Replica Analytics

AN AETION COMPANY

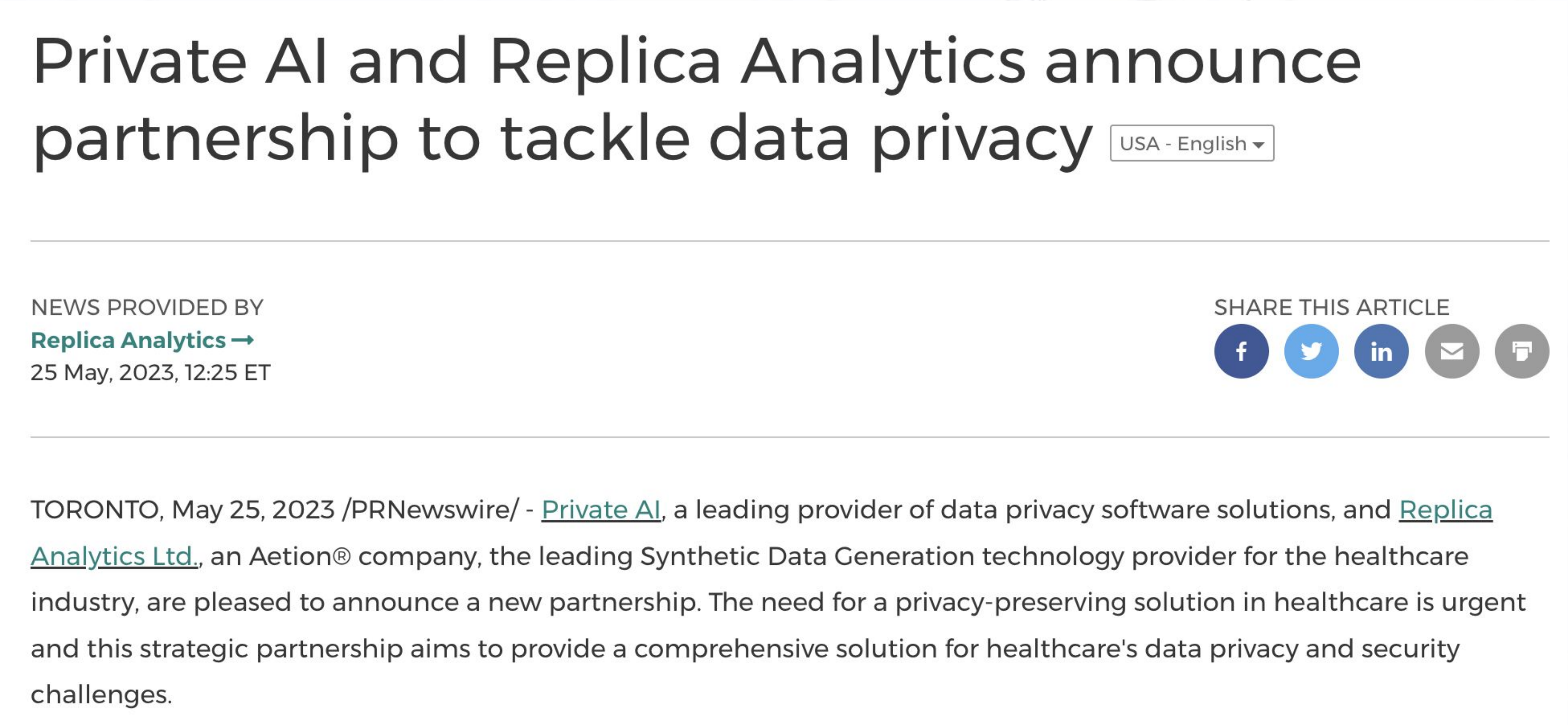
PRIVATE AI

DEMYSTIFYING HIPAA: TECHNIQUES FOR DE-IDENTIFYING COMPLEX MULTI-MODAL DATA

Presented by:

Lucy Mosquera
Sr. Director of Data Science,
Replica Analytics

Patricia Thaine
Co-Founder & CEO,
Private AI



Private AI and Replica Analytics announce partnership to tackle data privacy

NEWS PROVIDED BY
Replica Analytics →
25 May, 2023, 12:25 ET

TORONTO, May 25, 2023 /PRNewswire/ - Private AI, a leading provider of data privacy software solutions, and Replica Analytics Ltd., an Aetion® company, the leading Synthetic Data Generation technology provider for the healthcare industry, are pleased to announce a new partnership. The need for a privacy-preserving solution in healthcare is urgent and this strategic partnership aims to provide a comprehensive solution for healthcare's data privacy and security challenges.

PRIVATE AI

Replica Analytics
AN AETION COMPANY

Agenda

- Introduction to de-identification under HIPAA
- Safe harbor
- Expert determination
- Multimodal data
- Case study of multimodal data de-identification
- Questions

De-identification under HIPAA

PRIVATE AI

Replica Analytics

AN AETION COMPANY

“Health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual” - HSS

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/

De-identification under HIPAA

The process of de-identification, by which identifiers are removed from the health information, mitigates privacy risks to individuals and thereby supports the secondary use of data for:

- Comparative effectiveness studies
- Policy assessment
- Life sciences research
- & other endeavors

Examples of clinical trial datasets using anonymized data for the purpose of sharing data for research

- GlaxoSmithKline trials repository
- Project Data Sphere
- Yale University Open Data Access project
- Immport Immunology Database and Analysis Portal

Health Insurance Portability and Accountability Act

American legislation enacted in 1996 required the Department of Health and Human Services (HHS) to issue privacy regulations governing individually identifiable health information

U.S. Department of
Health and Human Services
Enhancing the health and well-being of all Americans

Standards for Privacy of Individually Identifiable Health Information (or the "Privacy Rule"), was finalized after two rounds of revision and public comment by 2000

This act applies to health plans, health care clearinghouses, and to any health care provider who transmits health information in electronic form

Protected Health Information

The Privacy Rule protects all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. The Privacy Rule calls this information "protected health information” (PHI).

"Individually identifiable health information" is information, including demographic data, that relates to:

- the individual's past, present or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual,
- and that identifies the individual or for which there is a reasonable basis to believe it can be used to identify the individual.13 Individually identifiable health information includes many common identifiers (e.g., name, address, birth date, Social Security Number).

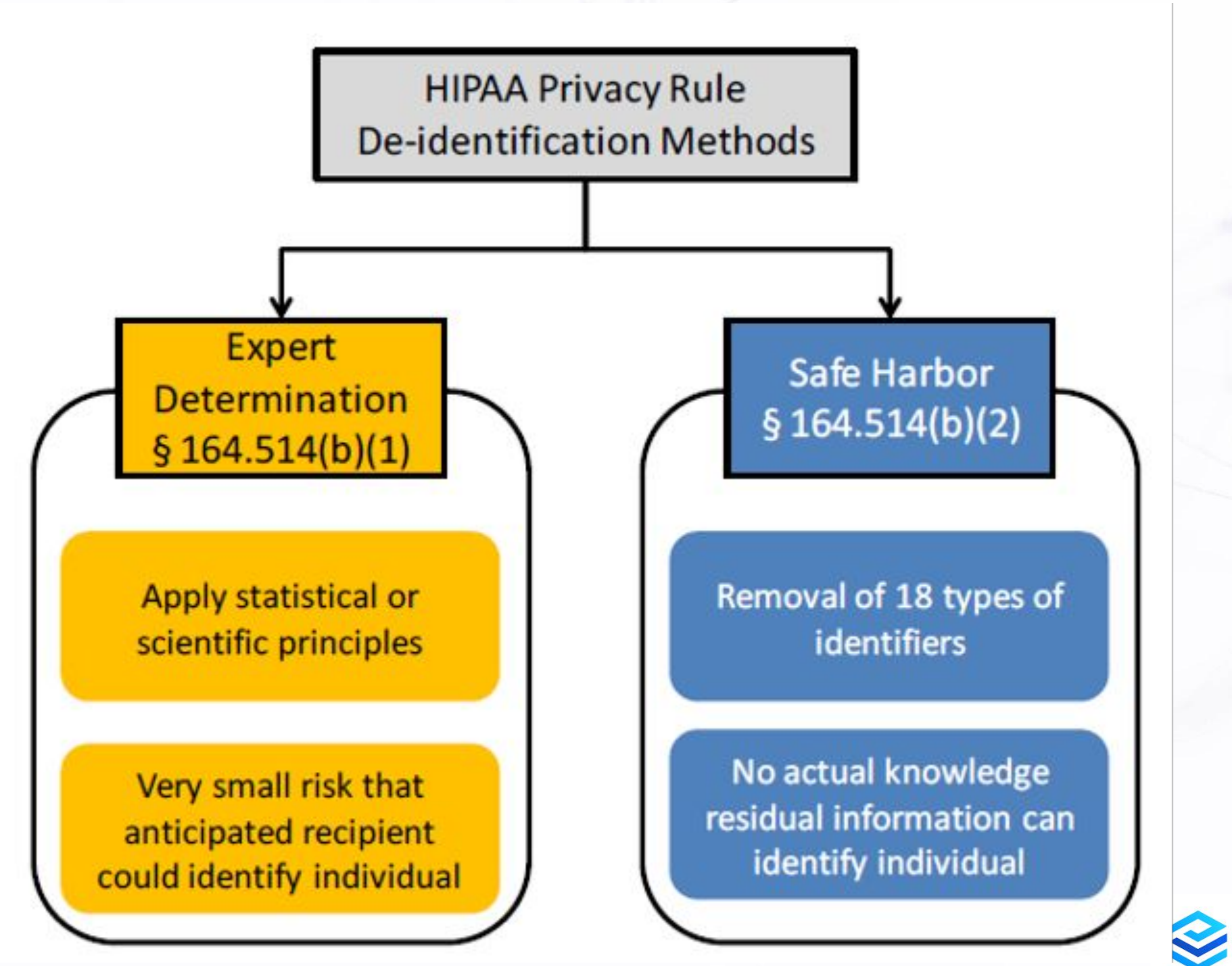
Why de-identify?

