.1 הערות פתיחה של פרופ' קרוצ'טה
14.....	4.2 'איר גולדברג, איתי דטנר ועמית הופרט
17.....	4.3 אורלי מנור
20.....	4.4 דני פפרמן
25.....	4.5 דוד שטיינברג
28.....	4.6 ענבל גולדשטיין
30.....	4.7 כרמית רפפורט
33.....	4.8 רוני טיארג'אן-אור
37.....	4.9 נוגה ליון
40.....	4.10 רון קנת וקלליה דיסריו

תקציר מנהלים

דוח זה מסכם דיון בשולחן העגול שנערך במוסד שמואל נאמן ב 13.4.2021 על רקע ההתנסות של המדינה בהתמודדות עם פנדמיה. השולחן העגול התקיים בשעות 14:00-17:00 וכלל מומחים עם מעורבות כזו או אחרת בניתוח נתונים שקשורים לקורונה. מטרת השולחן העגול הייתה לדון באתגרים, בכשלים וביתרונות של שימוש בנתונים ופיתוח מודלים המבוססים על שיטות סטטיסטיות בניהול מגפות ובהנגשת המידע למקבלי החלטות. הדוח כולל את סיכום המפגש ונספח עם הערות ששלחו חלק מהמשתתפים בכתב. מטרת הדוח היא להציף לקחים מההתנסות בפנדמיה לשימוש מקבלי החלטות והקהילה האקדמית בהווה ובעתיד. בסעיף סיכום הממצאים מפורטות 3 יוזמות אפשריות:

יוזמה מספר 1: קידום תקינה בבתי החולים וקופות החולים לקידוד אחיד של נתונים קליניים.

יוזמה מספר 2: שילוב סטטיסטיקאים בתהליכים שונים של קבלת החלטות במערכות שונות כגון בריאות, חינוך, תחבורה, איכות הסביבה, רווחה וכו'.

יוזמה מספר 3: קידום דיון ציבורי על הבעלות ורמת השקיפות של נתונים הנאספים במערכות הבריאות.

הרעיון לקיום שולחן עגול בנושא ההתנסות בישראל בניתוח נתוני קורונה עלה בשיחה טלפונית בין המארגנים, שלושה נשיאים לשעבר של האיגוד הישראלי לסטטיסטיקה. מוסד שמואל נאמן התנדב לשמש אכסניה למפגש, ונשלחו הזמנות למומחים בישראל עם מעורבות בנושא. ההזמנות נשלחו למומחים במכון גרטנר, במשרד הבריאות, בצה"ל, בקופות החולים, בלשכה המרכזית לסטטיסטיקה, באוניברסיטת תל אביב, בטכניון, באוניברסיטה העברית, במכון וייצמן, באוניברסיטת בן גוריון ובאוניברסיטת חיפה.

נושאי הדיון הוגדרו כדלקמן:

1. מהם האתגרים העיקריים בתהליך איסוף הנתונים ובמינוף המומחיות הסטטיסטית בעת מגפת הקורונה?
2. מהי התועלת הצפויה משימוש בניתוחים סטטיסטיים של נתוני הקורונה?
3. מהם המחסומים העיקריים שכרוכים בפיתוח ושימוש בכלי ניתוח מתחום הסטטיסטיקה בזמן טיפול במגפה?
4. איך ניתן להתגבר על המחסומים האלה?
5. באיזה אמצעים יש לנקוט כדי לפשט את הדרך להשגת נתונים רלבנטיים?
6. מהו תפקיד האקדמיה באופן כללי, ותפקיד המומחיות בתחום הסטטיסטיקה בפרט, בפיתוח כלי ניתוח ובהגדלת בסיס הידע לטובת מקבלי ההחלטות?
7. מה נחשבים כסיפורי הצלחה כתוצאה מניתוח נתוני קורונה תוך שימוש בכלי ניתוח מתחום הסטטיסטיקה?

המוזמנים שאישרו השתתפותם הוזמנו לשלוח מראש הערות כתובות. אלה מופיעות בנספח מס' 1. לפתיחת המפגש הוזמן פרופ' קורדו קרוצ'טה, נשיא האיגוד האיטלקי לסטטיסטיקה שתיאר את הנעשה שם ואחריו הציג פרופ' דניאל זייפמן, נשיא לשעבר של מכון וייצמן, את המלצות הועדה שעמד בראשה ועסקה בדו-שיח בין האקדמיה ומקבלי ההחלטות בממשלה ובכלל. סקירה זו שימשה רקע לדיון.

משתתפים במפגש

מארגנים:

רון קנת – מוסד שמואל נאמן

דוד שטיינברג – אוניברסיטת תל אביב

עדנה שכטמן – אוניברסיטת בן גוריון

מנחה:

ראובן גל - מוסד שמואל נאמן

פתיחה (בזום):

קורדו קרוצ'טה – נשיא האיגוד האיטלקי לסטטיסטיקה

משתתפים:

עירד יבנה – מוסד שמואל נאמן

דניאל זייפמן – מכון וייצמן

דני פפרמן – הלשכה המרכזית לסטטיסטיקה

יואב בנימיני – אוניברסיטת תל אביב

צבי ציגלר (חלקית) – הטכניון

יאיר גולדברג – הטכניון

עמית הופרט – מכון גרטנר

אורלי מנור – האוניברסיטה העברית

ניל גנדל (בזום) – אוניברסיטת תל אביב

מלכה גורפיין – אוניברסיטת תל אביב

כרמית רפפורט – אוניברסיטת חיפה

אורי שליט – הטכניון

נגה לזין – צה"ל

רוני טיארג'אן-אור – צה"ל

ענבל גולדשטיין – מכביטק

פתחה את הדיון פרופ' **עדנה שכטמן**, תוך הבעת התסכול הקיים בקרב חוקרי הקהילה הסטטיסטית על כך שלא נראה ששותפה באופן משמעותי בטיפול ובניתוח של הנתונים שנאספו וששימשו כמכוני מדיניות לאורך המגפה.

פרופ' **קורדו קרוצ'טה**, נשיא האיגוד האיטלקי לסטטיסטיקה, התארח באמצעות זום. הוא סיפר על מספר יוזמות והישגים של הקהילה הסטטיסטית באיטליה בשנה האחרונה. אלה כללו פעולות לארגון ולניתוח נתונים ממגוון תחומים (בריאות, חברה, כלכלה), שיתופי פעולה עם הצלב האדום ועוד מומחים, פעולות ליצור נגישות לנתונים רלוונטים, וגם הסברה לציבור הרחב כדי לעזור להם להבין ולהתמודד עם הסוגיות הקשורות לניתוח הנתונים.

בהמשך, הדיון בין המשתתפים היה עירני, פורה ולפעמים אף נוקב. הוא נגע במגוון נושאים והסיכום יאורגן לפי נושאים אלו.

1. היכולת להשפיע על מקבלי החלטות:

דניאל זייפמן ממכון וייצמן דיווח על וועדה שעמד בראשה מטעם האקדמיה הלאומית למדעים. תפקיד הוועדה היה להציע מתווה לשיתוף פעולה בין הקהילה האקדמית לבין מקבלי החלטות. הוא ציין שאין היום בישראל מסגרות טובות לנהל ולקדם חיבורים בין אקדמיה לממשלה (למעט במשרד הבטחון). הוא הדגיש את החשיבות שחוקרים באוניברסיטאות ידחפו להשמיע את תוצאות ומסקנות מחקריהם בקרב אנשי ממשל ולא יחכו שיפנו אליהם, שגישה לממשלה לא תיבנה רק מדיון בין אנשי אקדמיה. הוא ציין גם את החשיבות להתמחות בחשיפה במדיה. דניאל הדגיש שהדרך להגיע למקבלי החלטות היא דרך הציבור, והדרך להגיע לציבור היא דרך רשתות חברתיות וטוויטר. דניאל ציין שאמנם יש מדען ראשי בכל משרד ממשלתי, אבל זה גוף חסר משמעות ברוב המשרדים (למעט משרד הביטחון ומשרד החינוך). הוא סיפר כדוגמה על פרויקט "מדע על הבר", פרי יוזמתו כנשיא המכון, שעורר עניין גדול בציבור והנגיש מדע לציבור. **עמית הופרט** ממכון גרטנר שיבח את שיתוף הפעולה של הקהילה הסטטיסטית עם מוסדו אך ציין גם ששיתופי הפעולה לא תמיד התבטאו בדו"חות או במאמרים, לכן חשוב לבנות מערכת תמריצים על מנת למשוך אנשי אקדמיה לשיתופי פעולה. גם הוא ציין שהפיכת מחקר להשפעה זו אמנות שיש ללמוד ולהתמחות בה. **אורלי מנור**, יו"ר המכון הלאומי לחקר שרותי הבריאות ומדיניות הבריאות, דיווחה שלמכון היו הצלחות לחדור למקבלי החלטות, בעיקר על ידי יצירת צוותי עבודה, פרסום ניירות עמדה ממוקדים, ושימוש נבון של מומחים ביחסי ציבור. **אורי שליט** מהטכניון הוסיף שחוקרים מהאקדמיה שכן הופיעו במדיה לפעמים מצאו את עצמם מול לחצים ציבוריים שמקשים על המעמד המדעי. **נגה**

לוי, מפיקוד העורף, ציינה שיש לקחת בחשבון שמקבלי החלטות עסוקים וצריך לתפוס את תשומת ליבם תוך זמן קצר או קריאה רק של סיכום מתומצת. **רוני טיארג'אן אור**, ממחלקת מדעי ההתנהגות (ממד"ה) של צה"ל, הדגיש את הצורך לשמור על מסרים פשוטים על מנת להציף מסקנות מעל פני השטח. הוא ציין שהוא שומע טענות דומות על חוסר השפעה גם מחוקרים אחרים במדעי החברה. הוא ציין שגם בתוך צה"ל קשה מאוד להגיע למקבלי החלטות וגם שם יש צורך בבניית קשרים לאורך זמן וניצולם בעת משבר. עובדה זו צויינה גם על ידי **יואב בנימיני** מאוניברסיטת תל אביב. יואב ציין ששנת הקורונה הייתה איזמה לאנושות, אך הביאה עניין רב בנתונים ביוסטטיסטיים. למרבה הצער המקצוע כלל לא בלט (בעייני הציבור ואצל מקבלי החלטות) במתן תשובות. יואב דיבר גם על הקושי להביא בפני הציבור מסקנות בהן מידה לא מבוטלת של אי-ודאות (למשל רווח סמך), פן שמוכר לעולם הסטטיסטי אך לא תמיד מובן לציבור הרחב.

