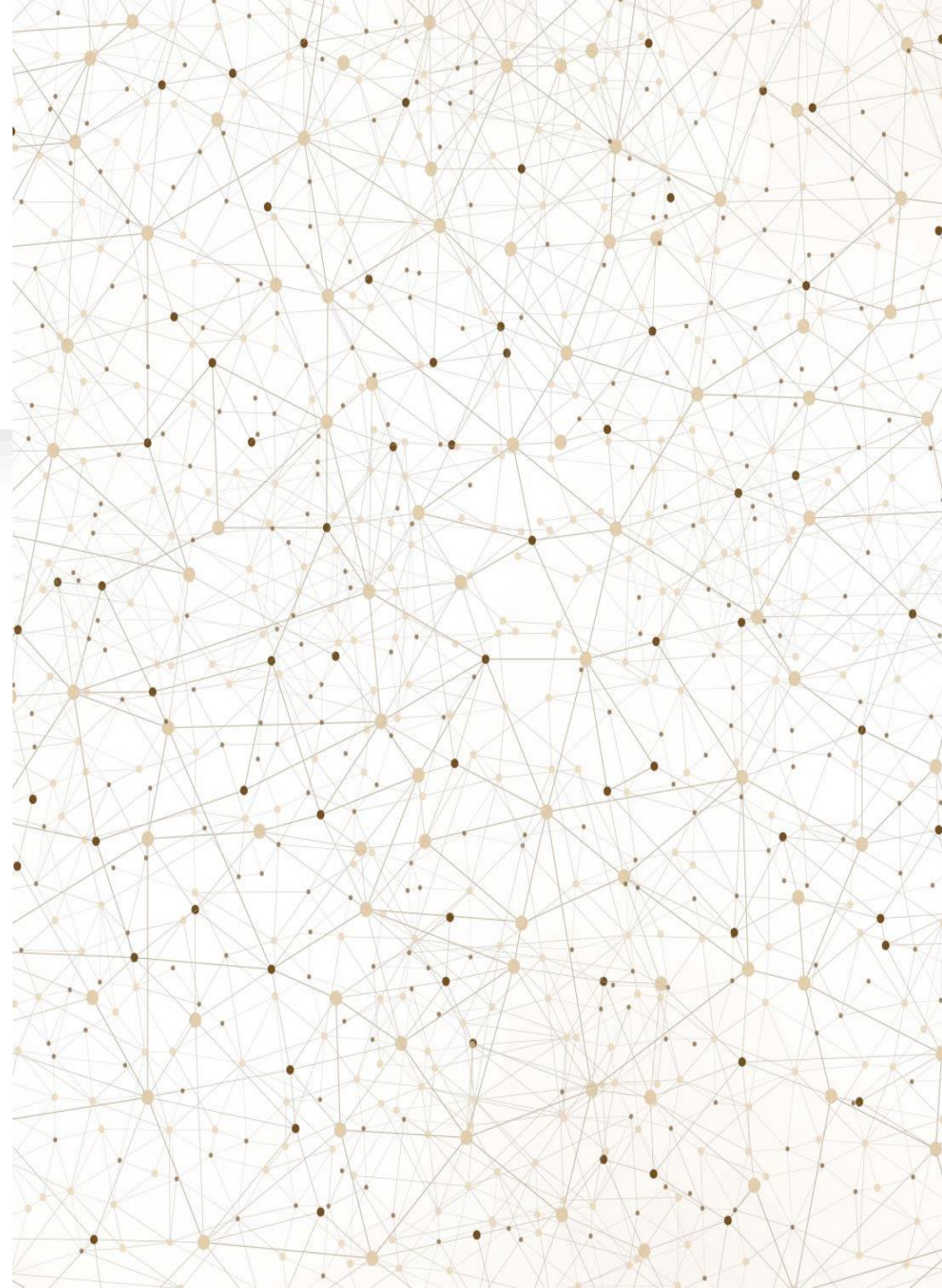


נתונים במחקר שיתופי פעולה עם גורמים חיצוניים

ד"ר אליסה לוברט
I-Metadata AI Center
סוקרת ועדת נתונים

אפריל 2024



מה בתכנית?