Health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual

The process of de-identification, mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors

De-identification allows responsible re-use of health data

Two paths to de-identification under HIPAA:



HIPAA Privacy Rule
De-identification Methods

Expert Determination
§ 164.514(b)(1)

Apply statistical or scientific principles

Very small risk that anticipated recipient could identify individual

Safe Harbor
§ 164.514(b)(2)

Removal of 18 types of identifiers

No actual knowledge residual information can identify individual

PRIVATE AI

Replica Analytics
AN AETION COMPANY

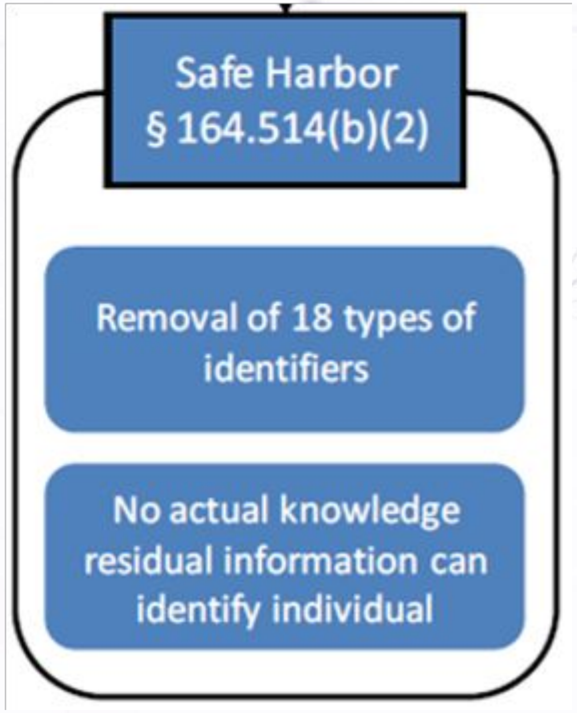
Safe Harbor

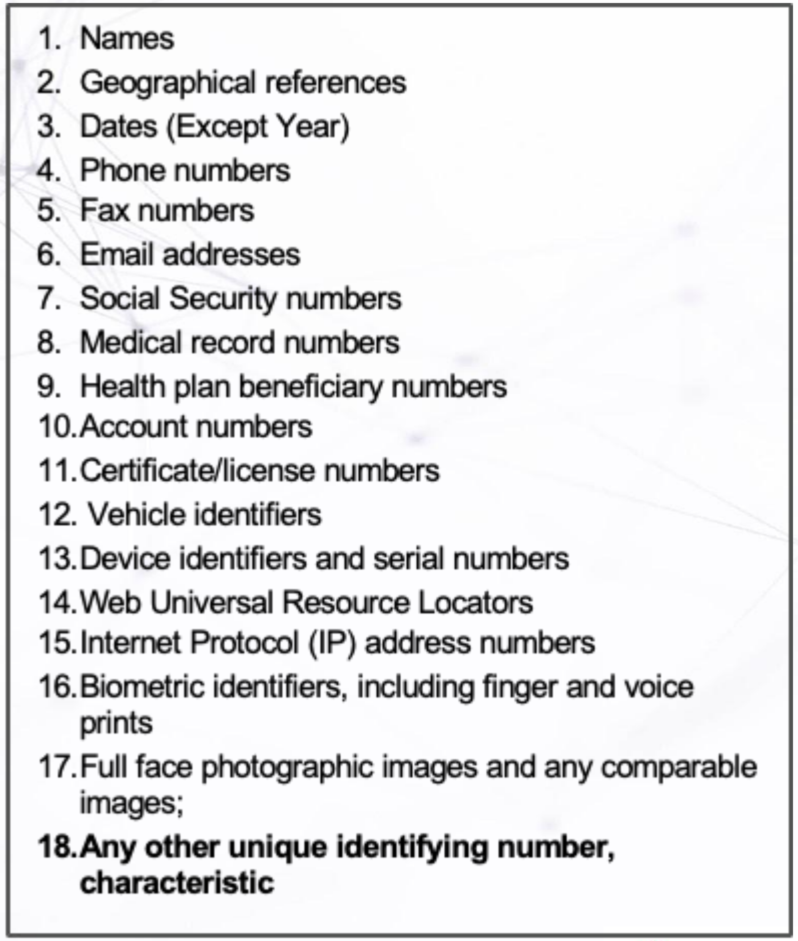
PRIVATE AI

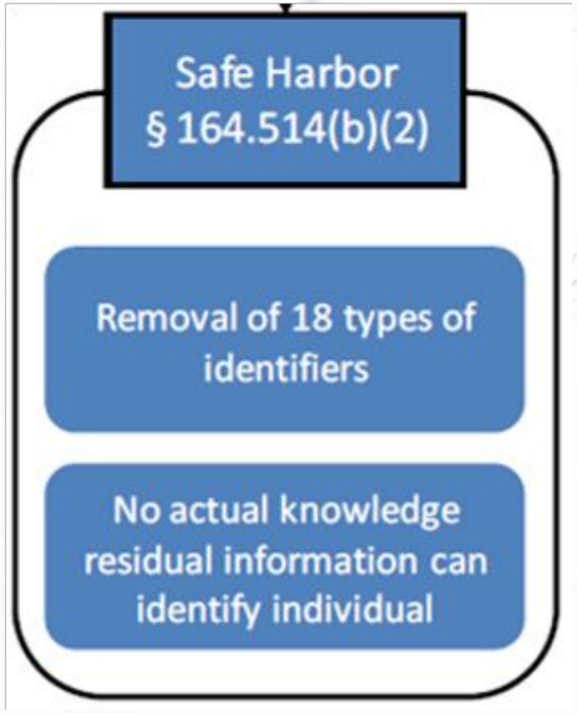
Replica Analytics

AN AETION COMPANY

Removal of 18 types of identifiers to mitigate re-identification risk


Safe Harbor
§ 164.514(b)(2)
Removal of 18 types of identifiers
No actual knowledge residual information can identify individual


1. Names
2. Geographical references
3. Dates (Except Year)
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social Security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images;
18. **Any other unique identifying number, characteristic**


Safe Harbor
§ 164.514(b)(2)
Removal of 18 types of identifiers
No actual knowledge residual information can identify individual

Actual Knowledge
Clear & direct knowledge means that remaining information could be used either alone or in combination with other information to identify an individual.

Example 1: Revealing Occupation
Former president of the State University coupled with age range

Example 2: Clear Familial Relation
Researcher may have family member in dataset.
Researcher may know many more details like the sequence of events

Example 3: Publicized Clinical Event
Publicized that patient had quintuplets in small town. This may be the only occurrence of quintuplets in that small town.

Example 4: Knowledge of a Recipient's Ability
Knowledge that the anticipated recipient of the data has table or algorithm to help id.

Example 5: Rare Diagnosis
Knowledge that there is someone with a rare diagnosis in the data

Safe Harbor

Advantages

- Easy to implement with relatively low privacy expertise
- Easy to automate for large structured datasets
- Recognized by regulators in the USA
- Objective criteria (except for the ‘no actual knowledge’ component)

Disadvantages

- May reduce utility of the data for health researchers (e.g., removal of dates may make time to event analyses impossible)
- May not sufficiently reduce risk for complex datasets
- May be overly restrictive for small samples of data
- Difficult to automate for multi-modal data (unstructured, images, audio, etc.)

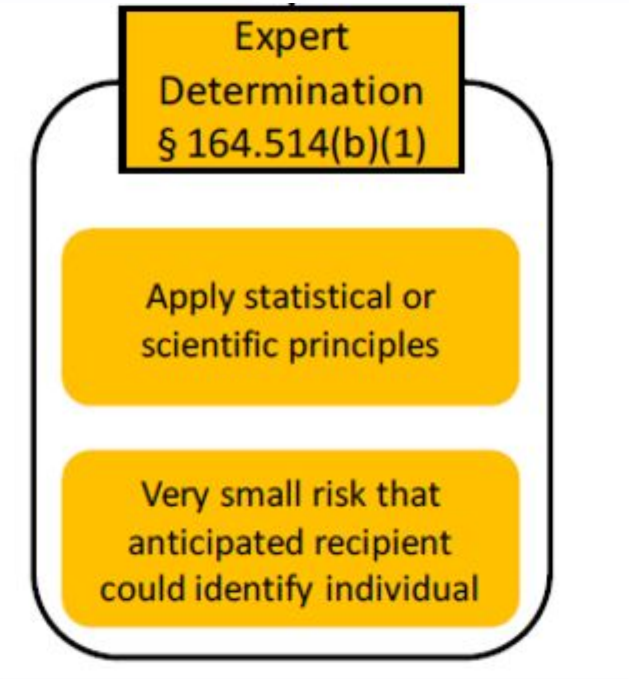
Expert Determination

PRIVATE AI

Replica Analytics

AN AETION COMPANY

Expert determination uses a risk-based approach to deem datasets de-identified


Expert Determination § 164.514(b)(1)

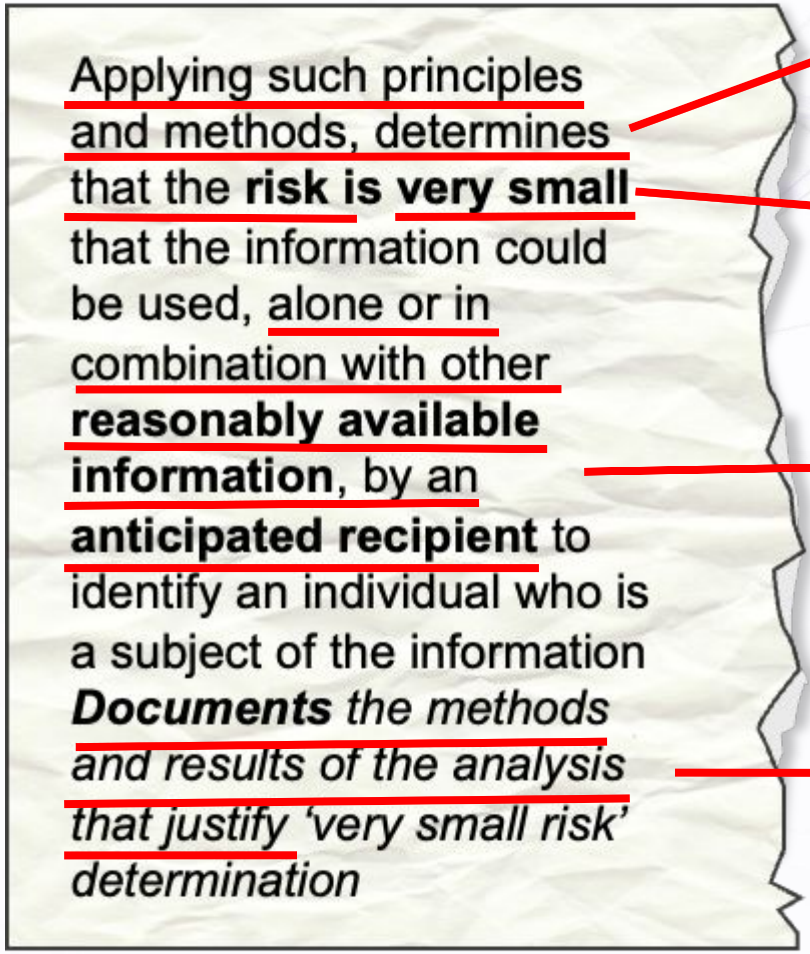
Apply statistical or scientific principles