||. **נתונים – נגישות, איסוף, סטנדרטיזציה: אורלי מנור** שיבחה את הפוטנציאל של מאגרי הנתונים הנאספים בישראל בידי משרד הבריאות, קופות החולים ובתי החולים. היא מקדישה מאמץ רב דרך המכון שהיא עומדת בראשו להפוך את המאגרים למשאב **לאומי**. בפרט היא ציינה כיוזמה מוצלחת את פרויקט תמנ"ע (תשתית מחקר לנתוני עתק), הפרוייקט של משרד הבריאות להנגשת הנתונים לחוקרים. דוברים אחרים הביעו פחות התלהבות מהפרוייקט, או חוסר היכרות איתו. **דני פפרמן**, הסטטיסטיקן הלאומי, תיאר את המאמצים הכבירים של הלמ"ס להמשיך את הפעילות הרגילה, למרות הקושי בביצוע סקרים בעת הקורונה, והצורך להוסיף סקרי נפשות ועסקים על מנת לקבל מידע תדיר יותר על ההשלכות החברתיות והכלכליות של המגפה. מידע זה משרת ישירות את משרדי הממשלה. דני ציין שהיו פחות הצלחות בתחום הבריאות. הצלחה אחת הייתה הסקר הסרולוגי בבני ברק, שתוכנן על ידי הלמ"ס. בימים אלו מתוכנן סקר סרולוגי ארצי של ילדים בגילאי 6-16 בשיתוף עם משרד הבריאות. **דוד שטיינברג** מאוניברסיטת תל אביב הבחין בין ניצול הנתונים הקיימים במערכת הבריאות, כפי שתיארה אורלי מנור, לבין איסוף יזום, כמו בסקרים שתיאר דני פפרמן. הוא הדגיש את החשיבות להפעיל כללים סטטיסטיים לאיסוף מדגמים על מנת לקבל מידע אמין. נציגי צה"ל בדיון -- **כרמית רפפורט**, **רוני טיארג'אן-אור** ו**נגה לוי** -- מעורבים, כל אחד/ת במקום שירותו/ה, בביצוע סקרים תדירים. גם הם ציינו את החשיבות של תכנון כל סקר, לרבות הנדגמים והתכנים הנבדקים, על מנת לענות על צרכים. **רון קנת** ממוסד שמואל נאמן העיר שהלמ"ס נמצא בתהליך של יצירת מסמך הנחיות על הדרך לבצע סקרים שאמור להיות מופץ לכל משרד ממשלתי שמעוניין לערוך סקר. מהלך דומה יכול להתקיים בתחום הנתונים הקליניים. **רוני טיארג'אן-אור** תיאר את המצב של הקורונה **כאינפוקאזיס**, בו עודף הנתונים לפעמים רק משבש את היכולת להבין ולהסיק מסקנות. **נגה לוי** דברה על מציאת האיזון בין זמינות (כלומר הבאת המידע במהירות, כאשר התוצאות ברות חשיבות) לעומת איכות (למשל הצורך לעכב מסקנות עד שיש מדגם גדול יותר, או עיכוב בגלל הקפדה על

היבטים מתודולוגים). **ענבל גולדשטיין** ממכביטק הציגה השוואה דומה, בתחום של פרסום מדעי, עם הצורך לאזן בין פרסום מהיר, אך לא מבוקר, במאגרים כמו arXiv, לבין ביקורת עמיתים, שלפעמים דורשת זמן רב מדי. היא גם העלתה בעיה כללית הקשורה לחוסר סטנדרטיזציה של נתונים. בעיה זו פוגעת במיוחד במאמצים לאחד נתונים ממקורות שונים, אך עלולה לבלוט גם בנתונים פנימיים של אותו ארגון. **מלכה גורפיין** מאוניברסיטת תל אביב חזרה על תובנה זו ואמרה שמודל שהיא פיתחה בגל הראשון של הקורונה נהפך ללא שמיש בגל השני כיוון ששינו הגדרות ולא הצליחו לספק לה נתונים שתאמו את הגדרות הגל הראשון. מלכה שיבחה את בתי החולים איתם עבדה על מידת שיתוף הפעולה במתן נתונים ועזרה בתחילת המגפה, אך ציינה שבהמשך שיתוף הפעולה נעצר. **ענבל גולדשטיין** העלתה הצעה ספציפית הנוגעת לאישור מחקרים המתבצעים על מאגרי נתונים גדולים, בהתאם להנחיות וועדת הלסינקי. היא ציינה שחסרות הנחיות איך לשחרר נתונים למחקרים כאלו תוך שמירה על פרטיות מידע רפואי אישי.

||| **ניתוחים ועיבודים סטטיסטיים:** מספר משתתפים ציינו את החשיבות של ניתוח נבון של הנתונים. **אורלי מנור** דברה על עבודתה בתחום הדבקה בקרב ילדים. **עמית הופרט**, **יאיר גולדברג** מהטכניון ו**איתי דטנר** מאוניברסיטת חיפה הפעילו מודלים למעקב אחר התקדמות המגפה. **מלכה גורפיין** ו**אורי שליט** פיתחו מודלים לחיזוי ששימשו את פרופ' ערן סגל, ממכון וויצמן, שהציג את תוצאותיהם בתקשורת. **דוד שטיינברג** הדגיש שנגישות של נתונים אינה מבטיחה שימוש חכם בהם והצביע על ניתוחים בעייתיים שתמכו במסקנות שגויות. **עמית הופרט** ביקר ספציפית את ההחלטה להפקיד בידי צה"ל מרכיב משמעותי של ניתוח נתוני הקורונה, בטענה שלא הייתה להם המומחיות הנדרשת לכך. **רון קנת** סיכם את הצורך והחשיבות בניתוח ובעיבוד הסטטיסטי בשל הפער שבין נתונים למידע, כאשר מידע מחייב עיבוד נכון של נתונים על מנת להגיע למסקנות.

לסיכום: לאורך כל הדיון המשתתפים חזרו לְמִסָּר הפתיחה של **עדנה שכטמן** על התסכול המקצועי בהתנהלות מול המגפה. נראה שהייתה הסכמה גורפת בין המשתתפים שיש חשיבות להקים מסגרת ארגונית שתוכל לשמש ככלי לרתום את המומחיות המקצועית והרצון הטוב של הקהילה האקדמית בכלל, והסטטיסטית בפרט, לתמיכה באיסוף וניתוח נתונים, והפקת מסקנות מהם בעת משבר. היתה הסכמה שצריך להקים מסגרת כזו עכשיו, כדי שנוכל להתמודד כראוי עם המשבר הבא.

מהדיון עלו ממצאים בשלושה תחומים:

1. איסוף וארגון נתונים

מתברר שאין סטנדרטיזציה מתאימה ובנוסף אין שקיפות לנתונים שלכאורה מהווים רכוש הציבור. היוזמה הממשלתית להקמת אגם נתונים, שציין דני פפרמן, עשויה לסייע בכיוון זה. ענבל גולדשטיין ציינה מספר תקנים שיכולים לשמש לאחידות קידוד נתונים קליניים.

יוזמה מספר 1: קידום תקינה בבתי החולים וקופות החולים לקידוד אחיד של נתונים קליניים.

2. שימוש מושכל בניתוחים סטטיסטיים

גורמים שונים שסיפקו הערכות למקבלי החלטות ולתקשורת על בסיס מודלים היו מעורבים כתוצאה מהיכרות איתם בעבודות קודמות, בתחומים שלא בהכרח קשורים לפנדמיה, או שהיו עם נוכחות תקשורתית גבוהה, למשל בטוטר. המסקנה היא שמעורבות אנשי מקצוע בקבלת החלטות סביב אירועים כגון פנדמיה מחייבת מעורבות מקדימה ולאורך זמן.

יוזמה מספר 2: שילוב סטטיסטיקאים בתהליכים שונים של קבלת החלטות במערכות שונות כגון בריאות, חינוך, תחבורה, איכות הסביבה, בינוי וכו' על בסיס קבוע

3. הנגשת המידע

המידע ששיקף את הנתונים הקיימים במערכות השונות לא נתפס בהכרח כרכוש הציבור. נראה שנדרש דיון רחב על הבעלות על נתונים, כפי שציינה אורלי מנור.

יוזמה מספר 3: קידום דיון ציבורי על בעלות ורמת השקיפות של נתונים הנאספים במערכת בריאות.

4. נספחים: הערות בכתב מחלק מהמשתתפים

4.1 הערות פתיחה של פרופ' קרוצ'טה

Round table on statistical analysis of Corona data

Introduction by Corrado Crocetta SIS president

Thank you, Professor Ron Kenett, for inviting me to this round table on statistical analysis of Corona data. Before starting I would like to present the Italian Statistical Society that I have the honor of managing since July 2020.

The Italian Statistical Society was founded in 1939 and it is the Italian leading organization for demographers, statisticians and data analysts. The Italian statistical society is the part of the federation of the European statistical societies FenstatS and it is associated to the international Statistical Institute ISI.

As you know, in February 2020, Italy was one of the first European country that had to face the pandemic emergency. We were not prepared for this catastrophe, that in terms of human life can be compared to the Second World War. At the beginning, the uncertainty was the most common sentiment and despite information coming from China, we did not foresee what was occurring to us. Along with the lack of medical information we had to face a lack of data about the number of swabs performed every day, the number of patients infected, hospitalized in intensive care units, and deaths.

During the lockdown, all media were presenting forecasting models estimated by chemists, physicists, astrophysicists and some statisticians and epidemiologists. Some of them were interested to foresee the end of the emergency, some others the maximum peak, but with the information available it was difficult to have affordable results. All members of the Italian Statistical Society were happy to give their contribution in this war against the virus by putting one's skills at the disposal of the community.

The Italian statistical society therefore acted on several fronts:

- a) Organizing 3 different working groups very active on social media, newspapers and televisions.
- b) Promoting statistical literacy about COVID-19 and organizing webinars to promote understanding of the most important definitions and statistical models. For example, on the ISI web pages it is possible to follow the online lessons of colleagues **Fulvia Mecatti**, elected member in the board of Italian Statistical Society and in the COVID-19 working group of the European Federation of National Statistical Societies (FENStatS) and **Clelia Di Serio**, full professor of Medical Statistics and Epidemiology at San Raffaele University Hospital and Swiss University (Lugano).
- c) Promoting epidemiological research in cooperation with Italian Red Cross and ISTAT, to estimate the percentage of asymptomatic patients in the population.
- d) Creating web pages dedicated to the dissemination of scientific articles on pandemic prediction models, infographics, data visualization tools, conferences and so on.