- מהי ועדת נתונים
- התממת נתונים
- מה בודקים בוועדת נתונים כתלות בסוג המחקר
- MDCClone

מחלקת מדעי הנתונים והבינה המלאכותית

- פיתוח מערכות תומכות החלטה מבוססות AI
- פיתוח אפליקציות על בסיס פלטפורמת נתונים חדשנית בענן
- שיתופי פעולה בתחום הדאטה עם שותפים חיצוניים
- שליפת נתונים והתממת קבצים מיוחדים
- MDCLONE
- פיתוח כלי NLP



מהי ועדת נתונים?

הגוף המקצועי הסוקר מחקרים העושים שימוש משני במידע בריאות

תפקידי הוועדה

שמירה על פרטיות המטופלים, בדגש על המחקרים בהם עלולה להיות פגיעה בפרטיותו של נושא המידע (שת"פ עם גורמים חיצוניים)

הועדה סוקרת:

- **מחקרי נתונים**
- **מחקרי שאלונים**
- **מחקרים ללא מוצר מחקר – הכוללים העברת נתונים או אלגוריתמים**
- **Amendments למחקרים מאושרים**

הועדה ממליצה על תיקונים בכל הקשור לסוגיות התממה, איסוף הנתונים, שמירת הנתונים, פלטפורמות להנגשת או העברת הנתונים.

הרכב ועדת הנתונים

- חברות ועדת הנתונים נמנות על חברי ועדת הלסינקי המוסדית
- לפי נוהל שימוש משני, חברי הועדה צריכים לכלול מומחים בתחום ניתוח המידע, בעלי ידע בסטטיסטיקה, המסוגלים להעריך רמת סיכון לזיהוי מידע ואיכות תהליכי התממת נתונים.
- חברות הוועדה הן נציגות מהיחידה למדעי הנתונים והבינה המלאכותית
- מנהלת ועדת הנתונים היא גב' ליאת נדאי ערד, מנהלת מחלקת מדעי הנתונים
- חברות הועדה סוקרות את המחקרים ומעבירות המלצותיהן לוועדת הלסינקי

כמות הנתונים במרכז הרפואי

CASES 17.5M+

PATIENTS 2.5M+	VITAL SIGNS 75M+		MEDICATIONS 45M+
LAB RESULTS 150M+	PATHOLOGY REPORTS 750K+	MICROBIOLOGY PANEL 3M+	
PROCEDURES 550K+	SURGERIES 450K+		DIAGNOSES 18M+
ULTRASOUND 280K+	CT 690K+	MRI 370K+	X-RAY 2M+

*הנתונים מעודכנים נכון לדצמבר 2023

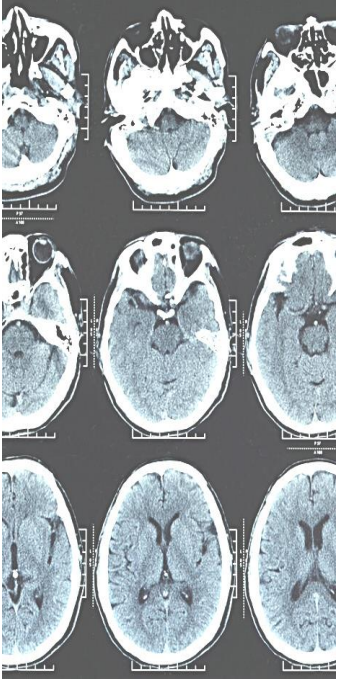
סוגי נתונים במרכז הרפואי

מידע מובנה

test code	Lab Test
82803159	Gases panel - arterial blood
84132150	Potassium - arterial blood
84132010	Potassium - blood
84132140	Potassium - whole blood
03556010	HCO3 (Bicarbonate) - venous b
03556140	HCO3 (Bicarbonate) - whole bl
03556150	HCO3 - arterial blood
03553140	pH - whole blood
03553010	pH - venous blood
03553150	pH - arterial blood
3555140	pO2 (Partial pressure of oxyge
3555570	pO2 (Partial pressure of oxyge
3555150	pO2 - arterial blood
3555010	pO2 (Partial pressure of oxyge
03554570	pCO2 (Partial pressure of carl
03554150	pCO2 - arterial blood
03554010	pCO2 (Partial pressure of carl
03554140	pCO2 (Partial pressure of carl