Very small risk that anticipated recipient could identify individual

Applying such principles and methods, determines that the **risk is very small** that the information could be used, alone or in combination with other **reasonably available information**, by an **anticipated recipient** to identify an individual who is a subject of the information

***Documents** the methods and results of the analysis that justify 'very small risk' determination*

Key components of expert determination



Applying such principles and methods, determines that the **risk is very small** that the information could be used, alone or in combination with other **reasonably available information**, by an **anticipated recipient** to identify an individual who is a subject of the information ***Documents*** *the methods and results of the analysis that justify 'very small risk' determination*

Method to measure risk

Definition of 'small risk'

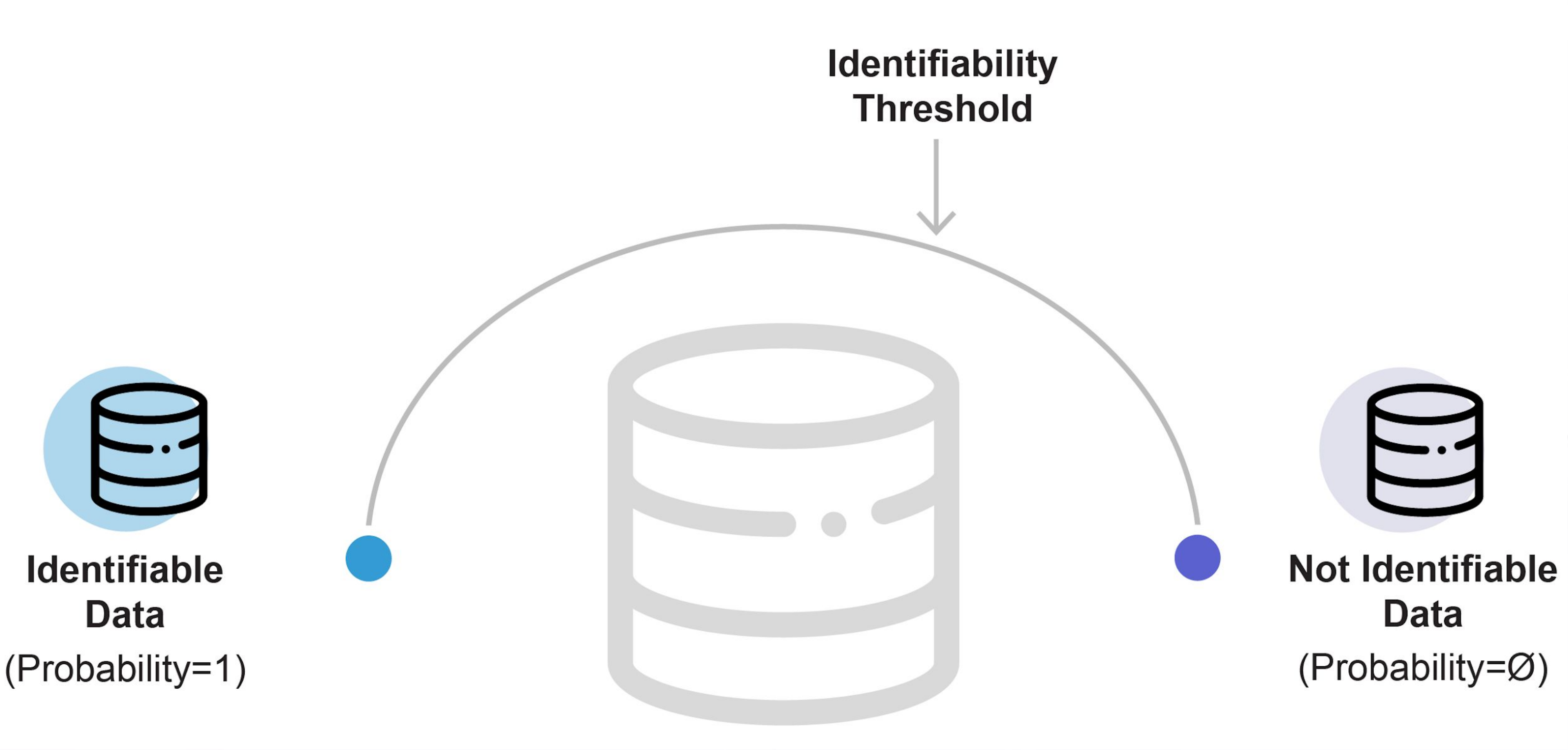
What information and motivation will an anticipated recipient have to re-identify

Document the methods, results, and justification

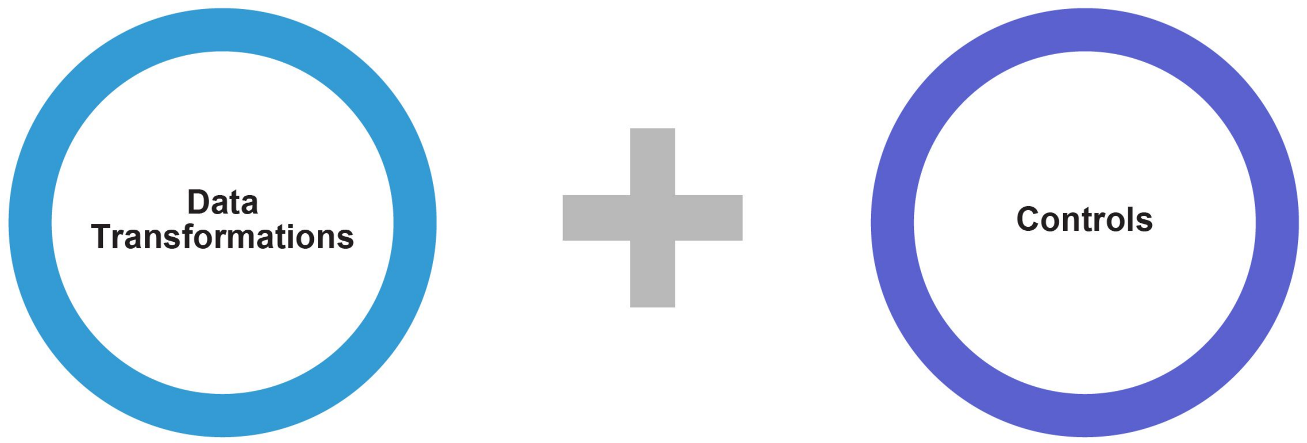
PRIVATE AI

Replica Analytics
AN AETION COMPANY

Expert determination aims to show a dataset has been de-identified so that the probability of re-identification is below an identifiability threshold



Multiple levers to manage risk



Data Transformations

Controls

- Encryption
- Generalization
- Suppression
- Addition of noise
- Microaggregation
- Synthetic data

- Security controls
- Privacy controls
- Contractual controls

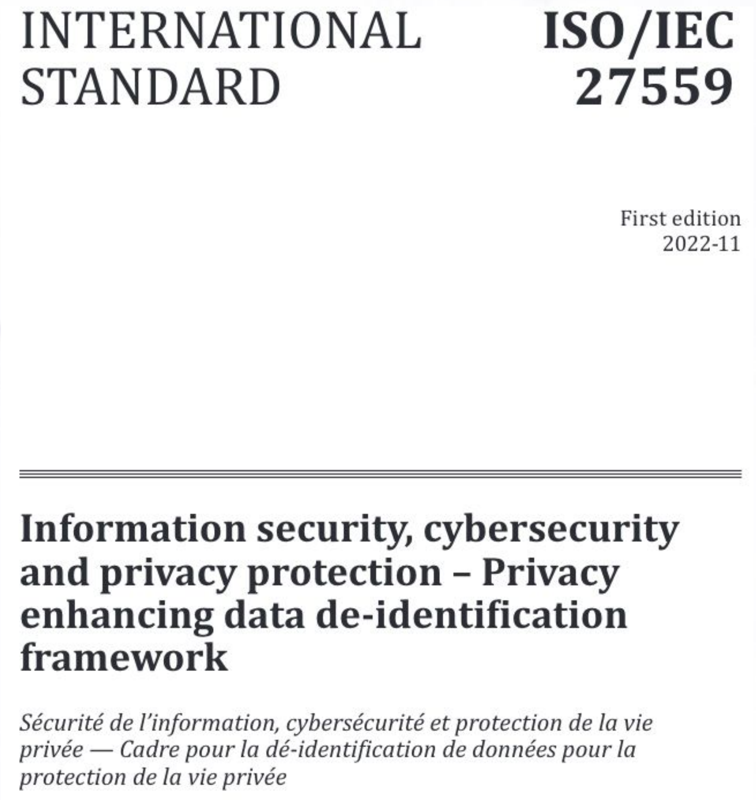
Three different ways for measuring disclosure risks

Qualitative: make a judgement call or have some rules of thumb / heuristics that can be used to decide whether the risk is acceptable

Quantitative using a model: a risk model is used to estimate the risk of a disclosure occurring

- Our focus and described in the ISO 27559 standard

Motivated Intruder test: this is also quantitative but involves launching a commissioned re-identification attack to probe the dataset



INTERNATIONAL STANDARD

ISO/IEC 27559

First edition
2022-11

Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework

Sécurité de l'information, cybersécurité et protection de la vie privée — Cadre pour la dé-identification de données pour la protection de la vie privée

PRIVATE AI

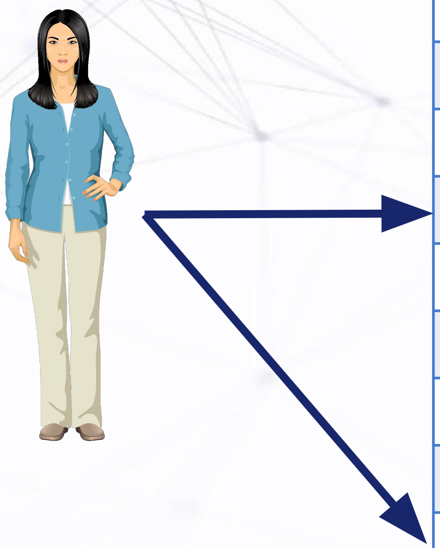
Replica Analytics
AN AETION COMPANY

Re-identification: finding a specific real person in the dataset

If I know my colleague Amy was hospitalized recently, I could try to find her record in a discharge dataset of prescriptions received