- e) Promoting the presentation of research project about the consequences of the pandemic on social life, economy and business. We are partners of the European project called Periscope (Pan -European Response to the Impacts of COVID - 19 and future Pandemics and Epidemics) that is coordinated by the colleague **Paolo Giudici**, elected member in the board of Italian Statistical Society. PERISCOPE will develop a comprehensive, user-friendly, openly accessible COVID Atlas and will guide policymakers and health authorities on promoting resilience and sustainability in the face of pandemics and epidemics. This project has a 10 million dollars budget and of 10 million dollars and counts on 32 European partners such as London School of economics, Karolinska institute and so on. We have also organized some webinars about futures studies because we are convinced that despite the lack of data, futurists can make use of various sources of data to study changes, measuring quantitative trends about populations, change in lifestyle, social values, emerging technologies and so on.
- f) We have devoted to the pandemic problems also the Statistical Festival that the Italian Statistical Society and the Italian National Institute of Statistics organize every year in Treviso. The title of the 2020 edition was: the world to come, ideas and measures to restart. I still remember the **Ron Kenett** plenary lecture at the Festival in 2019.
- g) We have promoted a petition using the website change.org asking people, not only SIS members, to subscribe our OPEN LETTER about the need for high-quality data for fighting against COVID-19 and the importance that people who have to analyse data have the right statistical competencies. In Italy, we have strong legislation to protect privacy. During this period, we have temporarily lost some fundamental rights such as the right to move from one region to another, to run economic activities such as bars, restaurants, cinemas, and so on, but public authorities do not allow us to access to individual patients' data. As statisticians, we are convinced that it is very important to respect privacy, but we know that if we can use the data collected by the health system, we can help public authorities to take decisions according to affordable and updated data. In one month, our document was signed by more than 1.400 people and we received the endorsement of some organizations promoting the same activities collecting the signature of about 40.000 individuals. Last January we organized a webinar with the consultant of prime minister Conte for the Italian sustainability program prof. **Filomena Maggino** and with the president of the COGIS commission who is responsible for coverage and the quality of statistical system prof. **Maurizio Vichi** during which we symbolically handed over our petition to the highest Italian authorities in the statistical field. The webinar had the attention of Italian media that are now pushing to have affordable data about the numbers of COVID-19.

I hope that the description of some activities of my scientific society can be useful for your conference and I will be happy to give you more information about the work and activity we have done.

The Italian media often refers that Israel may have reached “a sort of herd immunity” with about 5 million Israelis fully vaccinated and the number of daily coronavirus cases has plummeted by 97 percent.

I want to express my personal appreciation for the results you obtained. Your country represents a best practice and a good example to follow for Italians and I hope that also in Italy we will return to normality.

Professor Corrado Crocetta

- Full professor of Statistics, University of Foggia
- President of the Italian Statistical Society. from July 2020
- Director of the CIRPAS Populations, Health and Environment Research Center of the Universities of Bari, Foggia and Acavacata (Calabria).

4.2 יאיר גולדברג, איתי דטנר ועמית הופרט

11/4/2021

יאיר גולדברג (טכניון), איתי דטנר (אוני' חיפה), עמית הופרט (מכון גרטנר)

מדעי הנתונים בחרום - אתגרים והמלצות

מגפת הקורונה הפגישה את עולם מדעי הנתונים, על כל תתי התחומים שבו (סטטיסטיקה, מתמטיקה, מדעי המחשב, אפידמיולוגיה, ועוד) עם מצב חירום. בעוד שמצב חירום דורש קבלת החלטות מהירה באופן רציף על בסיס יומי ופעמים תוך יומי, מדעי הנתונים, תחום שאמור לאפשר קבלת החלטות מבוססת נתונים ועובדות ומתן המלצות לפעולה, מבוסס על תהליכים מורכבים ואיטיים. מדע הנתונים הינו תחום מולטידיספלינרי במהותו, הכולל היבטים "רכים", כגון יצירת שפה משותפת לכל העוסקים בתחום מסוים - אפידמיולוגיה במקרה הזה, ו"קשים" כגון איסוף, שמירה, הנגשה וניתוח נתונים אחראי, מעמיק, ומוכוון מטרה, כולל מתן הערכת אי הוודאות לגבי תוצאות שהתקבלו בניתוח.

לתפיסתנו, האתגר הגדול הינו בתחום איסוף הנתונים, איכותם והנגשתם, ורק אחר כך ניתוח כזה או אחר שנעשה על בסיסם. נציין שתי מטרות עיקריות לתהליך איסוף נתונים באופן כללי, לא דווקא לענין הקורונה:

1. יצירת מאגר מידע של "ground truth" - יש חשיבות לאיסוף נתונים שונים שיתנו תמונת מצב רחבה ככל האפשר על התחום בו עוסקים.
2. שמירת הנתונים באופן מושכל המאפשר הנגשתם לצורך דיווחים וניתוחים שוטפים לצורך קבלת החלטות ותמיכה בקביעת מדיניות מהטווח המיידי ועד הטווח הארוך. בנתונים רפואיים יש גם היבטים של אבטחת הנתונים, שמירה על פרטיות וזכויות החולה.

בדרך כלל המנגנונים המייצרים את הנתונים הם רבים ומגוונים, למשל מעבדות, בתי חולים, אפליקציות ממשלתיות של משרדי ממשלה כגון בריאות ועוד. בנוסף יש כמוכן נתונים ברשת שלרלוונטיים לניהול מגפה. הגופים שאוספים, שומרים ואמורים להנגיש את הנתונים גם הם רבים ומגוונים ופרוסים על פני מערכות מידע ארגוניות שונות של גופים שונים מקופות חולים, דרך בתי חולים ועד משרדים ממשלתיים כגון בריאות, למ"ס, חינוך ועוד.

מהתאור הקצר הנ"ל עולה שתהליך יצירת, שמירת והנגשת נתונים הינו מורכב מאוד עם היבטים תפעוליים מגוונים ולכן צריך פרוטוקול מסודר שמתאר את כל הגורמים המעורבים ביצירת הנתונים, איסופם והנגשתם ועושה סטנדרטיזציה של התהליך. כל פרוטוקול כזה צריך לקחת בחשבון היבטים של אבטחה ופרטיות.

כיום אף ארגון, אקדמי, ממשלתי או תעשייתי, יכול להחזיק בשטף את כל המומחים הרלוונטיים לתמיכה בתהליך מדע נתונים בריא ויעיל תחת כנפיו. מצד שני, בארץ יש מומחים באקדמיה, במגזר הציבורי, ובתעשייה שמכסים כל תחום אפשרי. ספציפית בסטטיסטיקה יש בארץ מומחים בסטטיסטיקה תעשייתית, ביוסטטיסטיקה, למידה סטטיסטית מודרנית, למידת מכונה, בינה מלאכותית, ניתוח נתוני עתק ועוד. בפרט, במצב חירום אין כיום גוף או מנגנון שיועד לארגן את כל המומחיות הזו בצורה מסודרת על מנת לתמוך בתהליך מדע הנתונים בכללותו: ייצור נתונים, איסופם, שמירתם, הנגשתם, ניתוחם, ודיווח לצורך קבלת החלטות וחזרה להשפעה על ייצור ואיסוף נתונים רלוונטיים במסגרת מעגל החיים הקלאסי של מדעי הנתונים.

בפועל מה שקורה במקרה שבו אין גוף מתכלל מאורגן, ואכן קרה, הוא ש"חבר מביא חבר" מה שיוצר תהליך לא אופטימלי וכנראה לא נעזר נכון בכל הידע והנסיון הקיים. בנוסף, זהו תהליך שמעלה חשש לחוסר שקיפות

ובעקבות כך איבוד אמון בתוצרים של צוות כזה או אחר, אמון שעליו קם ונופל ניהול והתמודדות מוצלחים עם אירוע חירום בסדר גודל כזה. אמון בנתונים ותובנות המבוססות עליהם הוא מרכיב קריטי בכלל בעולם מדעי הנתונים.

פתרון אפשרי - הקמת מנגנון "מילואים" לאומי של "אנשי נתונים" וסטטיסטיקאים שפועל בשגרה! ובחירום, עם התמריצים הרלוונטיים. נדגיש כי לדעתנו בכדי שמנגנון כזה יפעל בחירום הוא צריך לפעול במתכונת שגרה בכדי לפתח את היכולות לזמן אמת. לדוגמא, משרדי הממשלה יכולים לממן מספר מלגות לסטודנטים בכדי שיעבדו על בעיות שמטרידות את המדינה בשגרה ויש להם גם השלכות בחירום. כך יוצרו קשרי עבודה והיכרות בין המערכות. אפשר כמובן לחשוב על דוגמאות רבות נוספות ומנגנונים מגוונים ליצירת קשרי עבודה שוטפים.

הכותבים היו חברים בצוות מומחים מיוחד שהוקם בחודשים הראשונים של התפשטות מגפת הקורונה בישראל על מנת לייעץ למשרד הבריאות במידול מתמטי וסטטיסטי של היבטים שונים במגפה. הצוות כלל מומחים ממגוון רחב של מוסדות מחקר וארגונים אחרים ופעל תחת מכון גרטנר. אחד מתוצרי הצוות הוא דו"ח לגבי הערכת הדבקה בקרב ילדים ומבוגרים ומאמר שיצא בעקבות זה ב-PLOS Computational Biology. בנוסף הכותבים היו מנטורים בהאקאטון משותף של משרד הבריאות ואוניברסיטת חיפה בנושא הקורונה.

פרופ' אורלי מנור

פרופ' לביוסטטיסטיקה בבית הספר לבריאות הציבור של האוניברסיטה העברית, הדסה, ויושבת ראש חבר המנהלים של המכון הלאומי לחקר שרותי הבריאות ומדיניות הבריאות.

הערות רקע לשולחן עגול

1. נתוני מערכת הבריאות - "של מי לעזאזל הנתונים הללו?!" המכון הלאומי (לחקר שרותי הבריאות ומדיניות הבריאות) שבראשו אני עומדת עשה בשנים האחרונות עבודה מעמיקה בנושא ואף השפיע על מקבלי החלטות לפעול בנושא.

סדנה בינלאומית, כנס ים המלח, בסופו המנכ"ל דאז, משה בר סימן טוב, הקים מספר ועדות לקידום הנושא. נעשתה עבודה מרובה תוך שיתופי פעולה עם ישראל דיגיטלית, הרשות לחדשנות (המועצה להשכלה גבוהה ועוד). התקבלה הכרה ממשלתית בכך שנתוני הבריאות הדיגיטליים בישראל הם נכס אדיר ונדיר ונערכו מספר פרויקטים. יש הצלחות ויש נושאים שפחות התקדמו. כך נוצר פרויקט תמנ"ע (תשתית מחקר נתוני עתק) - תשתית לאומית למחקר ביג דאטה בבריאות לשרות קהילת המחקר והתעשייה. הנגשת מאגרי מידע בסביבת מחקר וירטואלית מאובטחת, התממה מתקדמת. דרך פלטפורמה זו גופים רבים קבלו נתונים בתקופת הקורונה, כולל כאלו שיעצו למקבלי החלטות. מניסיוני, הנתונים באיכות טובה, התקבלו מהר, נערך חיבור טוב עם נתוני למ"ס דרך אזורים סטטיסטיים. אבל: ישנן מספר בעיות- לא כל הגופים משתפים פעולה מכון של DATA SHARING. יש להמשיך להשקיע מאמץ בתחום הזה.