בדיקות מעבדה
דמוגרפיה
תרופות
אבחנות

תמונות



CT
MRI
US
Photo

טקסט חופשי

2 שנים, החלה אפינטור Q4-2015 בשל אנגימיילפומ
גון מוסתר]. לבבית אקו לב Q3-2016 תקין. ביצעה
תחת הטיפול. סוגר כליות 2 כליות קטנות עם פרנכימי
זאינטרסטיאלית. הומלץ להגיע למעקב חצי שנה ענ
ביקורת קרדילוג Q4-2018 - א.ק.ג. קצב סינוס. צ
ריביאלי, MR קל. לא הודגמו רבדומימות באופן ברור
ם מוסתר] - תקין 180-56, ממוצע 85. ללא פעילות
ניים בקרדילוגית מבוגרים. עשתה Q2-2018 סוגר בנ
רגיל, פרנכימה אטרופית ואקוגנית, מוקד אקוגני 3.6
נפרוזיס. דופלר כליות ללא עדות עקיפה להפרעה הפ
Q2: בסוף CT Q3-2016 ראות רחלוציה של הממ
זוגר שד בריאות הצללה עדינה באספקט עליון LLL ל
A. מוקדים דומים בבד ללא שינוי. שינויים בחוליות יצ
של החדר השמאלי, ללא שינוי בהשוואה לבדיקה קו
יות, ללא שינוי בהשוואה לבדיקה קודמת. ללא עדות
יב האדרתי בולט יותר בנגעים בשד הימני. אנגימייל
לא שינוי. היו אצל כירורג שד בקופת חולים. Q4-2018
ב Q3-2017 מטופלת במינרין ומאוזנת. נתן תקין
ש. Q2-2019 ביקורת ד"ר [שם מוסתר] - המשך מעק
:- הולטר ל"ד סה"כ תקין, לחץ דם ממוצע 125/76,
נפרוזיס, זרימות תקינות, שלפוחית שתן תקינה, קיב
ת לסף רגישות, חלבון לקראטנין 0.1, אין חלבון בני
ת דם Q3-2019 - גלוקוז 75, UREA 18.3, קראטנין

סיכומי אשפוז
פענוחי הדמיות/פתולוגיה

אחר



סרטי ניתוח
קבצי קול
EEG
ECG

שיטות התממה

ערבול (hash)



קידוד (Pseudonymization)



הרעשה (perturbation)



השמטה (suppression)



הזחה



מיסוך (masking)



התמרה (permutation)



דגימה (sampling)



הכללה (generalization)
אגרגציה (aggregation)



שיטות התממה – שדות מובנים

דמוגרפיה

בהתאם לסוג המידע

- קידוד
- הכללה
- התמרה

תאריכים

הכללה (שנה, רבעון,
חודש)

המרה (במספר המיצג
נקודת ייחוס - מספר
ימים מ...)
הזחה (בקבוע)

אזור גאוגרפי

הכללה (מחוז, עיר,
מיקוד)

פרטים מזהים

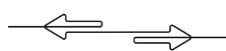
השמטה של כל ערך בר
זיהוי (שם ושם משפחה,
ארגונים...)