Using my knowledge that Amy is female and was born in 1989, I find two records that may be her

The risk I correctly find her record is 1 / 2 = 0.50



Sex	Year of Birth	NDC
Male	1985	009-0031
Male	1988	0023-3670
Male	1982	0074-5182
Female	1983	0078-0379
Female	1989	65862-403
Male	1981	55714-4446
Male	1982	55714-4402
Female	1987	55566-2110
Male	1981	55289-324
Female	1989	54868-6348
Male	1980	53808-0540

PRIVATE AI

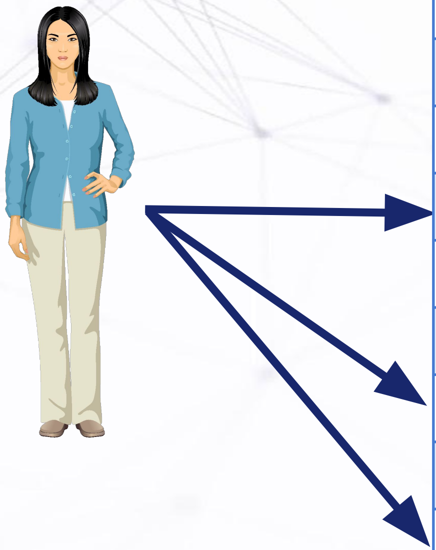
Replica Analytics
AN AETION COMPANY

Re-identification: finding a specific real person in the dataset

If I know my colleague Amy was hospitalized recently, I could try to find her record in a discharge dataset of prescriptions received

Using my knowledge that Amy is female and was born in 1989, on the generalized data, I find three records that may be her

The risk I correctly find her record is 1 / 3 = 0.33



Sex	Year of Birth	NDC
Male	1985-1989	009-0031
Male	1985-1989	0023-3670
Male	1980-1984	0074-5182
Female	1980-1984	0078-0379
Female	1985-1989	65862-403
Male	1980-1984	55714-4446
Male	1980-1984	55714-4402
Female	1985-1989	55566-2110
Male	1980-1984	55289-324
Female	1985-1989	54868-6348
Male	1980-1984	53808-0540

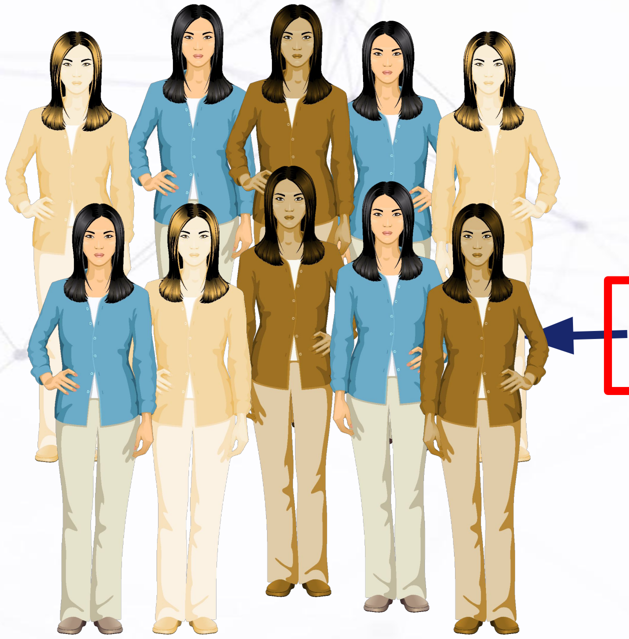
This is population to sample risk and can be calculated directly using the data you are de-identifying

Re-identification: matching records in the sample to real people

This compares a given record in the dataset to the real world population.

There may be 10 females hospitalized who were born in 1989

The risk I correctly select the right person to match the selected record is 1 / 10 = 0.10



Sex	Year of Birth	NDC
Male	1985	009-0031
Male	1988	0023-3670
Male	1982	0074-5182
Female	1983	0078-0379
Female	1989	65862-403
Male	1981	55714-4446
Male	1982	55714-4402
Female	1987	55566-2110
Male	1981	55289-324
Female	1989	54868-6348
Male	1980	53808-0540

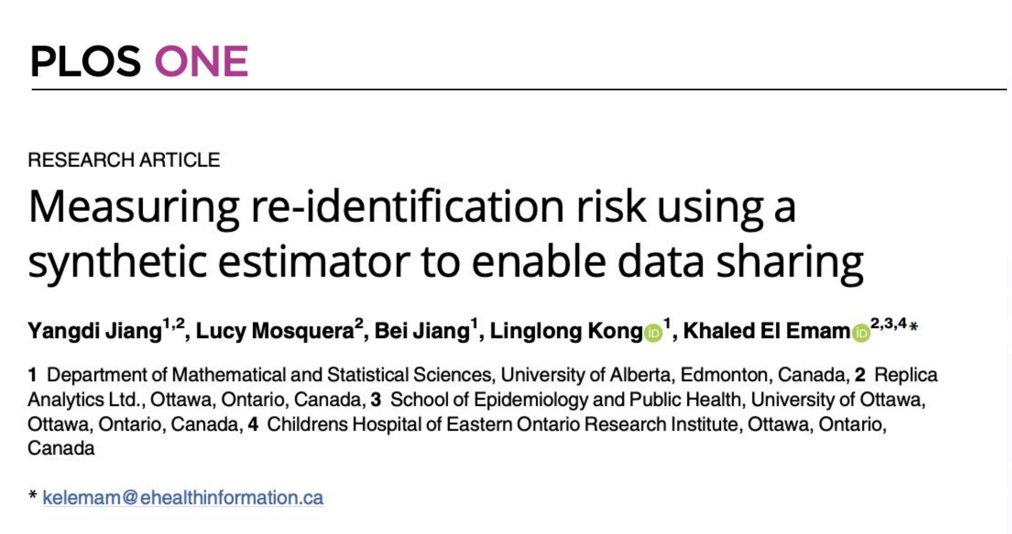
This is sample to population risk and <u>cannot</u> be calculated directly using the data you are de-identifying so you need an estimator!

PRIVATE AI

Replica Analytics
AN AETION COMPANY

Re-identification risk estimation using generative AI

Our risk estimator has been published in a peer reviewed journal and shown to have higher accuracy than a range of popular estimators



PLOS ONE

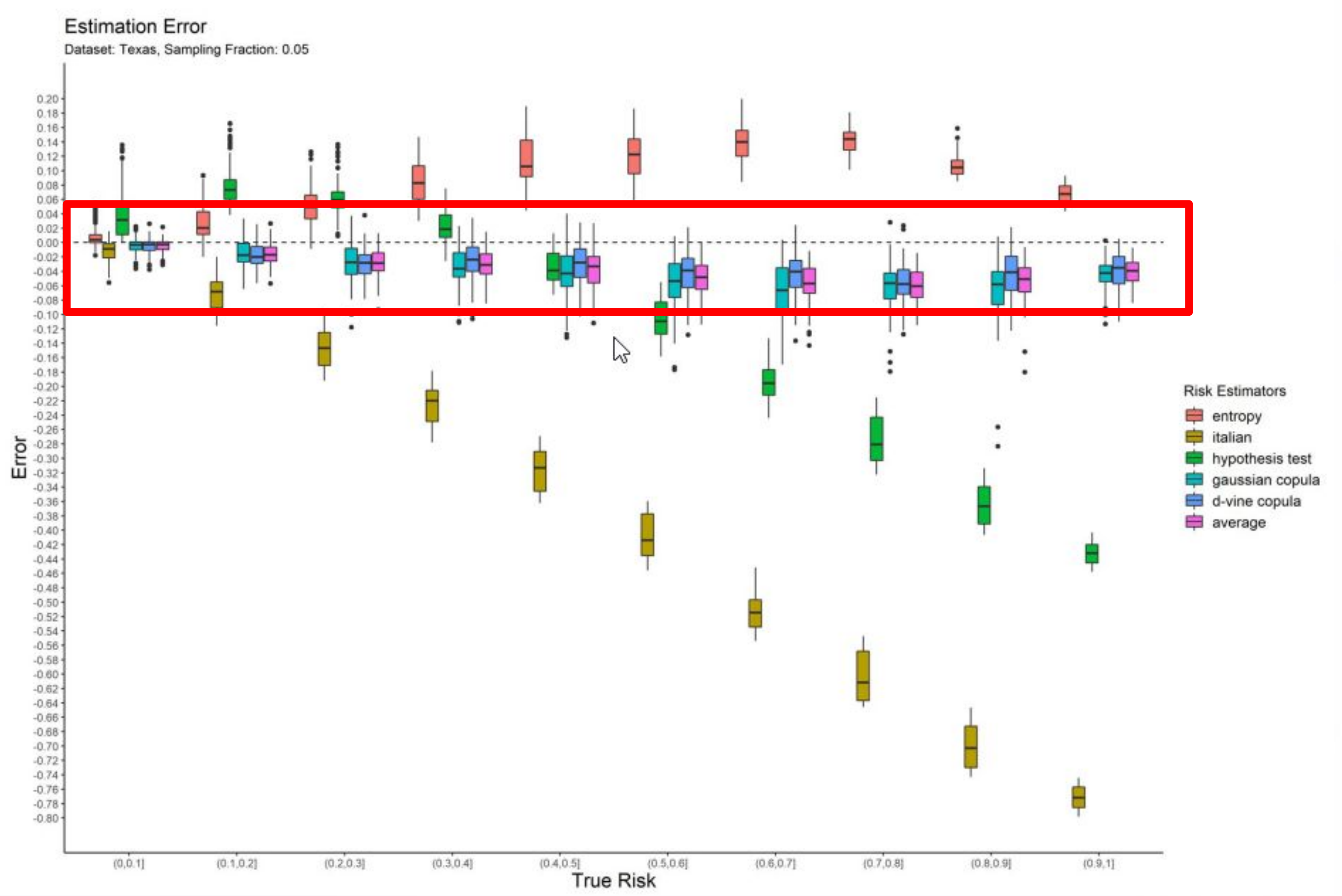
RESEARCH ARTICLE

Measuring re-identification risk using a synthetic estimator to enable data sharing

Yangdi Jiang[1,2], Lucy Mosquera[2], Bei Jiang[1], Linglong Kong[1], Khaled El Emam[2,3,4]*

1 Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada, 2 Replica Analytics Ltd., Ottawa, Ontario, Canada, 3 School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada, 4 Childrens Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada

* kelemam@ehealthinformation.ca



PRIVATE AI

Replica Analytics

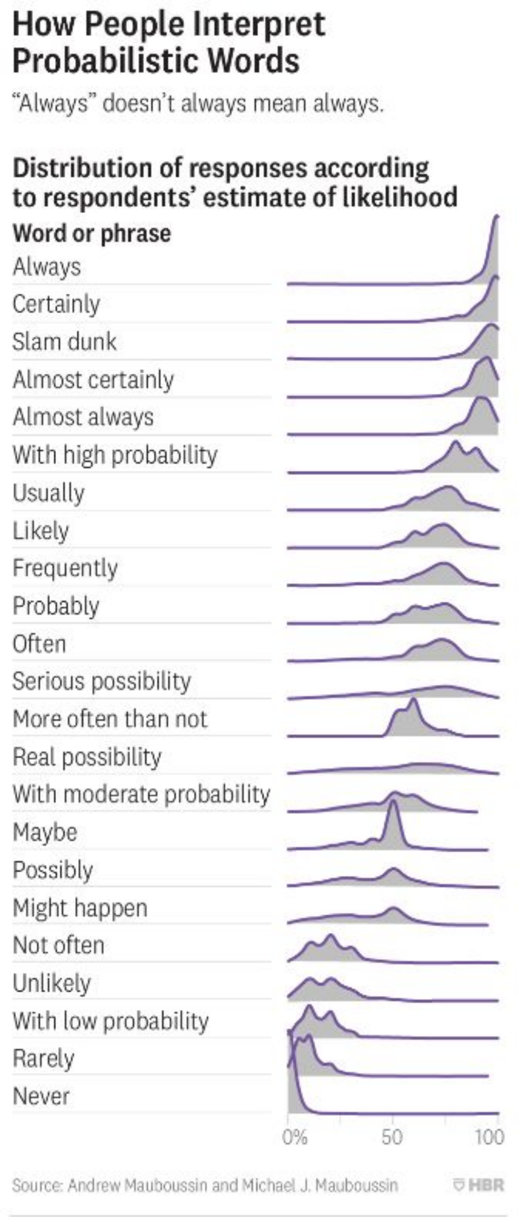
AN AETION COMPANY

Defining a ‘very small risk’

No single accepted definition of ‘very small’

Best practice for defining an acceptable threshold is to use an established value from guidance documents so that your approach is defensible.

Threshold ranges are included in the ISO 27559 standard, EMA and Health Canada have produced guidance for de-identification of clinical trial data



How People Interpret Probabilistic Words

“Always” doesn’t always mean always.

Distribution of responses according to respondents’ estimate of likelihood

Source: Andrew Mauboussin and Michael J. Mauboussin

PRIVATE AI

Replica Analytics
AN AETION COMPANY

Expert Determination

Advantages

- Allows for different risk mitigation strategies to be applied to meet the needs of end data users
- Manages risk quantitatively
- May produce higher utility data

Disadvantages

- Costly to implement
- Expert determination is not permanent to a dataset, it’s specific to an intended recipient
- Challenging to build a robust and defensible approach

Multimodal Data

PRIVATE AI

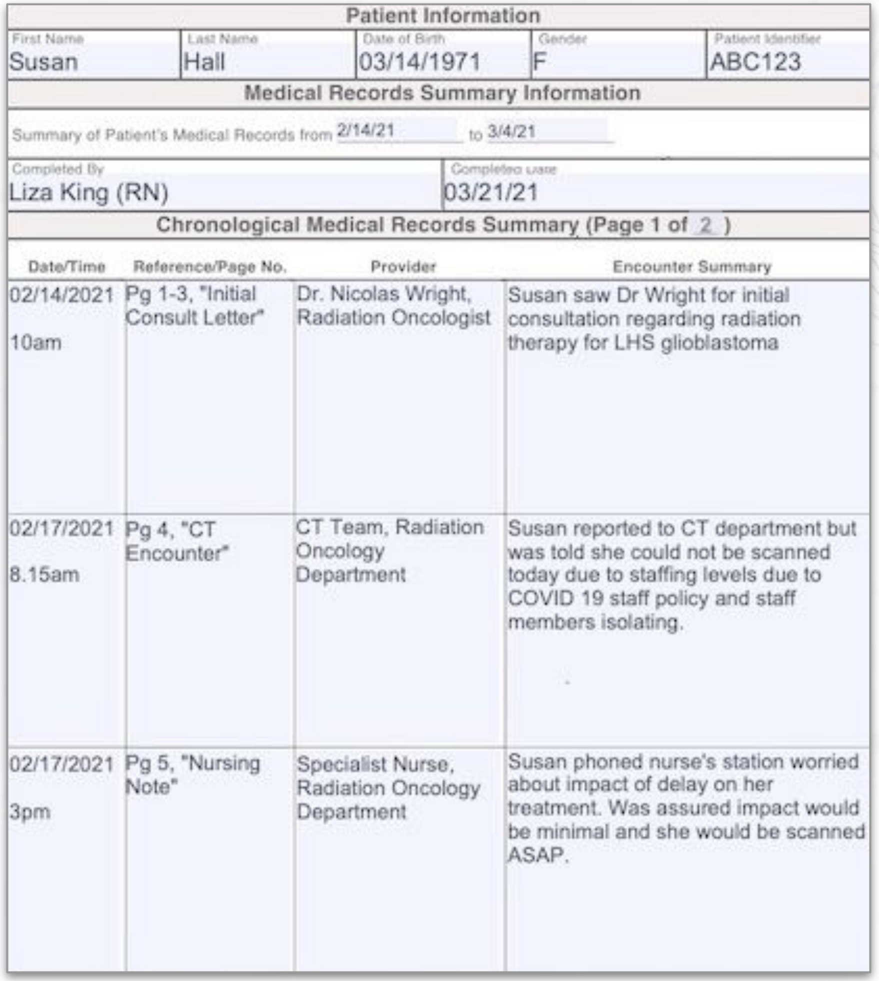
Replica Analytics

AN AETION COMPANY

ID	Birthdate	Date of visit	First	Last	Transcription
10339b10-3cd1-4ac3-ac13-ec26728cb592	1941-06-02	2022-08-05	Milo	Fadel	PREOPERATIVE DIAGNOSIS: ,Bladder cancer.,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA:, General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an **82-year-old** male who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the **San Antonio** OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left ________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.

Example: Medical Record Summary

Original



Patient Information

First Name	Last Name	Date of Birth	Gender	Patient Identifier
Susan	Hall	03/14/1971	F	ABC123

Medical Records Summary Information

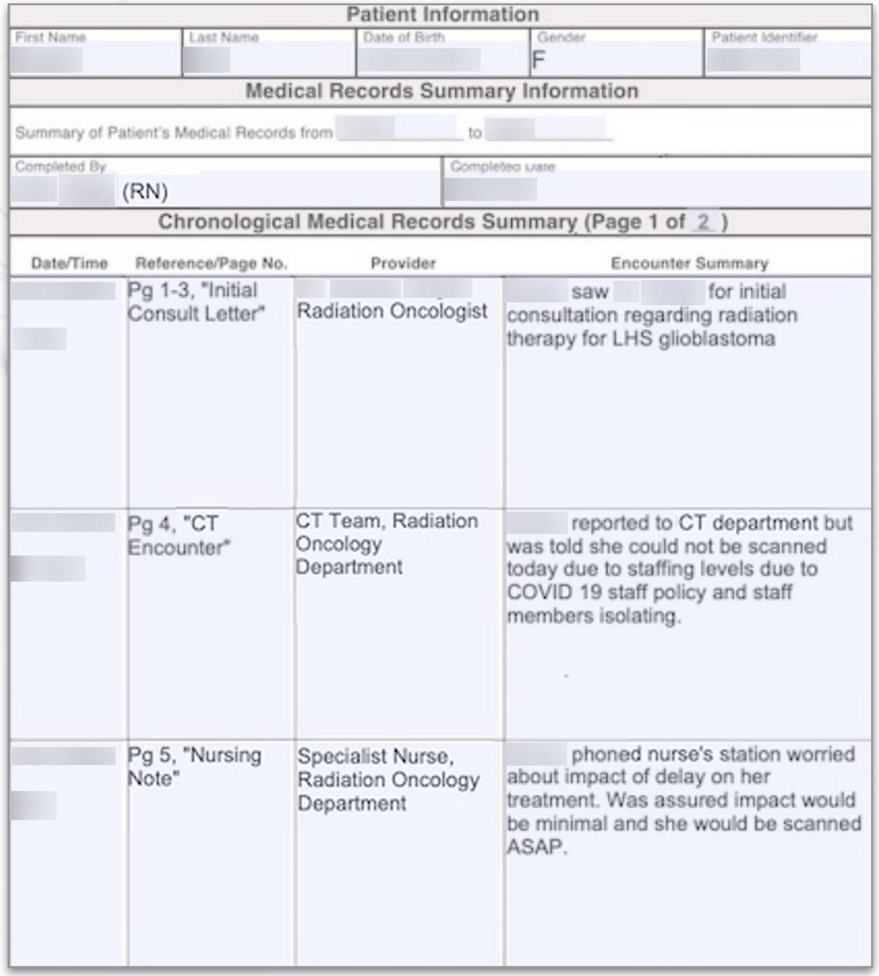
Summary of Patient's Medical Records from 2/14/21 to 3/4/21

Completed By	Completed Date
Liza King (RN)	03/21/21

Chronological Medical Records Summary (Page 1 of 2)

Date/Time	Reference/Page No.	Provider	Encounter Summary
02/14/2021 10am	Pg 1-3, "Initial Consult Letter"	Dr. Nicolas Wright, Radiation Oncologist	Susan saw Dr Wright for initial consultation regarding radiation therapy for LHS glioblastoma
02/17/2021 8.15am	Pg 4, "CT Encounter"	CT Team, Radiation Oncology Department	Susan reported to CT department but was told she could not be scanned today due to staffing levels due to COVID 19 staff policy and staff members isolating.
02/17/2021 3pm	Pg 5, "Nursing Note"	Specialist Nurse, Radiation Oncology Department	Susan phoned nurse's station worried about impact of delay on her treatment. Was assured impact would be minimal and she would be scanned ASAP.

HIPAA Safe Harbor De-identified



Patient Information

First Name	Last Name	Date of Birth	Gender	Patient Identifier
[redacted]	[redacted]	[redacted]	F	[redacted]

Medical Records Summary Information

Summary of Patient's Medical Records from [redacted] to [redacted]

Completed By	Completed Date
[redacted] [redacted] (RN)	[redacted]

Chronological Medical Records Summary (Page 1 of 2)

Date/Time	Reference/Page No.	Provider	Encounter Summary
[redacted] [redacted]	Pg 1-3, "Initial Consult Letter"	[redacted] Radiation Oncologist	[redacted] saw [redacted] for initial consultation regarding radiation therapy for LHS glioblastoma
[redacted] [redacted]	Pg 4, "CT Encounter"	CT Team, Radiation Oncology Department	[redacted] reported to CT department but was told she could not be scanned today due to staffing levels due to COVID 19 staff policy and staff members isolating.
[redacted] [redacted]	Pg 5, "Nursing Note"	Specialist Nurse, Radiation Oncology Department	[redacted] phoned nurse's station worried about impact of delay on her treatment. Was assured impact would be minimal and she would be scanned ASAP.