2. פעילות המכון הלאומי בתקופת הקורונה:

- המכון הלאומי הוציא קול קורא מיוחד להצעות מחקר בנושא הקורונה, נערכה הערכה מעמיקה אך מהירה, בכך קדמנו עריכת מחקרים- בכלל זה סקרים, מחקרים אפידמיולוגיים וכד'. 22 מחקרים יצאו לדרך. (תקציב ממוצע של כ-200,000 ₪).
- נערכו מספר מפגשי מדיניות - בהם עלו סוגיות של נתונים, קבלת החלטות ע"ס נתונים וכד'. הנטוורקינג סביב הארועים גם כן היה חשוב (חוקרים קבלו מידע על נושאי המחקר של עמיתיהם, ונוצר שיח בין מקבלי החלטות והאקדמיה).
- כנס ים המלח של מערכת הבריאות יערך בנושא הקורונה. סוגיות של הצורך בנתונים איכותיים לצורך קבלת החלטות יעלו בכנס.

3. פעילות נוספת שלי בתקופת הקורונה כללה -

א. נבחרתי ושמשתי כחברה בצוות בינ-אוניברסיטאי שהוקם על ידי נשיאי האוניברסיטאות כדי לסייע למקבלי החלטות הקבוצה עבדה בשלבים מוקדמים של המגפה- אפריל.

תובנות עיקריות: נערכה עבודה רבה וטובה בזמן קצר שכללה אינפוטטים סטטיסטיים בנושאי דגימה, רעיונות בנושא חלוקת הארץ לאזורים, שיטת רמזור ועוד (גם יואב היה חבר בקבוצה). בשל חסמים פוליטיים נראה היה שלא נערך דיון רציני בהצעות שהעלה הצוות.

- ב. יצרתי יחד עם עמיתים מבית הספר לבריאות הציבור ומהדסה קבוצת עבודה שהתמחתה בנושא הילדים ותפקידם במגפה- 10 איש. (כללה מומחים בדמוגרפיה, אפידמיולוגיה, סטטיסטיקה, רפואה, בריאות הציבור, מדיניות). התייחסנו למגוון סוגיות בהקשר לילדים: האם הם נדבקים, האם הם מדביקים, מה השפעת סגירת מוסדות החינוך ועוד. במסגרת זו שימשנו יועצים למשרד הבריאות, אוצר, הפרויקטור, מרכז המידע והידע, ועדות כנסת ושרים. (הצלחנו במידה סבירה)
ערכנו מעבר בעת משבר מאקדמיה "טהורה" ל advocacy או אקטיביזם
- הצענו ידע רב תחומי, זמן, מאמץ
 - ערכנו: ייעוץ מבוסס מחקר, תגובות מדעיות למחקרים/המלצות של גופים- גרטנר בני ברק אומדנים לתחלואת ילדים, סקר הסרולוגי בקופות, דו"ח מרכז המידע והידע בנושא ילדים, דו"ח משרד הבריאות בנושא ילדים, ערכנו שני ימי עיון- WEBENAR, איתרנו נתונים רלבנטיים (למנהל החינוך בירושלים היו נתונים טובים על שרשראות הדבקה, תמנ"ע)
 - כתבנו ניירות עמדה תוך שימוש משמעותי בתקשורת- תדרוכים של מגוון עיתונאים מובילים.

תובנות מפעילותי לגבי השימוש המיטבי בידע הסטטיסטי והאפידמיולוגי שנמצא באקדמיה בעת חרום כזו : יצירת מפגשים/כנסים, הקמת קבוצות עבודה, יצירת רשת מידע, קבוצות ווטסאפ, שימוש בתקשורת

IV איכות נתוני הבריאות- מספר תובנות מהתכנית הלאומית למדדי איכות לרפואת הקהילה בישראל.

התכנית הלאומית למדדי איכות לרפואת הקהילה בישראל מעריכה באופן שוטף את איכות הטיפול הרפואי שתושבי ישראל מקבלים מהקופות. את התכנית מובילה מנהלת אקדמית והיא נערכת בשיתוף פעולה מלא של כל הקופות. זוהי דוגמה ל DATA SHARING ושיתוף פעולה של גופים הנמצאים בתחרות. נתונים נאספים באופן קבוע מהתיקים הרפואיים הממוחשבים ב 4 הקופות. הקופות מכירות בנכס החשוב של נתוני המבוטחים (ראינו זאת המגפה). איכות הנתונים משתפרת כל הזמן. אך עדיין יש צורך רב בהרמוניזציה של נתונים.

4.4 דני פפרמן

ביתוחים סטטיסטיים של נתוני קורונה

השפעת המגפה על הפעילות של הסטטיסטיקה הרשמית: אתגרים והפקת לקחים

פרופ' דני פפרמן

בתקופת משבר, מקבלי החלטות, בין אם מדובר בממשלה, המגזר העסקי או אזרחים מן השורה, זקוקים לנתונים סטטיסטיים (**עובדות!!**), כדי לקבל החלטות מהירות ובעיקר **נכונות**. ניהול מגפת הקורונה (היה ועדיין) תלוי בזרימה מתמדת של נתונים איכותיים ועדכניים, על מנת שהממשלה תוכל לקבל החלטות נכונות בתחומים הבריאותיים, הכלכליים והחברתיים.

אתגרים מרכזיים שנצבו בפנינו על מנת שנוכל לספק נתונים כנדרש היו **צמצום משמעותי של כוח העבודה והצורך לעבוד מהבית**, עם גישה מוגבלת למערכות ייצור והפצה של הנתונים הסטטיסטיים החיוניים; הפסקה לחלוטין של שימוש בסוקרים **בביצוע ראיונות בבתי או עסקי הנדגמים, השבתת איסוף המחירים בקרב החנויות**, וסקרים במגזר העסקי שצומצמו בגלל פעילות עסקית שצומצמה או הושבתה כליל. כל זאת בתקופה בה ההתנהגות הכלכלית והחברתית היא תנודתית ביותר והמחייבת לכן סקרים עדכניים ומהירים.

האתגרים למערכת הסטטיסטית **אינם נעצרים ברמה הלאומית**. טיפול מקומי באתגרי המגיפה משבש את המערך הגלובאלי של הנתונים הסטטיסטיים ומכאן **יכולת ההשוואה בין מדינות**, המהווה בסיס להכוונה הכלכלית-חברתית הניתנת על ידי גופים בינלאומיים כמו ה-OECD, קרן המטבע, ארגון העבודה הבינלאומי והבנק העולמי. יוצא מכך שהחשיבות לשיתופי פעולה אינה רק עם בעלי העניין במישור המקומי, אלא גם עם המערכת הגלובאלית, כל זאת על מנת להבטיח יישום של ה"שיטות הטובות ביותר" לעריכת הסטטיסטיקה גם ובעיקר בעת משבר, תוך הרמוניזציה בין המערכות הסטטיסטיות והנתונים הסטטיסטיים.

הגופים הסטטיסטיים הבינלאומיים יצאו עם הנחיות מהירות לרציפות התפקודית והסטטיסטית של הסדרות המרכזיות כגון אינפלציה, תעסוקה, סקרים עסקיים וסקרי משקי בית, עריכת החשבונות הלאומיים ועוד. פרק מיוחד הוקדש למפקדים הלאומיים שרבים מהם היו אמורים להיערך בשנת 2020. במקביל, מדינות רבות ערכו סקרים חדשים ומהירים (עסקיים ובקרב משקי הבית) שנועדו לאספקת נתונים חשובים לניהול המגיפה, תוך לימוד הדדי על הנעשה אצל עמיתיהם במערכות הסטטיסטיות הלאומיות האחרות.

סוגיה מרכזית שעלתה באופן בולט על שולחן הדיונים בגין המגיפה הייתה יצירת סטטיסטיקה רשמית על בסיס "מקורות מידע אחרים" כמו "נתוני עתק", בנוסף למקורות המסורתיים של סקרים ישירים בקרב אוכלוסיית המטרה ו/או שימוש בנתונים מנהליים ממשלתיים. לדוגמא, קצירת נתונים באינטרנט. נושא זה ימשיך להיבחן ביתר שאת בשנים הקרובות, אך נתונים אלה אינם יכולים בשלב זה ובעתיד הנראה לעין להוות תחליף למערכת הסטטיסטית הקיימת והאיכותית. עם זאת, יש למצוא את הדרך למצות את הפוטנציאל בנתונים אלו, הן בתהליכי הייצור של הסטטיסטיקה הרשמית (מודלים לסיווג, זקיפה, בחירת כתובות וכדומה) והן, במידת האפשר, כסטטיסטיקה משלימה/תחליפית/מרחיבה. הנתונים החיצוניים הפוטנציאליים אינם נאגרים בהכרח למטרות סטטיסטיות ועבודה מדעית סטטיסטית רבה עדיין נדרשת על מנת להפיק את המיטב ממקורות המידע החדשים.

להלן אפרט בקצרה את הפעולות המיוחדות שנערכו בלמ"ס בישראל אבל בשלב זה אומר שהאתגר המרכזי של כל המערכות הסטטיסטיות בעולם היה לתחזק, לשפר ולהרחיב את המערכות הסטטיסטיות הקיימות, בתקופה בה המשאבים שעמדו לרשותן בכוח אדם ותקציבים דווקא צומצמו וניזוקו באופן משמעותי והנתונים שהועמדו לרשותם על ידי המשרדים האחרים הוגבלו מאוד מסיבות שונות, לא תמיד ענייניות.

האתגרים במערכת הסטטיסטית הלאומית בישראל

הלמ"ס מצאה את עצמה, מיד עם תחילת המגיפה, פועלת תחת גזירות קשות. כוח העבודה שעמד לרשותה **צומצם לשליש בתקופת הסגר הראשון**, והעבודה מהבית (שגם לה היו מגבלות של מערכות טכנולוגיות והגנת הסייבר) נכללה בתוך השליש המותר; עיוות זה תוקן מאוחר יותר כאשר עובדים מן הבית לא באו על חשבון עובדים שנדרשו להגיע למקום העבודה, כאשר במקביל פיתחנו יכולות לעבודה על מערכות מוגנות מהבית. נערכנו, אפוא, עם כוח העבודה המצומצם להמשיך ולהפיק את האומדנים המרכזיים (החודשיים). חיש מהר התגלו האתגרים הלוגיסטיים (שינוי כפוי של שיטות איסוף הנתונים, עסקים סגורים ומכאן חוסר מענה, וכו'), אך הלמ"ס **לא הפסיקה למעשה אף פעילות סטטיסטית חיונית בתקופה זו**. עם זאת, אחת הבעיות המרכזיות של ניהול המשק והכלכלה על בסיס נתונים סטטיסטיים איכותיים היא שבזמנים רגילים נתונים אלה מתייחסים לתקופות קודמות ולא ל"יום האתמול". שינויים בפעילות הכלכלית-חברתית בזמנים רגילים אינם מהיום למחר, אלא קורים באופן הדרגתי, והוויתור על נתונים בזמן אמת לטובת שלמות ואיכות הנתונים שמתייחסים לתקופות קודמות, לרוב מקיים את הגדרת האיכות של fit for use.