קידוד תעודת זהות או
מספר אשפוז להחליף
במספר סידורי+רשימה
מקשרת

שיטות התממה – טקסט חופשי

[שם מוסתר] – יליד 1998, [מצב משפחתי מוסתר], הופנה להערכה נפרולוגית בעקבות אבחנה קלינית של Tuberous Sclerosis Complex / TSC מגיל הילדות ללא בירור מולקולרי גנטי וצילום בקרב בנות/בני המשפחה החולים. במקרה אנמנסטית ובדיקה גופנית, המטופל אכן עונה על קריטריונים קליניים של TSC. לאור זאת, הוסבר כי מומלץ על יעוץ גנטי עם בירור מולקולרי יעודי. כמו כן, הוצג לו האופק של מעורבות נפרולוגית במקרה של TSC, כולל מחלת כליות ציסטית / יתר לחץ דם ופאוכרומוציטומה / AML / ממאירות כליתית. כמו כן, הוסבר אודות דגם הורשה AD של TSC, אשר בעטיו קיים סיכון של 50 אחוז להישנות בכל הצאצאים של [שם מוסתר]. הוסבר בדבר האפשרות למניעת הישנות במסגרת הליך אבחנתי טרום לידת / טרום השרשתי. בדיקה גופנית: לחץ דם 130/85 מ"מ"כ; עור – ריבוי Angiofibroma בפנים עם Cephalic Plaque מימין; שיניים עם Pits; ללא מעורבות ציפורניים או נגעים בולטים בבדיקות העור; לב-קולות לב סדירים ללא אוושות ברורות. רקע משפחתי: # אב [טקסט מוסתר] – מחלת כליה ספוגית על רקע Amyloidosis. # אחות של סבתא אמהית – אפילפסיה. # אב [טקסט מוסתר] TSC ואפילפסיה; NET בטחול ולבלב. # דודים / דודות אמהיות – לסלקס אבחנה של TSC. # בן דוד אמהי – TSC; אפילפסיה. # סב אמהי – TSC; שחמת כבד; Multiple Myeloma. בדיקה מעבדה ודימות: # קריאתני 0.64 מג' לדל' ללא הפרעה אלקטרוליטארית נלוות; כולסטרול גבולי. # ספירת דם – המוגלובין 16.3 גרם לדל' עם MCV 87 ללא ציטופניה נלוות. # הולטר לחץ דם Q1-2021 – ממוצע כולל 120/69 מ"מ"כ; ממוצע בשעות ערות 125/72 מ"מ"כ; ממוצע בשעות השינה – 102/56 מ"מ"כ. # סונר בטן Q1-2021 ([ארגון מוסתר]) ממצא היפראקוגאי בקוטב עליון של כליה שמאל. לא ניתן לשלול בוודאות אנגיomiוליפומה או ארטיפקט בלבד. # MRI Q1-2021 – כליות תקינות בגודל ובמיקום מרכזות סמכותיות ללא הידרונפרוזיס. ללא ממצא האדירתי חשוד טללא ממצא ושמני החשוד ל-AML; הסננה שומנית דיפוזית. # MRI מח Q1-2020 ([ארגון מוסתר]) – נודולים סבאפנדימלים בהיקף החדרים הלטרלים כולל מוקדים ב[שם מוסתר]יים הפרונטאליות בסמוך לאזור פורמן ע"ש מונרו דו"צ, האדירה של התהליך ב[שם מוסתר] הפרונטאלית מימין. מוקדי אות גבוה ברצפי T2 בחומר הלבן דו"צ תת קורטיקלים, ללא שינוי - יכולים להתאים להמרטומות של טוברוס סקלרוזיס. # בדיקת עיניים מ-2021 – ללא עדות לממצא ברשתית. # [טקסט מוסתר] ב [ארגון מוסתר]. # השכלה – בגרות מלאה; קורס מקצועי בתחום המחשבים. במהלך בית הספר עבר אבחון פסיכודידקטי בעקבות קשיים לימודיים. המלצות: # יעוץ גנטי ובירור מולקולרי על ידי ריצוף הגנים TSC1/2 עם בדיקת חסרים ותוספות; יעוץ גנטי בכל הריון ובתכנון הריון. # MRI כליות כל 1 שנה. # HRCT של הריאות, במידה ולא בוצע בעבר לבירור בסיסי של LAM וציסטות ריאתיות. # אקו לב עם דופלר אארוטה. # ביוכימיה כל 1/2 שנה – תפקודי כליות, אלקטרוליטים, רבבות זרחן, סידן, מגנזיום, חומצה אורית; תפקודי כבד עם אלבומין; TSH. # ספירת דם כל 1/2 שנה; חומצה פולית, B12. # בדיקת שתן כל 1/2 שנה – כללית, יחס חלבון לקריאטנין. # מעקב רב מערכתי במרפאות, כולל: נירולוגיה / עיניים / שיניים / עור, כולל לבירור האפשרות לטיפול מקומי במעכבי mTOR. # הערכה דיאטטית לאור כבד שומני; העשרת כלכלה בכורכום / קפה / שמן זית. # המנעות מחשיפה לתכשירים נפרוטוקסיים / NSAIDs / חוקנים עתירי זרחן / חומר ניגוד על בסיס יוד. # מעקב נפרולוגי בעוד 1/2 שנה. # הפנית כל בנות/בני המשפחה החולים להערכה נפרולוגית ויעוץ גנטי; בכל הריון ובתכנון כל הריון, מומלץ על יעוץ גנטי קדם. [שם מוסתר]