PRIVATE AI

Replica Analytics

AN AETION COMPANY

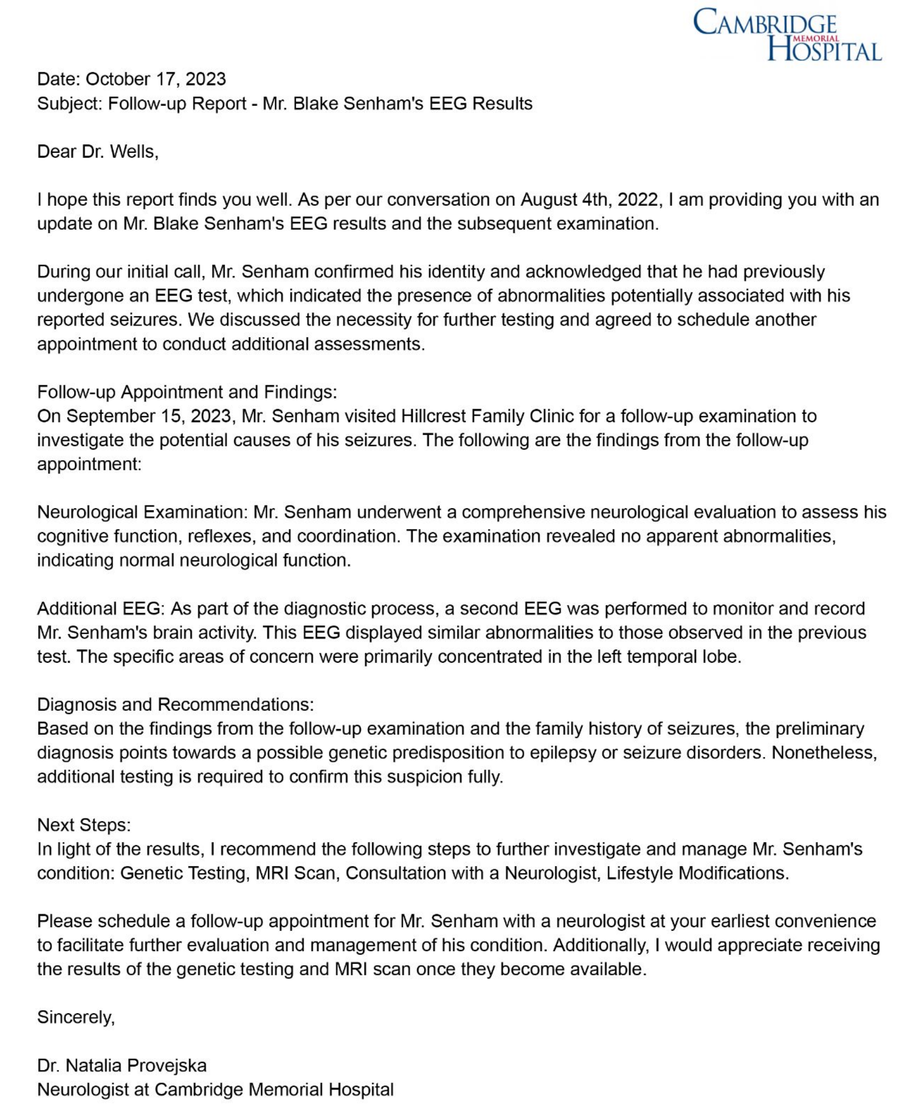
Caution with document de-identification

- Insufficient blurring may be reversible
- Optical Character Recognition is imperfect: doctor’s handwriting may be difficult to obfuscate
- Surprising types of data can be in documents: DICOM images, prescription scans, etc.

Caution with video and audio de-identification

- Face blurring removes risk of machine-based facial recognition, but clothing or other elements in the video may be recognizable to someone who knows the patient

- While PII can be fairly reliably bleeped out of the audio track, there is still not enough research on reliable methods for speech de-identification. That is, modifying the voice so it is no longer a biometric. As a result, someone who knows the patient might recognize their voice or machine-based voice recognition could be used to re-identify the individual.



CAMBRIDGE MEMORIAL HOSPITAL

Date: October 17, 2023
Subject: Follow-up Report - Mr. Blake Senham's EEG Results

Dear Dr. Wells,

I hope this report finds you well. As per our conversation on August 4th, 2022, I am providing you with an update on Mr. Blake Senham's EEG results and the subsequent examination.

During our initial call, Mr. Senham confirmed his identity and acknowledged that he had previously undergone an EEG test, which indicated the presence of abnormalities potentially associated with his reported seizures. We discussed the necessity for further testing and agreed to schedule another appointment to conduct additional assessments.

Follow-up Appointment and Findings:
On September 15, 2023, Mr. Senham visited Hillcrest Family Clinic for a follow-up examination to investigate the potential causes of his seizures. The following are the findings from the follow-up appointment:

Neurological Examination: Mr. Senham underwent a comprehensive neurological evaluation to assess his cognitive function, reflexes, and coordination. The examination revealed no apparent abnormalities, indicating normal neurological function.

Additional EEG: As part of the diagnostic process, a second EEG was performed to monitor and record Mr. Senham's brain activity. This EEG displayed similar abnormalities to those observed in the previous test. The specific areas of concern were primarily concentrated in the left temporal lobe.

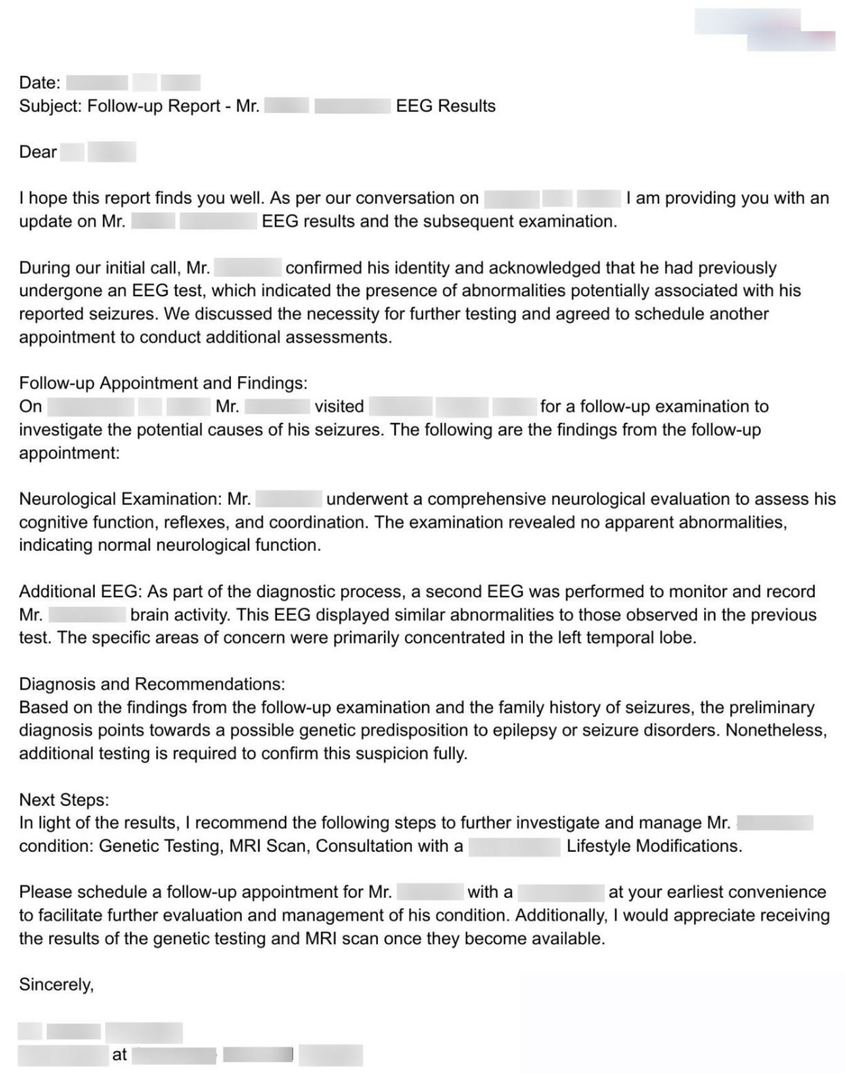
Diagnosis and Recommendations:
Based on the findings from the follow-up examination and the family history of seizures, the preliminary diagnosis points towards a possible genetic predisposition to epilepsy or seizure disorders. Nonetheless, additional testing is required to confirm this suspicion fully.

Next Steps:
In light of the results, I recommend the following steps to further investigate and manage Mr. Senham's condition: Genetic Testing, MRI Scan, Consultation with a Neurologist, Lifestyle Modifications.

Please schedule a follow-up appointment for Mr. Senham with a neurologist at your earliest convenience to facilitate further evaluation and management of his condition. Additionally, I would appreciate receiving the results of the genetic testing and MRI scan once they become available.

Sincerely,

Dr. Natalia Provejska
Neurologist at Cambridge Memorial Hospital



[redacted]

Date: [redacted]
Subject: Follow-up Report - Mr. [redacted] EEG Results

Dear [redacted]

I hope this report finds you well. As per our conversation on [redacted] I am providing you with an update on Mr. [redacted] EEG results and the subsequent examination.

During our initial call, Mr. [redacted] confirmed his identity and acknowledged that he had previously undergone an EEG test, which indicated the presence of abnormalities potentially associated with his reported seizures. We discussed the necessity for further testing and agreed to schedule another appointment to conduct additional assessments.

Follow-up Appointment and Findings:
On [redacted] Mr. [redacted] visited [redacted] for a follow-up examination to investigate the potential causes of his seizures. The following are the findings from the follow-up appointment:

Neurological Examination: Mr. [redacted] underwent a comprehensive neurological evaluation to assess his cognitive function, reflexes, and coordination. The examination revealed no apparent abnormalities, indicating normal neurological function.

Additional EEG: As part of the diagnostic process, a second EEG was performed to monitor and record Mr. [redacted] brain activity. This EEG displayed similar abnormalities to those observed in the previous test. The specific areas of concern were primarily concentrated in the left temporal lobe.