אתגרים אלו טופלו במספר דרכים:

א. **ביצוע סקרים ייעודיים המכסים אירועים של יום האתמול**, כמה שיותר קרוב לזמן אמת (חשיבות מירבית לתקופות של סגר ויציאה מהסגר), בקרב עסקים ומשקי בית, על מנת לתת מידע למקבלי ההחלטות בניהול המגיפה. בצענו סקר עסקים ב 11 מועדים שונים, וסקר נפשות ב 4 מועדים שונים.

ב. **הכפלת מועדי הפרסום של נתונים סטטיסטיים חשובים** (מעבר מפרסום חודשי לפרסום דו שבועי- סקר כח אדם על מצב התעסוקה וסקר אמון צרכנים). נראה אולי פשוט אך חייב שינויים בתכניות הדגימה כך שבכל שבועיים נקבל מדגם מייצג;

ג. שימוש במקורות מידע אחרים (במידת האפשר) כדי להשלים מידע חסר בנתונים הסטטיסטיים הקלאסיים, כמו שימוש בנתוני כרטיסי אשראי להשלמת נתוני סקר הוצאות משקי הבית לעדכון משקלות מדד המחירים לצרכן, או שימוש בנתונים מקוונים כדי לאמוד שינויים במחירים במקום ביקור בחנויות.

ד- אתגר מרכזי נוסף, המיוחד ללמ"ס, הוא **העבודה מן הבית** שנכפתה על המערכות הממשלתיות, דבר שחייב מעבר לרכש והתארגנות טכנולוגית. הגנת הסייבר, אבטחת המידע והגנה על הפרטיות הם תנאים הכרחיים לפעילותנו. מערכות היצור הסטטיסטיות מופרדות לחלוטין מן העולם החיצוני ועבודה מאומצת על פני כמה חודשים נדרשה על מנת לעצב את הארכיטקטורה המיוחדת הנדרשת כדי שעובדי הארגון האמונים על הכנת הנתונים הסטטיסטיים יוכלו לעשות זאת מהבית. אגב, פעילות זו של כניסה לנתונים מאובטחים מהבית משמשת היום כמודל למשרדים אחרים.

היכן התגלו אתגרים שהיו קשים יותר להתגבר עליהם?

עד עתה ציינתי את האתגרים המרכזיים והיכולת לצלוח אותם תוך שיתוף פעולה עם עמיתינו בארץ ובעולם. תחום אחד שהיה מאתגר יותר ועם פחות הצלחה היה **בתחום הסטטיסטיקה של הבריאות**. לאורך כל חודשי המשבר, הלמ"ס ניסתה לתרום מהידע והיכולות שלה בביצוע סקרים, אך ההצלחה הייתה חלקית, כאשר אנו יודעים שלשכות סטטיסטיות בעולם שיחקו תפקיד מרכזי יותר באיסוף, עיבוד והנגשת הנתונים הסטטיסטיים לטובת הניהול הבריאותי של המגיפה. בחלק מהמדינות נערכו, כבר בתחילת הדרך, סקרים בריאותיים מיוחדים להבנת ההתפשטות של המגיפה באוכלוסייה. הלמ"ס עשתה מאמצים ניכרים כדי לשפר את הסטטיסטיקה השוטפת שהיא מעבדת בתחומי הבריאות כמו פרסום נתוני תמותה ונתונים של עודף תמותה, אך בקשותיה והצעותיה החוזרות ונשנות לסיוע בעריכת סקרים על התפשטות המגפה ומאפייניה לא צלחו במיוחד. יחד עם זאת, הצלחנו בשיתוף עם משרד הבריאות, מכון גרטנר ועיריית בני ברק לבצע סקר סרולוגי בבני ברק ובימים אלו ממש אנו נערכים

יחד עם משרד הבריאות לבצוע סקר סרולוגי ארצי בקרב ילדים בגילאי 6-16, שאני מאוד מקווה שנוכל לחזור עליו בעוד כמה גלים בהמשך.

נושא כאוב נוסף בקשר עם משרד הבריאות היה **בהעברת נתוני תחלואה למאגר נתוני הלמ"ס** כך שנוכל להצליבם עם הנתונים המצויים ברשותנו ולהפיק אומדנים הנדרשים על ידי גופים ממשלתיים אחרים. גם כאן חשוב לי לומר שלאחר חודשים רבים של בקשות חוזרות ונשנות, אושר לנו ממש לאחרונה לקבל את נתוני הנבדקים השלייליים של בדיקות הקורונה כך שנוכל לעבד ולפרסם שיעורי תחלואה באזורים סטטיסטיים או בהתאם לאפיוני עסקים, שנתבקשנו לחשב על ידי משרד האוצר והמל"ל.

ניתוחים סטטיסטיים בתקופת הקורונה

מעצם מהותה והגדרת פעילותיה בחוק, הלמ"ס אינה עוסקת בדרך כלל בניתוחים סטטיסטיים כפי שהם מתבצעים באקדמיה, מכוני מחקר ואפילו במוסדות כמו בנק ישראל. יחד עם זאת, אציין שתי דוגמאות שאנו כן עוסקים בהן ושיכולות להיחשב כניתוחים סטטיסטיים.

1- חישוב נתונים מנוכי עונתיים ומכאן נתוני מגמה בתקופת הקורונה- לא אכנס כאן לשיטות החישוב בזמנים רגילים אך ידוע לכולנו שנתוני השנה האחרונה מתנהגים באופן שונה לגמרי מבשנים קודמות. האם העונתיות בשנה האחרונה דומה במשהו למה שחישבנו בשנים אחרות? האם לא עדיף להסתפק בנתונים המקוריים, שבחלק מהסדרות כמו בחשבונאות הלאומית הם בעלי חשיבות משנית? היינו צריכים לבדוק ולהחליט באופן פרטני לגבי כל סדרה לחוד מה לעשות.

2- חישוב חודשי של עודף תמותה בגין הקורונה שהפכה למצרך מאוד מבוקש- בתחילה פשוט החסרנו מנתוני התמותה בשנת הקורונה את נתוני התמותה של השנה הקודמת או ממוצעי כמה שנים קודמות, כפי שהומלץ על ידי הארגונים הבינלאומיים, אך בהמשך השתמשנו במודל של תחזית תמותה שפותח על ידי **קבוצת חוקרים ממוסדות אקדמאיים ואחרים** על מנת לחזות את שיעורי התמותה שהיינו מצפים לקבל לולא מגפת הקורונה, ומאז אנו משווים את התמותה למעשה לאותם תחזיות.

Data, Statistics and Public Trust

Roundtable Discussion, 13 April

David Steinberg, Tel Aviv University

The last 12 months have been unprecedented in Israel, and around the world, in the public exposure to data, their analysis and their use in guiding public policy. In the past, when first year students were asked what they think about when they hear the word “statistics”, the immediate answers were election polls and sports statistics. In the last year that changed dramatically – the universal answer was infection and mortality counts, what is “R”, and how we should react to them, both at the individual level and at the national level.

Here are several issues that I think are important for our discussion, all related to questions of data, statistics/analysis and public trust.

- What data are important? How should they be gathered?
- What data should be available to the public and to the research community? What is needed to facilitate availability?
- What information must accompany “raw data” to ensure that users understand the context and are able to use the data wisely and ethically?
- What is needed to increase public trust in data and the conclusions reached from them?

Most data that are collected serve administrative goals; they are not collected for statistical analysis or decision making. Some of these data proved essential in providing a context to understand the specific data related to the pandemic. Of more interest here is the unique data that was required to track the pandemic and guide policy: daily case counts and mortality counts, number of hospitalized patients, data relating to the economic and social impact of closure policies.

Here are some specifics. Early in the pandemic data pipelines appeared to often be nonexistent or slow and inefficient. Official organizations that needed data did not get them. Better infrastructure is needed to facilitate rapid construction of data pipelines for emergencies.

Some critical data is available only by structured accumulation efforts. Two good examples are the surveys of serology response, first in B’nei Braq and then nationwide. The first survey was a carefully designed probability sample; the second a “convenience sample” that took advantage of individuals who had regularly scheduled blood tests. Probability samples are more difficult, and more expensive, to carry out. But they carry the great advantage of providing a clear picture of the surveyed population, whereas convenience samples suffer from biases that are typically impossible to correct. What data should be collected needs careful thought. The MIT Technology Review (5 Apr 2021) describes a heavily funded program (\$1.75 billion) for sequencing viral variants, but without the ability to link the sequences back to other data that would make them useful for policy.

Data usually needs careful analysis to become valuable input to decision making. That means getting data to those who can contribute to the analyses. A good example is the dashboard that is available from the Ministry of Health (<https://datadashboard.health.gov.il/COVID-19/general>). The dashboard summarizes much useful data. A weakness of the dashboard is that some very relevant information is not provided. Another is that downloaded data appear to be limited to the time window that is shown on the screen.

Availability of valid data is crucial for public trust. The pandemic has been a remarkable breeding ground, not just for the SARS-cov-2 virus, but also for viral network posts that purport to provide essential information but too often mislead and confuse. The best antidote to this is when official sources are open about making valid data available. That has not always happened. Many posts presented anecdotal evidence that highlighted claimed dangers of vaccination; these posts convinced some Israelis to avoid vaccination. Official data on adverse reactions to vaccinations cannot be found on the dashboard, nor have I found them from any internet searches.

Data, carefully analyzed, can be extremely useful for policy. Too often analysis is not careful, especially when someone is trying to “make the case” for an argument. Moreover, “data” cannot be viewed here in a narrow context. They must include all the scientific background that is relevant to understanding what the numbers mean. I cite three examples. (1) Early in the pandemic, purely empirical analyses, giving excellent fits to the reported case numbers in many countries, predicted that SARS-cov-2 would die out naturally after about 90 days. These analyses ignored well-established theories in virology that showed that predicted we would see repeat waves of infection when restrictions were eased. (2) One of the analyses by Prof. Lass claimed that total mortality in January 2021 was actually lower than in many other recent Januaries. There were two crucial errors. Most important, the analysis used mortality data reported by the Israel Central Bureau of Statistics in early February 2021; but (as was stated on the CBS website), the data at that time were far from complete. A full recording of January was ready only at the end of February, and had about 20% more deaths than the early file. Second, the graph of January mortality from previous years had peaks and valleys. More discerning analysis would have discovered that all the peaks were in years with severe influenza, whereas 2021 had almost no influenza at all. (3) An anti-vaccination website posted an article linking the heavy death toll from SARS-cov-19 in late January and February to the vaccination campaign itself, claiming that the vaccines were elevating the death toll. The analysis completely ignored the even heavier (relative to expectations) mortality in October, long before vaccination had been initiated, and the high incidence of infection in early January.

David Steinberg is professor emeritus of statistics at Tel Aviv University. His primary area of research is the statistical design of experiments and he is interested in general questions of scientific study design.