קידוד (Pseudonymization)



התמרה (permutation)

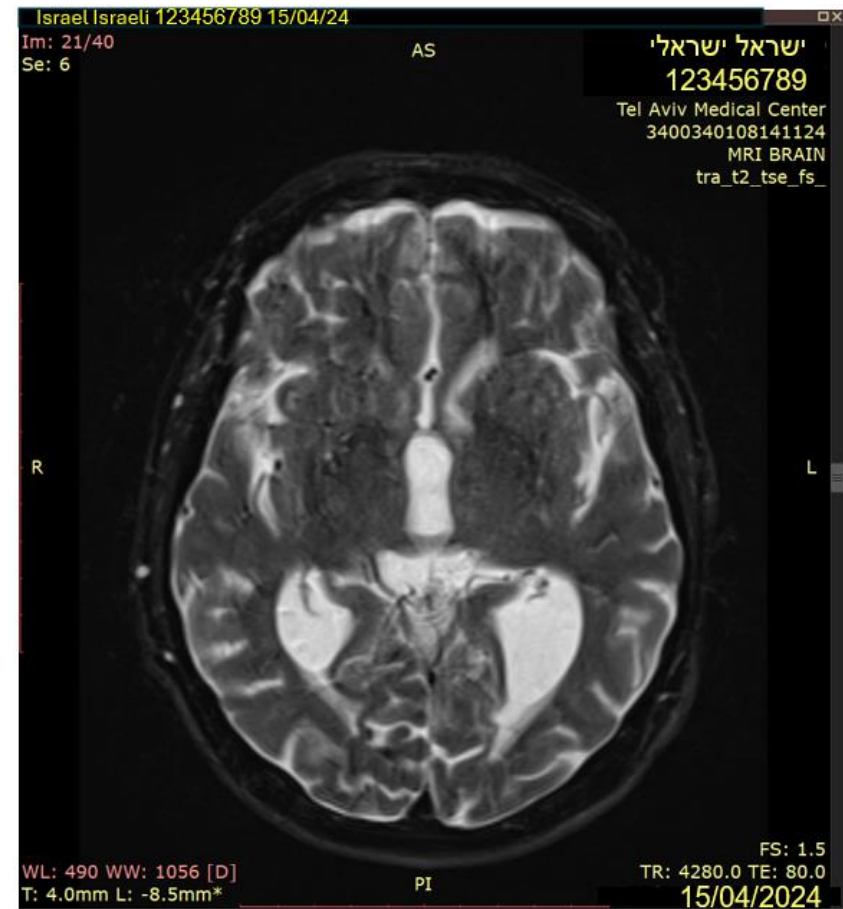
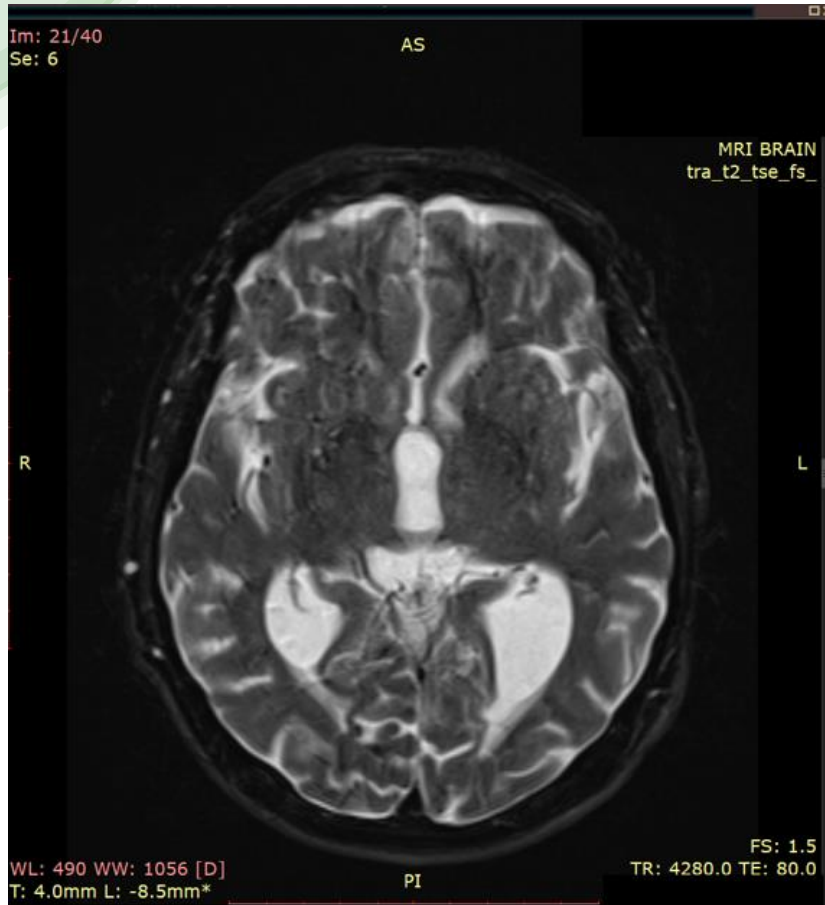