Diagnosis and Recommendations:
Based on the findings from the follow-up examination and the family history of seizures, the preliminary diagnosis points towards a possible genetic predisposition to epilepsy or seizure disorders. Nonetheless, additional testing is required to confirm this suspicion fully.

Next Steps:
In light of the results, I recommend the following steps to further investigate and manage Mr. [redacted] condition: Genetic Testing, MRI Scan, Consultation with a [redacted] Lifestyle Modifications.

Please schedule a follow-up appointment for Mr. [redacted] with a [redacted] at your earliest convenience to facilitate further evaluation and management of his condition. Additionally, I would appreciate receiving the results of the genetic testing and MRI scan once they become available.

Sincerely,

[redacted]
[redacted] at [redacted]

PRIVATE AI

Replica Analytics
AN AETION COMPANY

Case Study:
De-identification of Multimodal Data

PRIVATE AI

Replica Analytics

AN AETION COMPANY

ID	Birthdate	Date of visit	First	Last	Transcription
10339b10-3cd1-4ac3-ac13-ec26728cb592	1941-06-02	2022-08-05	Milo	Fadel	PREOPERATIVE DIAGNOSIS: ,Bladder cancer.,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA:, General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an 82-year-old male who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the San Antonio OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left ________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.

De-Identification Using Safe Harbor

PRIVATE AI

Replica Analytics

AN AETION COMPANY

ID	Birthdate	Date of visit	First	Last	Transcription
10339b10-3cd1-4ac3-ac13-ec26728cb592	1941-06-02	2022-08-05	Milo	Fadel	PREOPERATIVE DIAGNOSIS: ,Bladder cancer.,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA:, General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an 82-year-old male who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the San Antonio OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left _________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.

Medical Record Number

Date

Name

Geographic reference smaller than a state

Automated entity detection across the data to identify any instance of the 18 identifiers under Safe Harbor

ID	Birthdate	Date of visit	First	Last	Transcription
[ID_NUMBER_1]	1941	2022	[FIRST_NAME_1]	[LAST_NAME_1]	PREOPERATIVE DIAGNOSIS: ,Bladder cancer.,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA:, General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an 82-year-old male who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the Texas OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left ________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.

Date

Name

Medical Record Number

Geographic reference smaller than a state

All identifiers present have been removed so the data is de-identified under Safe Harbor

PRIVATE AI

Replica Analytics

AN AETION COMPANY

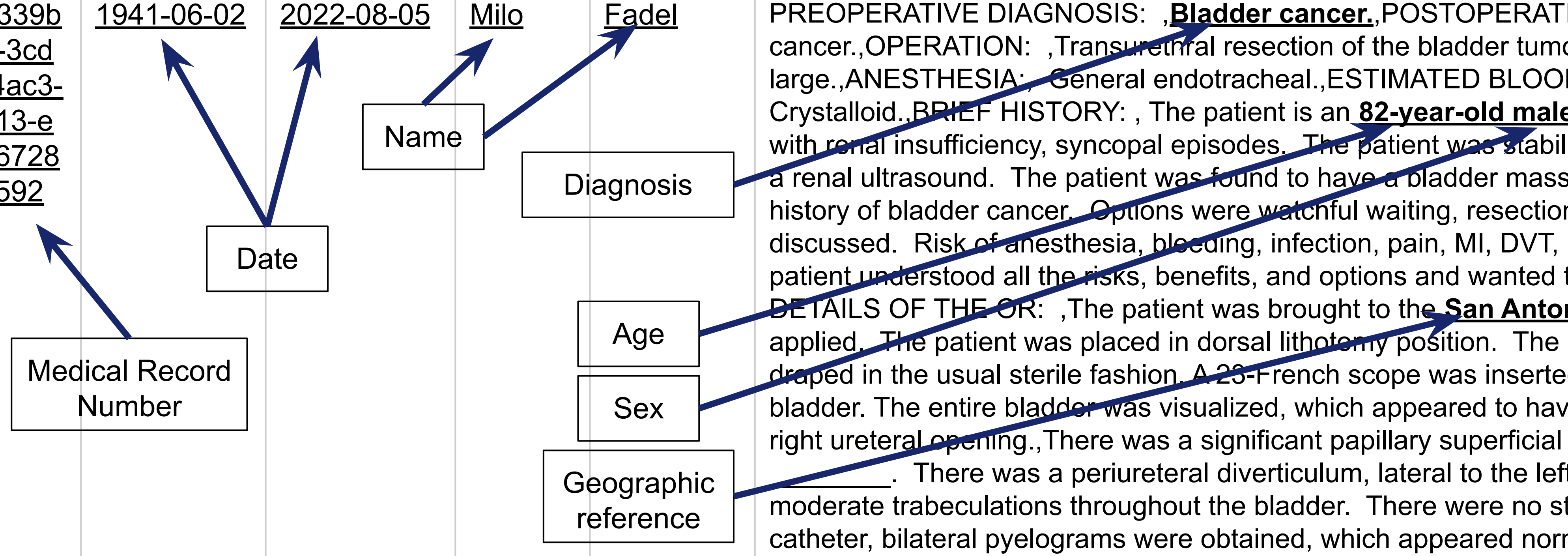
De-Identification Using Expert Determination

PRIVATE AI

Replica Analytics

AN AETION COMPANY

ID	Birthdate	Date of visit	First	Last	Transcription
10339b10-3cd1-4ac3-ac13-ec26728cb592	1941-06-02	2022-08-05	Milo	Fadel	PREOPERATIVE DIAGNOSIS: ,**Bladder cancer.**,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA: , General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an **82-year-old male** who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the **San Antonio** OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left ________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.



Name

Diagnosis

Date

Age

Sex

Medical Record Number

Geographic reference

Automated entity detection across the data to identify any individually identifiable health data

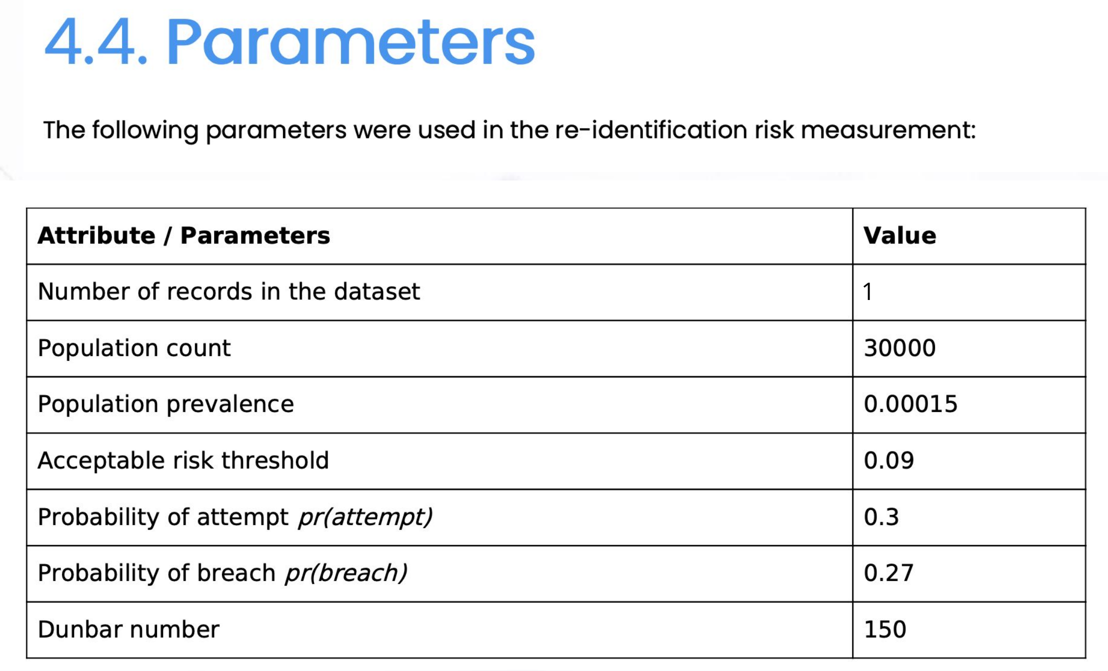
Example re-identification risk assessment

Information present an attacker may use to re-identify:

- Direct Identifiers: ID, name → Must be removed
- Quasi-Identifiers: Date of birth, age, sex, location, major diagnosis, date of visit

Assess the context of data sharing to get other key parameters and input into risk assessment software. Includes:

- Population size
- Probability of an attack occurring



4.4. Parameters

The following parameters were used in the re-identification risk measurement:

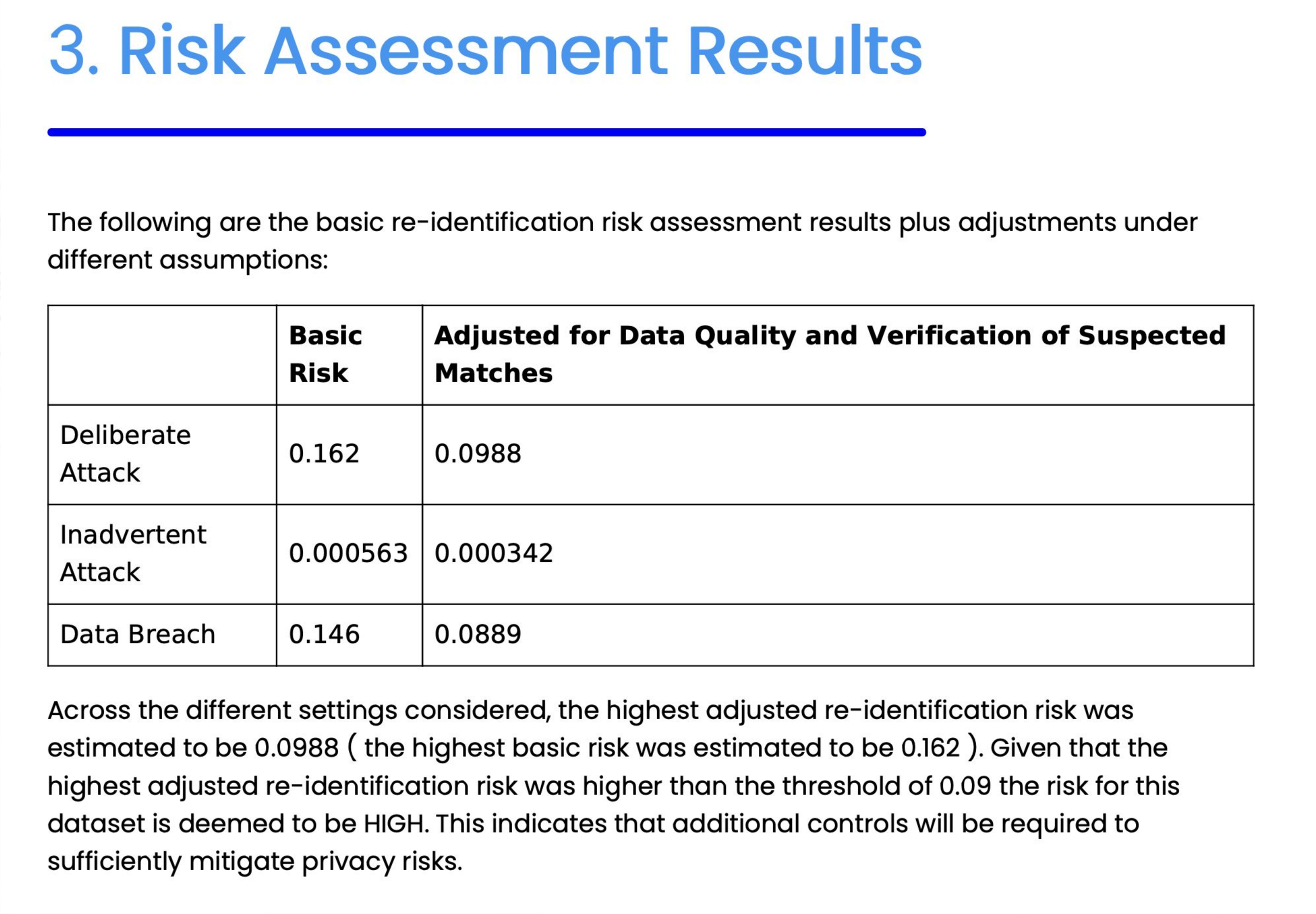
Attribute / Parameters	**Value**
Number of records in the dataset	1
Population count	30000
Population prevalence	0.00015
Acceptable risk threshold	0.09
Probability of attempt *pr(attempt)*	0.3
Probability of breach *pr(breach)*	0.27
Dunbar number	150