חוקרת ביחידה לאפידמיולוגיה של מכון המחקר והחדשנות מכבי שירותי בריאות מזה כעשור. עוסקת בעיקר במחלות כרוניות של הגיל המבוגר.

נושאים לדיון:

הקורונה היא מחלה חדשה שבה רב הנסתר על הגלוי. בהתאם עלו המון סימני שאלה על התפשטות, התנהגות לאורך זמן, התנהגות באוכלוסיות שונות, גורמי סיכון להדבקה, להחמרה, לתגובה לטיפול וכו'. איסוף הנתונים בעיניי מהווה החולייה החלשה והכי תובענית בזמן. אי לכך בחרתי להתייחס בהערותי בעיקר לנושא "מהם האתגרים העיקריים בתהליך איסוף הנתונים...?" ולנושא "באיזה אמצעים יש לנקוט כדי לפשט את הדרך להשגת נתונים רלוונטיים?"

- עוד לפני שאפשר לשלב מומחים בסטטיסטיקה נדרשים נתונים מגוונים איכותיים על אוכלוסייה מייצגת וגדולה דיה (בעיקר בשלבים המוקדמים כאשר התחלואה באוכלוסייה היתה נמוכה במיוחד). המחקר תמיד צמא לדטה ודוחף לאיסוף דטה (גם בשגרה ולא רק בעת משבר). אלא שצפה בעיה אשר מוכרת גם מלפני האפידמיה: דטה פרוס ולא רציף- חלקו קהילה (מחלות רקע שקדמו להדבקה, צריכת תרופות קודמת), חלקו אשפוז (חום, סטורציה, שלל מדדים קליניים מנטרים ברמה יומית). אתגר פוליטי (מידע הוא כוח), אתגר רגולטורי (הלסינקי, התממה), אתגר טכנולוגי (כספות, אחסון). רוב התשתיות לא ערוכות לשיתופי דטה אד הוק.
- גם כשמתאפשרים חיבורי דטה בין מרכזיים (למשל בתיווך mdclone), חשוב ללמוד להכיר את הדטה של הצד השני, הוא לא דווקא self-explanatory ולא מגיע עם "data/manual" dictionary. יש לדעת מה המפתח למיזוג טבלאות, מה רמת האמינות של כל נתון, יש נתון שמבוזר על פני כמה וכמה טבלאות וצריך לאגם/לתקף. גם נתון שנשמע על פניו "פשוט" מתגלה כמסועף. אלו תהליכים שאמנם קשה לדחוס בזמן קצר או לטקסט כתוב, אך היכרות טובה עם אופן איסוף המידע חיונית לשימוש נכון בו ופרשנות נכונה של התוצאות.
- נניח שצלחנו את שלב הנגשת הנתונים, האם הם "מוכנים" לניתוח סטטיסטי? הקורונה היא מחלה חדשה- עוד אין "שפה" אחידה אוניברסלית, אין סטנדרטיזציה, כל מרכז רפואי ממציא סרגל אחר, קידוד אחר בהתייחס לאותו נתון. (דוגמא נגדית- סוכרת- טרמינולוגיה מוכרת, מבוססת).
- אשליית המידע הקיים- הנתון נגיש אך בטקסט חופשי ללא דפוס קבוע.
- קושי בתיאום בין יוזמות מקבילות. חוסר ביד מכוונת מלמעלה. שלל אנליסטים גויסו או התגייסו ביוזמתם מפאת צו השעה אבל לא קל להכתיב טבלה מנהלת (מי עושה מה כרגע). כפילויות רבות, לעתים גם פנימית בתוך ארגון (יד ימין לא יודעת מיד שמאל).

כדאי להזכיר גם סוגיה שלא נוגעת לאיסוף או שיתוף הנתונים:

- הפיתוי לעסוק בנושא חם או לספק מענה ללחץ הציבורי/כלכלי. לא כל מי שיש לו את ה-know how בעבודה עם ביג דטה/דטבייס מתמצא בשיטות מחקר המתאימות למחלות זיהומיות. חשוב לקחת בחשבון ערפלנים פונטציאליים כדי לאפשר הסקת מסקנות אחראיות. יש גורמים שפעלו עצמאית מתוך חיפזון לשחרר מהר פוסטים או מאמרים. יצירה מאולתרת של צוותי עבודה מולטי-דיסציפלינרים מצריכה זמן שכפי הנראה לא תמיד התאפשר.

לעניות דעתי כדאי לפעול להרמוניזציה של אופן תיעוד נתונים רפואיים במערכת הבריאות הישראלית, עם ובלי קשר לפרוץ המגפה. כמו כן חשוב לקדם תשתיות של שיתוף מידע ברמה לאומית. קלישאת "השלם גדול מסכום חלקיו" מתאימה גם כאן. שילוב ידיים מעין זה יתרום ודאי למחקרים של מחלות נדירות (שלא ניתן לבצע במרכז רפואי אחד בלבד). בדיקות רפואיות המבוצעות רק בחלק מהמרכזים ועוד כהנה וכהנה.

שולחן עגול בנושא ניתוחים סטטיסטיים של נתוני קורונה – מוסד נאמן 13.4.21

מאת: ד"ר כרמית רפפורט

אודותי

ד"ר כרמית רפפורט היא חוקרת, מנחה ויועצת בתחום ניהול מצבי חירום, המשכיות עסקית ורציפות תפקודית. ד"ר רפפורט היא חברת סגל והמרכזת האקדמית של התכנית לתואר שני בניהול מצבי חירום ואיזורי אסון בחוג לגיאוגרפיה ולימודי סביבה באוניברסיטת חיפה וגם מנהלת את המכון לרגולציה של מצבי חירום ואסון במרכז האקדמי למשפט ולעסקים.

כרמית קיבלה את הדוקטורט מהטכניון- מכון טכנולוגי לישראל בשנת 2011. פרסמה עשרות מאמרים בכתבי עת מקצועיים בתחומי התמחותה – התנהגות אוכלוסיה בחירום, חוסן חברתי, מנהיגות ותקשורת במצבי חירום ואסון. כמו כן, זכתה לתקציבי מחקר ועבודות ייעוץ מטעם האיחוד האירופי, משרד התיירות, משרד הבטחון, משרד המדע, הטכנולוגיה והחלל ועוד.

התייחסות לנושאי השיח

א. מהם האתגרים העיקריים בתהליך איסוף הנתונים ובמינוף המומחיות הסטטיסטית בעת מגפת הקורונה

בתקופת התפשטות המגיפה ניכר עושר של נתונים, ברמה הלאומית והבינלאומית, בנושאים מגוונים. הנתונים נחלקו לנתוני בריאות ומדיניות ולנתוני סקרים בקרב הציבור. אחד האתגרים שעלו הוא נקודות הזמן בו נאספו הנתונים, לאור השינויים בהתפשטות המגיפה במקומות שונים במדינות שונות וכן לאור מדיניות ניהול המגיפה שהשתנתה באופן תדיר. אתגר נוסף, הוא ההיבט החברתי-תרבותי אשר היתה לו משמעות רבה בהתפשטות המגיפה, למשל בהקשר למשמעת הציבורית בשמירה על הנחיות ההתקהלות. לנתוני הבריאות ישנו קשר לרקע החברתי-תרבותי שאותו יש לבחון כחלק מתהליך איסוף וניתוח הנתונים.

ב. מהי התועלת הצפויה משימוש בניתוחים סטטיסטיים של נתוני הקורונה?

המגיפה, אשר התפשטה במקומות שונים בעולם, תקפה אוכלוסיות ספציפיות (מבוגרים ובעלי מחלות רקע), והיתה מדבקת מאד. לפיכך, ניתוחים סטטיסטיים יכולים להועיל בבחינת השוואות שונות, החל מבינלאומיות ברמה התמודדות מערכות בריאות, ועד מקומיות לשם התאמה מיידית של ניהול המגיפה. לא פחות מכך חשובה הלמידה ממקום למקום – בהקשר לשינויי מזג האויר ביבשת אחת לעומת אחרת או הפעלת מדיניות כזו או אחרת. לאור ניתוח הנתונים ניתן להשפיע על המדיניות בטווחים המיידים והארוכים יותר. בנוסף, לאור חשיבות נושא התנהגות האוכלוסיה בהקשר למגיפה והתפשטותה ישנה תועלת רבה בחקר התנהגות אוכלוסיה כתגובה למדיניות הממשל, ברשתות החברתיות וההתנהלות היומיומית כחלק מתהליך של זיהוי צרכים ומתן מענים.

ג. מהם המחסומים העיקריים שכרוכים בפיתוח ושימוש בכלי ניתוח מתחום הסטטיסטיקה בזמן טיפול במגפה?

איסוף הנתונים וניתוחם במהירות ובאופן אשר ישרת את מקבלי ההחלטות דורש שיתוף פעולה בין חוקרים למשרד הבריאות וגופים ממשליים אחרים המעורבים בניהול המגיפה. שיתוף פעולה כזה כרוך במתן גישה לנתונים בזמן אמת וכך שילוב של חוקרים בדיוני מקבלי ההחלטות על מנת לספק תמנת מצב מיידית ומדוייקת לגבי הנושאים הנידונים

איך ניתן להתגבר על המחסומים האלה?

• באיזה אמצעים יש לנקוט כדי לפשט את הדרך להשגת נתונים רלבנטיים?

שיתוף אנשי מקצוע מתחום המחקר והסטטיסטיקה בקרב מקבלי ההחלטות, יצירת מאגר מאוחד של נתונים המתעדכן תדירות

• מהו תפקיד האקדמיה באופן כללי, ותפקיד המומחיות בתחום הסטטיסטיקה בפרט, בפיתוח כלי ניתוח ובהגדלת בסיס הידע לטובת מקבלי ההחלטות?

לאקדמיה תפקיד חשוב ביצירת חיבורים בין עולמות מחקר שונים לבין ניהול המגיפה. חוקרים שונים יכולים לתרום לזיהוי תהליכים ודפוסים מתהווים בזמן אמת כמו גם בהשוואת נתונים מתקופות שלפני המגיפה.

• מה נחשבים כסיפורי הצלחה כתוצאה מניתוח נתוני קורונה תוך שימוש בכלי ניתוח מתחום הסטטיסטיקה?

ביחד עם פרופ' רון קנת בוצע ניתוח נתוני תחלואה ונתוני התנהגות בישראל ובהשוואה בין ישראל ואיטליה – שתי מדינות אשר מדיניות ניהול המגיפה בהן היה שונה בתכלית. בשימוש בכלים סטטיסטיים מתקדמים (עצי החלטה, ניתוח משוואות מבניות, רשתות בייסיאניות, סימלציות) הראו כיצד בתרחישי התנהגות שונים התחלואה השתנתה. מחקר יישומי כזה יכול לספק כלי לחיזוי התנהגות האוכלוסייה וכן יאפשר ביסוס מדעי לבחירה בין חלופות עבור קובעי מדיניות.