מיסוך (masking)



*** "לא ניתן למסור שדות מידע לא מובנה, אלא אם ניתן להתמימם באופן מספק, להנחת דעתה של הוועדה לשימוש מחקרי במידע..."

שיטות התממה – תמונות



השמטה (suppression)



מה נבדק בועדת נתונים?

דרכי השמירה על פרטיות המשתתפים וסודיות המידע שנאסף, כולל הגורם האחראי על שמירת המידע, דרך העברתו, אופן שמירתו, משך השמירה וכו'



מחקר פנימי

שת"פ חיצוני הכולל
הנגשת נתונים

שת"פ חיצוני הכולל
העברת נתונים

מחקר פנימי

חוקרים פנימיים בלבד

**מהם כלל הנתונים
הנאספים?**

לציין נתונים מיוחדים
(תמונות, סריקות, קבצי קול,
נתונים ביומטריים..)

**אופן התממת הנתונים?
(כולל נתונים מיוחדים)**

מקורות המידע

(קמיליון, נמ"ר, PACS...)

**מהן השיטות
הסטטיסטיות?**

אם הסטטיסטיקאי חיצוני,
מדובר במחקר שת"פ

**האם נוצרה רשימת
קידוד?**

מי אחראי על יצירת ההתממה,
מי מחזיק במפתח הקידוד?

**במידה ויש איסוף
נתונים פרוספקטיביים
– לוודא שיש טופס
החתמה על הסכמה
מדעת**

היכן יאוחסנו הנתונים?
לציין נתיב לשמירת הקובץ
אין לשמור נתונים על המחשב
האישי, דיסק און קי, מחשב נייד,
עננים פרטיים

מחקר פנימי

Amendment – המחקר יעבור סקירה מחודשת בידי ועדת הנתונים במקרים:

•מחקר פנימי בו מבוקשת הנגשת/העברת נתונים לגורם חיצוני

שימוש ב – ChatGPT

•ניתן לבצע את המחקר ב-GPT4 דרך הענן המאובטח של איכילוב במייקרוסופט

•כרוך בהסדרה של סעיף תקציבי מול מחלקת המחשוב והדרכה על אופן השימוש

•הנתונים שיועלו יהיו מותממים ולא יכילו מידע מזהה בטקסט שיוזן לתוך המערכת.

•אין לעשות שימוש מחקרי, הכולל הזנת נתונים רפואיים של מטופלים ל ChatGPT בגירסה

החינמית בדפדפן

מחקר עם שת"פ חיצוני – הנגשה

מתן גישה לגורם חיצוני לצפייה ועיבוד של נתונים מותממים, ללא יכולת הורדת הנתונים למחשב חיצוני לארגון

קרי: אין העברה של קבצי נתונים מחוץ לכתלי הארגון



מחקר עם שת"פ חיצוני - הנגשה

החוקרים הניגשים לנתוני המקור הם פנימיים בלבד

מהי מהות שיתוף הפעולה?
הנגשת או מסירת נתונים?

מי מרכז את הנתונים ומבצע את האנליזה?
כמה מרכזים משתתפים במחקר?

מקורות המידע בהם ייעשה שימוש?
(קמיליון, נמ"ר, PACS...)

מהם כלל הנתונים הנאספים?
האם יש נתונים מיוחדים?