Initial re-identification risk

Even after transforming ID and name, the re-identification risk may still be higher than an acceptable threshold

In this example the highest estimated risk was 0.162

In this case additional transforms are required to de-identify



3. Risk Assessment Results

The following are the basic re-identification risk assessment results plus adjustments under different assumptions:

	Basic Risk	**Adjusted for Data Quality and Verification of Suspected Matches**
Deliberate Attack	0.162	0.0988
Inadvertent Attack	0.000563	0.000342
Data Breach	0.146	0.0889

Across the different settings considered, the highest adjusted re-identification risk was estimated to be 0.0988 (the highest basic risk was estimated to be 0.162). Given that the highest adjusted re-identification risk was higher than the threshold of 0.09 the risk for this dataset is deemed to be HIGH. This indicates that additional controls will be required to sufficiently mitigate privacy risks.

PRIVATE AI

Replica Analytics
AN AETION COMPANY

ID	Birthdate	Date of visit	First	Last	Transcription
3456789098765432345678	****-**-**	2022-08-05	John	Smith	PREOPERATIVE DIAGNOSIS: ,**Bladder cancer.**,POSTOPERATIVE DIAGNOSIS: , Bladder cancer.,OPERATION: ,Transurethral resection of the bladder tumor (TURBT), large.,ANESTHESIA: ,General endotracheal.,ESTIMATED BLOOD LOSS: , Minimal.,FLUIDS: , Crystalloid.,BRIEF HISTORY: , The patient is an **80-89-year-old male** who presented to the hospital with renal insufficiency, syncopal episodes. The patient was stabilized from cardiac standpoint on a renal ultrasound. The patient was found to have a bladder mass. The patient does have a history of bladder cancer. Options were watchful waiting, resection of the bladder tumor were discussed. Risk of anesthesia, bleeding, infection, pain, MI, DVT, PE were discussed. The patient understood all the risks, benefits, and options and wanted to proceed with the procedure. DETAILS OF THE OR: ,The patient was brought to the **San Antonio** OR, anesthesia was applied. The patient was placed in dorsal lithotomy position. The patient was prepped and draped in the usual sterile fashion. A 23-French scope was inserted inside the urethra into the bladder. The entire bladder was visualized, which appeared to have a large tumor, lateral to the right ureteral opening.,There was a significant papillary superficial fluffiness around the left ________. There was a periureteral diverticulum, lateral to the left ureteral opening. There were moderate trabeculations throughout the bladder. There were no stones. Using a French cone tip catheter, bilateral pyelograms were obtained, which appeared normal. Subsequently, using 24-French cutting loop resectoscope a resection of the bladder tumor was performed all the way up to the base. Deep biopsies were sent separately. Coagulation was performed around the periphery and at the base of the tumor. All the tumors were removed and sent for path analysis. There was an excellent hemostasis. The rest of the bladder appeared normal. There was no further evidence of tumor. At the end of the procedure, a 22 three-way catheter was placed, and the patient was brought to the recovery in a stable condition.The patient is to follow-up with Dr. X in seven days.

Name

Diagnosis

Date

Age

Sex

Medical Record Number

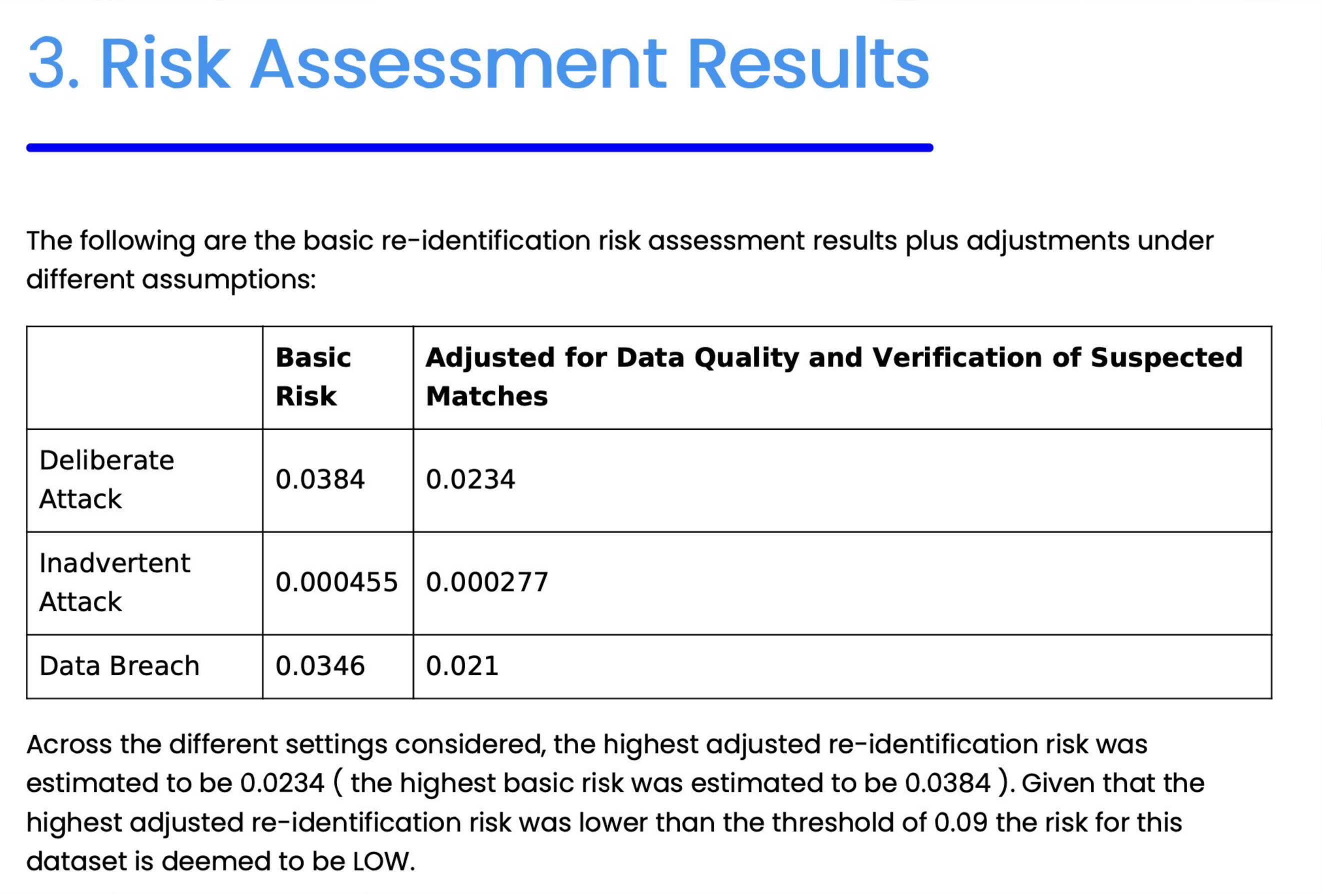
Geographic reference

Mitigate risk by replacing ID and name with random values, suppressing DOB and generalizing age

Final re-identification risk assessment

Maximum risk is now 0.038 which is below the acceptable risk threshold so the data is deemed to be de-identified

The choice of transforms to mitigate risk could be customized to meet the needs of the end data user, in this case dates associated with the treatment were prioritized to allow for time to event analyses



3. Risk Assessment Results

The following are the basic re-identification risk assessment results plus adjustments under different assumptions:

	Basic Risk	**Adjusted for Data Quality and Verification of Suspected Matches**
Deliberate Attack	0.0384	0.0234
Inadvertent Attack	0.000455	0.000277
Data Breach	0.0346	0.021

Across the different settings considered, the highest adjusted re-identification risk was estimated to be 0.0234 (the highest basic risk was estimated to be 0.0384). Given that the highest adjusted re-identification risk was lower than the threshold of 0.09 the risk for this dataset is deemed to be LOW.

Key takeaways

De-identification under HIPAA can be accomplished via Safe Harbor or Expert Determination

- Safe Harbor is a rigid checklist approach to de-identification of data
- Expert Determination is a framework to manage re-identification risks in data that requires detailed documentation and may use multiple strategies for risk mitigation (e.g., controls like contracts and cybersecurity as well as data transformations like generalization, suppression, or synthetic data generation)

De-identification can be applied to complex multi-modal health datasets to facilitate responsible re-use

Scaling de-identification requires robust technological solutions

Questions?

PRIVATE AI

Replica Analytics

AN AETION COMPANY

Thank you!

Contact:

lmosquera@replica-analytics.com
patricia@private-ai.com

PRIVATE AI

Replica Analytics
AN AETION COMPANY