רוני טיארג'אן – אור

ר' תחום מו"פ תשתיתי בממד"ה-צה"ל, וחוקר יישומי עצמאי. חבר הועד המנהל של האגודה הסוציולוגית הישראלית כנציג החוקרים היישומיים, וחבר הועד המנהל של אגודת חוקרי צבא-חברה בישראל. עד לשנה שעברה - ר' מרכז הסיקור הצה"לי בממד"ה, במשך כעשרים שנה. עומד בפני סיום עבודת דוקטורט במדע המדינה באוניברסיטת שטרסבורג.

תחומי עיסוק ועניין מחקרניים מרכזיים: דעת קהל ותקשורת בהקשרי לחימה ומשבר, יחסי צבא-חברה, מתודולוגיה מחקרית בדגש על תחום הסקרים, מחקרי הערכה ומדיניות.

במהלך תקופת ההתמודדות עם הקורונה עוסק בין השאר בניהול הסקרים הפנים-צבאיים בהקשר להתמודדות עם קורונה, עד היום בוצעו 17 סקרים בהיקף נרחב.

טל. 052-9468215

Orroni6@gmail.com

על סטטיסטיקה באינפו-כאוס - נושאים לדין:

להלן מספר לקחים ו/או מחשבות ונקודות עליהן ארצה לעמוד מתוך התנסותי בעת ההתמודדות עם הקורונה, כאשר חלק משאלות המחקר שהועלו על ידי מארגני הכנס ידונו במשתמע או באופן ישיר מתוך הבנת הנושאים הנ"ל.

אתגר הקשב למחקר ונתונים במשבר: מי סובל ממנו?

• **האמנם אתגר של סטטיסטיקאים? ... של מדעי החברה \ הטבע? ... או אתגר של האקדמיה בכלל? נחשפתי במספר מקומות לדיונים זהים לזה העומד בבסיס הכנס, שבמסגרתם עלתה תפיסה ש"לא מקשיבים לנו די", "אנחנו לא די מעורבים או משפיעים". בדומה לדיון הנוכחי, חוויתי דיון כמעט זהה בקרב חברי החוקרים בדיסציפלינות אחרות עמם הייתי בקשר (בעיקר מתחום מדעי החברה).**

מוצע למקד את מקור האתגר: האם הוא אמנם של הדיסציפלינה הסטטיסטית, של דיסציפלינות מסוימות או של האקדמיה בכלל?

מאפייני השדה בעת המשבר:

• **פער זמנים קצר במיוחד בין מחשבה, קבלת החלטות וביצוע בעת משבר, במיוחד בארגונים המנהלים באופן ישיר את המשבר, ובמיוחד בארגון שאני עובד בו (צה"ל), ולא רק בעת ההתמודדות עם מגפת הקורונה.**

- **אינפו-כאוס:** כאוס של אינפורמציה (מונח שתבע יחיאל לימור): מסביר להבנתי באופן הטוב ביותר את ההתארגנות של המחקר בתוך האירוע, לפחות בתחילתו: מידע רב, שגוי בחלקו ו/או חלקי המעורב במידע "אמין". במציאות כזו נוצרת תחושה שהמידע הוא בלתי ניתן להכלה: סותר ומסתיר, מתחרה, מגיע ממגוון מקורות פורמליים ולא פורמליים, מקורותיו לא תמיד ברורים ו/או אמינים – וכל המקורות מתחרים על אזנם של החוקרים ושל מקבלי החלטות, קשה להבין מה "אמין" ומה פחות. האינפו-כאוס מתגבר מאוד בהשפעת ה-new media.

נושא זה היה נוכח גם בעת משברי עבר (בעיקר בתחילת אירועי לחימה), אך בהיקף מצומצם יחסית.

רוב הנושאים שאכתוב עליהם מכאן קשורים לנושא זה. משמעויות:

- כאנשי מקצוע העוסקים במחקר וידע, איך אנו לא הופכים לעוד "מקור קטן" ולא משמעותי של מידע בבלייל האינפו-כאוס?
- איך יוצרים קשב ארגוני?
- איך אפשר בכל זאת להפיק אמירות משמעותיות, מורכבות, ו/או מקוריות?
- לעתים נוצר צורך במידע על נושא ספציפי שעניין לרגע את מקבלי החלטות, דרש מאתנו החוקרים עבודה ניכרת ומיד נעלם העניין, עוד בטרם הספקנו להעביר איזושהי מסקנה מסודרת לגביו.

- **מגפה כלל עולמית- משמע: טשטוש גבולות:**

- **כולם שותפים:** ביומיום הצבא הוא ארגון סגור שלעתים קרובות מעדיף לבצע באופן עצמאי את תהליכי החשיבה שלו, ולעתים לשלב מומחים בתהליכים מסוימים. בשונה מכל אירוע אחר, במהלך ההתמודדות עם הקורונה הייתה פריצת גבולות: השתתפתי ושיתפתי בפורומים מגוונים ושונים, בצבא ומחוצה לו, עם נציגי ממשלה, אקדמיה, גופים אחרים – על בסיס יומיומי כמעט. גם אלה היו חלק מתהליך יצירת האינפו-כאוס.

אני מאמין שמדובר בהזדמנות למינוף שיתופי הפעולה לעתיד.

- **הארגון (שלי) במצב אנומלי:** בהמשך ישיר לנאמר קודם, צה"ל הוא בדרך כלל הארגון הממונה על ניהול משברים שאנו "רגילים" אליהם, קרי משברים ארגוניים. כאן היה משבר שלא הובל על ידי הצבא אך השפיע עליו מאוד.

- **אלמנט הזמן:** בעת התרחשות משבר, מתי מקשיבים לנו יותר ומתי פחות? אלמנט הראשוניות, תקופת ההתרגלות והתקופה שלאחר המשבר. קיים צורך בתכנון השימוש בנתונים (פשוטים ומורכבים) על ציר ניהול המשבר.

מאפייני המחקר באינפו-כאוס:

- **זילות הנתונים, המחקר והמידע:** להערכתי בעת התרחשות אינפו-כאוס, המחקר, ובוודאי סטטיסטיקה ונתונים סטטיסטיים הופכים בתודעה של מקבלי החלטות לזולים, נפוצים מדי. ייתכן שמדובר בתהליך החורג מזמני משבר. שתי אפשרויות משוערות להתפתחות:

- **הפסימי:** תהליך ארוך של כרסום בערך המחקר והמידע.
- **האופטימי:** עניין גובר בנתונים ממקורות אמינים ש"שברו" את הכאוס.

- **בחויה שלי, הכלי הנדרש והיעיל ביותר בעת משבר הוא סטטיסטיקה פשוטה, עיתית, הבוחנת התרחשויות יומיומיות, בתדירות גבוהה מאוד.**
למרות מספר רב של ניסיונות במהלך ההתמודדות עם הקורונה – למקבלי החלטות אין סבלנות ו/או זמן ו/או קשב לניתוחים מורכבים, סטטיסטיים או אחרים, בעת המגפה (או כל משבר אחר). רוח הדברים היא: "תגידו את הטיעון ודי". אווירה כזו לא מעודדת חשיבה מורכבת או מעמיקה.
- **תיאוריה, ביקורת וחשיבה מורכבת:** דיון אקדמי ו/או מורכב תוך כדי האירוע הינו חסר סיכוי, לאור השדה שתואר. איזה דיון תיאורטי בכל זאת נוכח שם?...
 - סוגיות של תחום ה"חוסן" היו הרלבנטיות ביותר, ורק לפעמים.
 - עיסוק בסוגיות פרקטיות, אקטואליות מתוך נקודת מבט אינטר-דיסציפלינרית.
 - תיאוריות סוציולוגיות, מדע-מדינתיות, או מתחום מדע התקשורת מורכבות ועוסקות בחברה "מלמעלה" ו"מגבוה" – למעשה חסרות סיכוי לקבל איזשהו קשב בעת משבר.
 - אין קשב לגישה ביקורתית, שנתפסת כ"מעצבנת" ולא תורמת, במיוחד אם היא לא מלווה בהמלצות פרקטיות.
 - אין קשב לנתונים שהם מעבר לבסיסיים ולאמירות מורכבות (מבקשים את החומר באופן "לעוס").
- **נתונים:** אוספים הרבה מאוד נתונים, ולא ממצים את הניתוח בעת המשבר. בעת כתיבת שורות אלה, בסוף המשבר, נראה שזה הזמן לעיבוד מורכב יותר.

דילמות מקצועיות:

- המתח שבין מהירות ורלבנטיות למקצועיות ודיוק.
- המתח שבין הטהרנות הדיסציפלינרית לבין אינטר-דיסציפלינריות. אני סבור שאינטר-דיסציפלינריות תמיד עדיפה.
- לאיזה **סוג עבודות** יש יותר קשב ארגוני ולאילו פחות בעת משבר? איך ניגשים למקבלי החלטות בארגון בעת משבר?
- **חוסר סבלנות למורכבות וביקורתיות בעת משבר.** כיצד ניתן להתגבר על מחסום זה בעת המשבר?... לדעתי לא ניתן. פשטות מנצחת בעת משבר, למקצועיות מיוחסת חשיבות משנית, על כל המשתמע מכך.
- **איסוף נתונים:** הפער בין האיסוף היעיל והמסיבי (ובצבא: כזה שלא עומדים בפניו אתגרים משמעותיים) לניתוח הראשוני הפשטני והלא מעמיק.
- **טכנולוגיה, DS והשפעתם (...?) על האיסוף והניתוח:** בין שיווק ל"אמת".
- מחסור בסיפורי הצלחה בניתוח נתונים מורכב בהקשר של **סקרים בצה"ל** בעת ההתמודדות עם הקורונה. האמנם תמונה זו הייתה רק בצה"ל ורק בכל הנוגע לנתוני סקרים?
- **קיים צורך \ הכרח לבנות את המקום שלנו לפני המשבר,** כלומר: המבנה הארגוני, היכרות מוקדמת ומארג הקשרים בין ארגונים בעת משבר הרבה יותר חזק מאשר הסמכות המקצועית. צריך להתמקם מראש.



שולחן עגול בנושא: ניתוחים סטטיסטיים של נתוני קורונה

מסמך מקדים

ייעוד ותפקידים – מדור מחקר ותודעה

מדור מחקר ותודעה בפיקוד העורף נמצא תחת ענף התנהגות אוכלוסייה במחלקת אוכלוסייה בפיקוד. המדור עוסק במחקר יישומי בתחום מדעי ההתנהגות. רוב המחקר הכמותי, והניתוח הסטטיסטי במדור, נערכים על בסיס סקרי עמדות הציבור הנאספים בשגרה ובחירום. במהלך השנה האחרונה, נוספו מקורות נתונים רבים נוספים ונעשה שימוש נרחב ומגוון יותר בנתונים, לשם תמיכה בתהליכי קבלת החלטות בפיקוד העורף.