אופן התממת הנתונים
באחריות החוקר הראשי

מהי הפלטפורמה להנגשה?
אחת מתוך המאשרות

מהן השיטות הסטטיסטיות?

במידה ויש איסוף נתונים פרוספקטיביים – החתמה על טופס הסכמה מדעת

ההנגשה תתאפשר בכפוף לחוזה

פלטפורמות מאושרות להנגשה

- **חדר מחקר מאובטח בענן AWS** שבבעלות ביה"ח
- מאפשר העלאת דאטה והקמת מכונות לניתוח הנתונים (R, SPSS, פייתון, למידת מכונה)
- **REDCap** - של ביה"ח ובתנאי שההגדרות לא מאפשרות הורדת הנתונים למחשב חיצוני
- **באחריות החוקר הראשי** להתאים את ההרשאות כך שבכל נקודת זמן רק מורשים שעובדים במחקר יהיו בעלי גישה.

Basic Privileges	
Expiration Date (if applicable)	(D/M/Y)
Highest level privileges:	
Project Design and Setup	☐
User Rights	☐
Data Access Groups	☐

Privileges for Viewing and Exporting Data							
Data Viewing Rights pertain to a user's ability to view or edit data on pages in the project (e.g., data entry forms, reports). Users with 'No Access' Data Viewing Rights for a given instrument will not be able to view that instrument for any record, nor will they be able to view fields from that instrument on a report. Data Export Rights pertain to a user's ability to export data from the project, whether through the Data Exports page, API, Mobile App, or in PDFs of instruments containing record data. Note: Data Viewing Rights and Data Export Rights are completely separate and do not impact one another.							
	Data Viewing Rights			Data Export Rights			
	No Access (Hidden)	Read Only	View & Edit	No Access	De-Identified*	Remove All Identifier Fields	Full Data Set
My First Instrument	☐	☒	☐	☒	☐	☐	☐

* De-identified means that all free-form text fields will be removed, as well as any date/time fields and Identifier fields.

פלטפורמות מאושרות להנגשה

- **חדר מחקר מאובטח בענן AWS** שבבעלות ביה"ח
 - מאפשר העלאת דאטה והקמת מכונות לניתוח הנתונים (R, SPSS, פייתון, למידת מכונה)
- **REDCap** - של ביה"ח ובתנאי שההגדרות לא מאפשרות הורדת הנתונים למחשב חיצוני
 - באחריות החוקר הראשי להתאים את ההרשאות כך שבכל נקודת זמן רק מורשים שעובדים במחקר יהיו בעלי גישה.
- **OneDrive** - של ביה"ח ובתנאי שההגדרות לא מאפשרות הורדת הנתונים למחשב חיצוני
- פלטפורמות נוספות ייבחנו מול הממונה על אבטחת מידע (גב' עדינה סולומון)

גב' אביטל סילברסטון מצוות Cloud-Devops מאגף מערכות מידע ודיגיטל תתמוך בבקשות החוקרים בסוגיות הנגשת נתונים avitalw@tlvmc.gov.il

מחקר עם שת"פ חיצוני - העברה

מסירת נתונים לגוף חיצוני למרכז הרפואי - הוצאת נתונים מחוץ לכתלי הארגון
כולל העברת תמונות, הדמיות, טבלת נתונים, הזנת נתונים בeCRF, REDCAP, שאלונים וכו'



מחקר עם שת"פ חיצוני - העברה

ככלל, מדיניות בית החולים הינה הנגשת נתונים ויש להימנע ממסירת נתונים ככל האפשר

ההצדקה למסירת הנתונים צריכה להיות מנומקת היטב, למשל:

- השתתפות במחקר רב מרכזי
- מומחיות ייחודית של הגורם החיצוני בביצוע האנליזה (עיבוד תמונה, AI, סטטיסטיקאים מובחרים וכו')
- איסוף נתונים ממספר מרכזים רפואיים על מנת לייצר מדגם בעל ערך סטטיסטי
- מחקר העוסק בכמות קטנה של מטופלים הסובלים ממחלה נדירה
- יצירת מאגר נתונים בינלאומי עבור חקר מחלה ספציפית
- יש לוודא כי האפשרות של הנגשת הנתונים נבחנה ונמצאה כלא מתאימה, בטרם הוחלט על מסירת הנתונים

מחקר עם שת"פ חיצוני - העברה

**כיצד מתבצעת
ההתממה?**

כולל הנתונים המיוחדים

**מהם כלל הנתונים
הנמסרים?**

(טבלאי, תמונות, סריקות, ביומטרי,
מידע רגיש..)

**מיהו הגורם אליו מועברים
הנתונים?**

מי מרכז את המחקר

**היכן יישמרו הנתונים
וכיצד יאובטחו במוסד
החיצוני?**

**באיזו פלטפורמה יועברו
הנתונים?**

מהי כמות המטופלים?

מעל 100 מטופלים יש
להעביר לאישור של מנהלת
וועדת נתונים

**מסירת הנתונים תהיה
בכפוף לחוזה**

לא תהיה העברת נתונים
לצד ג'

**מי יהיו בעלי הרשאות גישה
לנתונים במוסד החיצוני?**

מחקר עם שתפ חיצוני - העברה

- הנתונים יהיו מותממים – אין להוציא נתונים מזוהים או ברי זיהוי מחוץ לכותלי בית החולים
- במידה וקיימת רשימה מקשרת – תישמר בידי החוקר הראשי
- **במחקרי "ללא מוצר"** – לרוב מדובר במחקרים בהם מתבצעת מסירת דגימות של מטופלים ואיסוף נתונים פרוספקטיבי, בכפוף לטופס הסכמה מדעת.
- במידה ובנוסף לנתונים הפרוספקטיביים יבקשו לבצע איסוף נתונים רטרוספקטיבי מתוך תיקי המטופלים, יש להתייחס לנתונים אלו כאל "מחקר עם שת"פ חיצוני – העברה".
- **במחקר שאלונים** בו השאלונים העוברים לגורם חיצוני – לוודא שמירת שאלונים בצורה מותממת, ללא תעודת זהות, שם, טלפון, מייל וכו' בגוף השאלון.
- יש לוודא שהפלטפורמה של השאלונים קיבלה אישור של הממונה על אבטחת מידע
- **מסירת נתונים לסטטיסטיקאי חיצוני**
- יש לבחון אפשרות של הנגשת נתונים
- ניתן לשלוח נתונים מותממים בלבד
- **Amendment** המבקש להגדיל את כמות הנתונים המועברים ייבחן בידי חברות הועדה

פלטפורמות מאושרות למסירת נתונים

OneDrive •

- יחליף את המייל המאובטח, מאפשר שליחת נתונים בנפח גבוה, מאובטח בידי מחלקת המחשוב

REDCap •

- עם הרשאות להורדת נתונים לגורמים חיצוניים

- eCRF של יזם המחקר אליו מוזנים ישירות נתונים מותממים באופן ידני

- **במחקרי שאלונים**, בתנאי שהנתונים המוזנים אינם מכילים מידע מזהה של המטופלים:

Microsoft forms •

גב' אביטל סילברסטון מצוות Cloud-Devops מאגף מערכות מידע ודיגיטל תתמוך בבקשות החוקרים בסוגיות הנגשת נתונים avitalw@tlvmc.gov.il

MDCLONE



מייל לפניות:

MDCclone@tlvmc.gov.il



ניתן לייצא את
הנתונים
לשותף חיצוני

בכפוף לאישור
העברת נתונים
מועדת
הלסינקי



חוקר חיצוני
יכול לקבל
הרשאת גישה
לשרת

בכפוף לאישור
הנגשה מועדת
הלסינקי



מורידים את
הנתונים
כאקסל
בתיקייה
האישית
בשרת

הסביבה
בשרת מכילה
R, פייתון,
SPSS



כשרואים שיש
את הנתונים
הרצויים פונים
לוועדת
הלסינקי
לקבלת אישור
מחקר

גישה לנתוני
מקור
מותממים



המערכת
נמצאת
בגישה
באמצעות
VPN
Remote
desktop



בדיקת
היתכנות
למחקר על
נתונים
סינתטיים
(בטרם קיים
אישור
הלסינקי)



פלטפורמה
לגישה
למידע קליני
בסביבה
מאובטחת.
מכילה מידע
מעולמות תוכן
מגוונים
שמתעדכנים
באופן תדיר

תודה

שאלות?