ייעוד המדור:

מתן תשתית ידע, מבוססת מחקר, לתמיכה במתן מענה מיטבי לעורף בעת שגרה וחירום. באמצעות ביצוע מחקרים אמפיריים ושימוש בפלטפורמות טכנולוגיות לשם ניתוח נתונים ממקורות מגוונים.

מתודולוגיות עבודה מרכזיות - בשגרה ובחירום

1. סקרים חיצוניים: נערכים ע"י חברת סיקור (דגימת משתתפים ואיסוף נתונים) באמצעות פאנל אינטרנטי או סקרים טלפוניים עבור אוכלוסיות ייחודיות ומנותחים במדור.
2. סקרים פנימיים: נערכים במדור באופן מלא של כלל שלבי ביצוע סקר, ומופצים ע"י הרשות המקומית ישירות באמצעות לינק לתושבים.
3. במקביל למחקר הכמותי, מתקיים גם מחקר איכותני, הכולל: קבוצות מיקוד, ניתוחי רשת ורשתות חברתיות ועוד.
4. אגמי"ם (אגירת מידע ומיצוי נתונים): גוף מודיעיני בפקע"ר שמהווה סביבת אגירה ועיבוד מידע לאומית המורכבת מסנסורים ומקורות מידע רבים, יכולת בינה מלאכותית שמאפשרת חיזוי ותמיכה בקבלת החלטות. כחלק מכך, המדור נמצא בשיתוף פעולה עם אגמי"ם ליצירת שכבות מידע שייסעו לבניית תמונת המצב של התנהגות אוכלוסייה וקבלת החלטות בחירום.

נקודות רלוונטיות לשיח ודיון

מהם האתגרים העיקריים בתהליך איסוף הנתונים ובמינוף המומחיות הסטטיסטית בעת מגפת הקורונה?

- 📍 "אינפו כאוס" – אתגר ההתמודדות עם ריבוי מקורות ונתונים: זיהוי הרלוונטיים ביותר, אתגר החיבור בין קבצי נתונים הבנויים באופן שונה.
- 📍 חתירה לייצוגיות בסקרים פנימיים ברשויות המקומיות (מכסות- ע"פ איזה קריטריון להגדיר את המדדים לייצוגיות המתאימה ביותר לרשות).
- 📍 מתח תמידי בין יצירת השפעה Vs. התעקשות על מתודולוגיה- פער בין צורך בממצאים מידיים ונקודתיים עבור ה"כאן ועכשיו" לבין רצון לבצע מחקר מעמיק הדורש ניתוח סטטיסטי מורכב שדורש זמן ומתייחס לתוצאות רחבות יותר.
- 📍 איסוף נתוני סקרים באופן שיאפשר השוואה בין אירועים מסוגים שונים, כולל היכולת להשוות לאחר.
- 📍 הצגת הנתונים למקבלי החלטות: אתגר ההצגה הפשוטה של נתונים מורכבים, ציפייה של מקבלי ההחלטות לאמירות מובהקות.



מדיניות שמשנתנה לעתים תכופות ; תלות במשאבים.

מהי התועלת הצפויה משימוש בניתוחים סטטיסטיים של נתוני הקורונה?

תמיכה בתהליכי קבלת החלטות בזמן אמת, וניתוח מורכב בדיעבד עבור קבלת החלטות עתידיות.

איסוף מידע על עמדות ותפיסות הציבור בתחומים שונים, גם כאלה המשיקים למצבי חרום אחרים. לדוגמא: למידה מעמיקה על הישמעות הציבור להנחיות עבור אירועים עתידיים (שאינם בריאותיים).

שיפור בתהליכי סיקור ממוקד אוכלוסיות עם מאפיינים ייחודיים (חברה ערבית, חברה חרדית ועוד).

מהם המחסומים העיקריים שכרוכים בניתוח ושימוש בכלי ניתוח מתחום הסטטיסטיקה בזמן טיפול במגפה?

(אי)יכולת השוואה בין שיטות דגימה ואיסוף שונות (טלפוני VS. מקוון; מקוון פאנל VS. מקוון תפוצה חופשית וכו').

צורך בזמן תגובה מהיר- עבור קבלת החלטות בזמן אמת (כצורך ממשי וגם כחלק מתרבות של הארגון הצבאי).

ריבוי גורמים – המציגים מודלים סטטיסטיים אשר עשויים להיתפס כסותרים – עלולים לבלבל את הציבור ואת מקבלי ההחלטות.

מה נחשבים כסיפורי הצלחה כתוצאה מניתוח נתוני קורונה תוך שימוש בכלי ניתוח מתחום הסטטיסטיקה?

ניתוח נתוני למ"ס אודות רשויות מקומיות – מאפייני הרשות ונתונים אוכלוסייה – לשם אפיון אינדיקטורים של פגיעות חברתית. האינדיקטורים צורפו לנתוני התחלואה לשם יצירת מדד המאפשר הערכה של יכולת ההתמודדות הפוטנציאלית של כל רשות, ומכאן לתמיכה בתהליכי קבלת ההחלטות אודות משאבים ותמיכה לרשויות.

מול הרשויות המקומיות והמחוזות: עבודה עם אחריות מקצועית מבוססת נתונים; למידה משותפת של צרכי השטח בניתוח נתונים; יכולת העמקה בהמלצות לרשות המקומית, באמצעות סיוע בתהליכי סיקור ברשויות המקומיות.

נוגה לוין

ראש מדור מחקר ותודעה

פיקוד העורף

4.10 רון קנת וקלליה דיסריו

Some comments on statistical analysis of Corona data¹

Ron S. Kenett¹ and Clelia di Serio²

¹ KPA, Samuel Neman Institute, Technion, Israel and University of Turin, Italy

² Università Vita-Salute San Raffaele, Milan, Italy,

Background

These notes were prepared as background to a round table designed to discuss, with some retrospective, various aspects of statistical analysis and statistical expertise in the Corona pandemic. In considering the topics raised in the round table we would like to distinguish between three related topics:

1. The various stages of statistical intervention
2. The methods and tools for statistical intervention
3. The role of the statistician and biostatistician

The first author has been involved in an analysis of population mobility data and health data from Israel and Italy. The second author is leading the biostatistics group of the San Raffaele hospital system in Italy and is president of the International Biometric Society - Italian Region. These notes reflect on the combined experience of the authors.

Statistical intervention in a pandemic

The Covid-19 pandemic is still considered as an *emergency*. However, one cannot really talk about an emergency, when a health situation lasts for more than one year and includes an evolution of policy strategies, population behavior and scientific knowledge. The role of a medical researcher, in general, has been evolving from handling a first *panic* phase (March 2020), when the search for diagnostic tools, medication approaches and hospitalization procedures was almost *emotional* and based on *trial and error* strategies, until today, when the focus is on understanding the vaccine effect and on planning the best vaccine strategies. In parallel, the role of statisticians and biostatisticians has been evolving to face different challenges and needs of biomedical research, in the hope of mitigating the dramatic impact of COVID-19 on society, the health and the economic systems.

Phase I: Data collection and storage

During the first phase (February 2020-June 2020) statisticians working in the Italian hospitals have been involved in designing and building a biobank for storing COVID 19 data. Initially, this was complicated by the lack of accuracy of diagnostic tools. No unique criteria were provided for defining variables (even the definition of “positive” was not unique and comparable). At that stage, statisticians were coordinating data collection and storage trying to ensure some quality control. In Israel, Prof Gamzu was appointed as COVID19 project coordinator in July 2020. One of his first steps was to reorganize the ministry of health data base. The updated dashboard is available [here](#). Because of politics, Gamzu was not able to implement differential restrictions (the red light system) and restrictions were applied universally, if needed, or not. Overcontrol was also prevalent due to lack of awareness and methods for process control methods such as change point detection.

¹ Prepared as notes for a round table at the Samuel Neman Institute, Technion, 13/4

Phase II: Data analysis

During a second phase (June 2020-November 2020), statisticians in Italy have been devoting their work to analysing clinical data and data from basic research. Basically, this was a very important phase where many statisticians tried to find a modelling approach, and extrapolate from information on epidemiology parameters such as mortality, fatality rate, incidence rate and predictors of disease evolutions. The search for appropriate statistical approaches, in dealing with highly correlated covariates and variables, has been a primary goal to help biomedical research identify COVID19 risks factor. In Israel, physicists and computer scientists lead modeling efforts. See for example a [paper](#) by Eran Segal in Nature communications. Eli Waxman, a professor of Physics at the Weizmann Institute led the national Israeli team that formulated the National Security Council's exit [strategy](#). A call by the Israeli National Statistician, and general manager of the central bureau of statistics, to run a national serological survey to map infection rates was largely ignored. In fact, Seigal Sadeszki, who was the director of public health in the Health Ministry, [stated](#) on TV that testing is not useful as it does not cure people.

Phase III: Design and surveillance

In a third phase, Italian statisticians have been involved in study design (planning new prospective studies) and surveillance studies. This includes i) handling issues on sampling (that is designing a representative statistical sample to follow longitudinally evolution of immunity related factors) ii) identification of individual predisposing factors to define not only surveillance procedures for frail groups of infected people but to develop vaccination strategies and iii) to identify priority groups for vaccination. In Israel, with a world record deployment of vaccinations (Pfizer), a major study by the Clalit healthcare system assessed vaccination efficiency, see [paper](#) by Ran Balicer.

Some retrospective considerations

The COVID19 pandemic created a situation where data driven policies are much needed. This required access to data that is actionable and setting up information systems for collecting, storing and presenting data, and validated predictive models, to support decision makers such as hospital managers, education system administrators and economic system leaders. In Italy, 1400 statisticians and researchers signed a petition to make COVID19 data open access. A special meeting of the Italian statistical society was held to deal with COVID19 [issues](#). The COVID19 pandemic provides statisticians an opportunity to review their role in a wide range of domains. In particular, the effective investigation of disease outbreak is a domain where methods and tools can be improved [1]. On a deeper scientific level, the assessment of evidence should be considered a key competency of statistical work [2]. In a more general context, the growing role of data science needs to be considered [3], perhaps by getting statisticians to focus on a wider picture of information quality [4].

References

- [1] Ronald D. Fricker Jr and Steven E. Rigdon (2020) Investigating a disease outbreak, *Significance*, 17:2, 14.
- [2] Andrew Gelman (2020) Assessing Evidence vs. Truth in the Coronavirus Pandemic, *Chance*, 33:3, 58-60,
- [3] Ron S. Kenett and Thomas C. Redman (2019) *The Real Work of Data Science: Turning data into information, better decisions, and stronger organizations*, Wiley.
- [4] Ron S. Kenett and Galit Shmueli (2016) *Information Quality: The Potential of Data and Analytics to Generate Knowledge*, Wiley.

בריאות



מוסד שמואל נאמן
למחקר מדיניות לאומית

טל. 04-8292329 | פקס. 04-8231889
קרית הטכניון, חיפה 3200003
www.neaman.org.il