Assessing and minimizing re-identification risk in data derived from health records (the cartoon version)

Gregory Simon
Kaiser Permanente Washington Health Research Institute

Supported by Cooperative Agreement U19 MH092201

KAISER PERMANENTE®

Outline:

- Motivating example
- Legal requirements
- What actually creates re-identification risk?
- Methods for assessing and mitigating risk
- Back to example

KAISER PERMANENTE®

Use case – MHRN Suicide Risk Prediction Models

- Models predicting risk of suicide attempt or suicide death within 90 days of outpatient mental health visit
- Developed and validated using data from 20 million outpatient visits in 7 health systems
- Surprisingly good prediction accuracy, substantially outperforming existing tools
- But we suspect (and hope) someone else could do better

Suicide Risk Prediction Dataset (1 record per visit)

- Demographics (sex, 5 age categories, race, ethnicity)
- Visit year
- Health system (i.e. state of residence)
- Approximately 150 dichotomous predictors regarding:
  - MH/SUD diagnoses (e.g. diagnosis of depression in last 90 days)
  - MH medications (e.g. prescription for antipsychotic in last 5 yrs)
  - MH utilization (e.g. ED visit for MH diagnosis in last year)
  - Hx of suicidal behavior (e.g. ED visit for injury/poisoning in last yr)
- Outcomes
  - Non-fatal suicide attempt within 90 days of visit (in broad categories)
  - Suicide death within 90 days of visit (in broad categories)

What the law requires:

- De-identified data
  - Does not contain direct or indirect identifiers
  - Can be shared without formal Data Use Agreement
  - Presumed to have very low (acceptable) reidentification risk
- Limited data
  - Contains indirect identifiers
  - Cannot be shared without formal Data Use Agreement
  - Presumed to have higher (unacceptable) reidentification risk

Data can be considered de-identified or “safe for sharing” if:

- Safe Harbor method
  - Does not contain any of the 18 forbidden elements
  - Does not contain other known secondary identifiers
- Expert Determination method
  - An “expert” with knowledge of these data and broader data ecosystem determines risk is “not greater than very small”
  - This standard could be stricter than the Safe Harbor method – if you know that risk is greater than “very small”
  - BUT don’t worry – listening to this presentation doesn’t make you an official expert

Is our suicide risk prediction dataset safe for sharing?

- It contains none of the 18 forbidden elements
- We don't have direct knowledge of potential secondary identifiers
- So we can say we're in that "safe harbor"
- BUT, we should aspire to a higher standard than not breaking the law
- And I'd like to keep my job
- SO, we should ask:
  - What really is the risk of re-identification?
  - How can we reduce it?

Structure of our data

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

KAISER PERMANENTE®

Where is the danger in these data?

Not here in the sensitive places

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

But here, in the ordinary places

KAISER PERMANENTE®

The key distinction: unique vs. identifying

- Exact value of my last 5 bank transactions
  - Very likely unique to me
  - But not identifying unless you already have my bank records
- My 9-digit zip code and year of birth
  - Could be unique (or close to unique) to me
  - Widely available
- It's not the private stuff that creates risk. It's the public stuff linked to the private stuff.

Applied to our dataset:

- The re-identification risk doesn't come from sensitive things that nobody knows:
  - History of suicide attempt in prior 90 days
  - Diagnosis of drug use disorder in prior year
  - Diagnosis of schizophrenia at index visit
- It comes from ordinary things that people could know:
  - Age group
  - Race/Ethnicity
  - State of residence

Example: Linkage to state mortality data

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Name	State	Year	Age	Sex	Race	Hisp
A……. B…	WA	2012	16	M	WH	Y
C….. D…..	WA	2012	55	M	WH	N
D…. E….	WA	2012	62	M	WH	N
H….. I….	WA	2012	19	F	AS	N
J…. K….	WA	2012	81	F	BL	Y
L…. M…	WA	2012	40	F	WH	N

KAISER PERMANENTE®

Confusion about risk due to “small cell sizes”

It’s not about the frequencies within a column

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Over-estimates risk in a small dataset (5 records out of 200 = 2.5%, not very unique)
Under-estimates risk in a large dataset (In 20 million records, none will have counts <6)

KAISER PERMANENTE®

Confusion about risk due to "small cell sizes"

And it's not about the uniqueness of an entire row

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

With only 20 dichotomous predictors, number of cells = 1,048,576
Many (if not most) of those cells are certain to have few members

Confusion about risk due to “small cell sizes”

It’s about uniqueness in the “potentially linkable” parts of a row

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

And matching uniqueness in some identified external data source

KAISER PERMANENTE®

Risk is always defined in relation to linkable external data

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
WA	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
CA	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
MI	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
MN	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
HI	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
OR	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
CA	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
MN	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
WA	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
CO	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
CA	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

Name	State	Year	Age	Sex	Race	Hisp
A……. B…	WA	2012	16	M	WH	Y
C….. D…..	WA	2012	55	M	WH	N
D…. E….	WA	2012	62	M	WH	N
H….. I….	WA	2012	19	F	AS	N
J…. K….	WA	2012	81	F	BL	Y
L…. M…	WA	2012	40	F	WH	N

KAISER PERMANENTE®

Which of these “small cell sizes” would create risk of re-identification state mortality data?

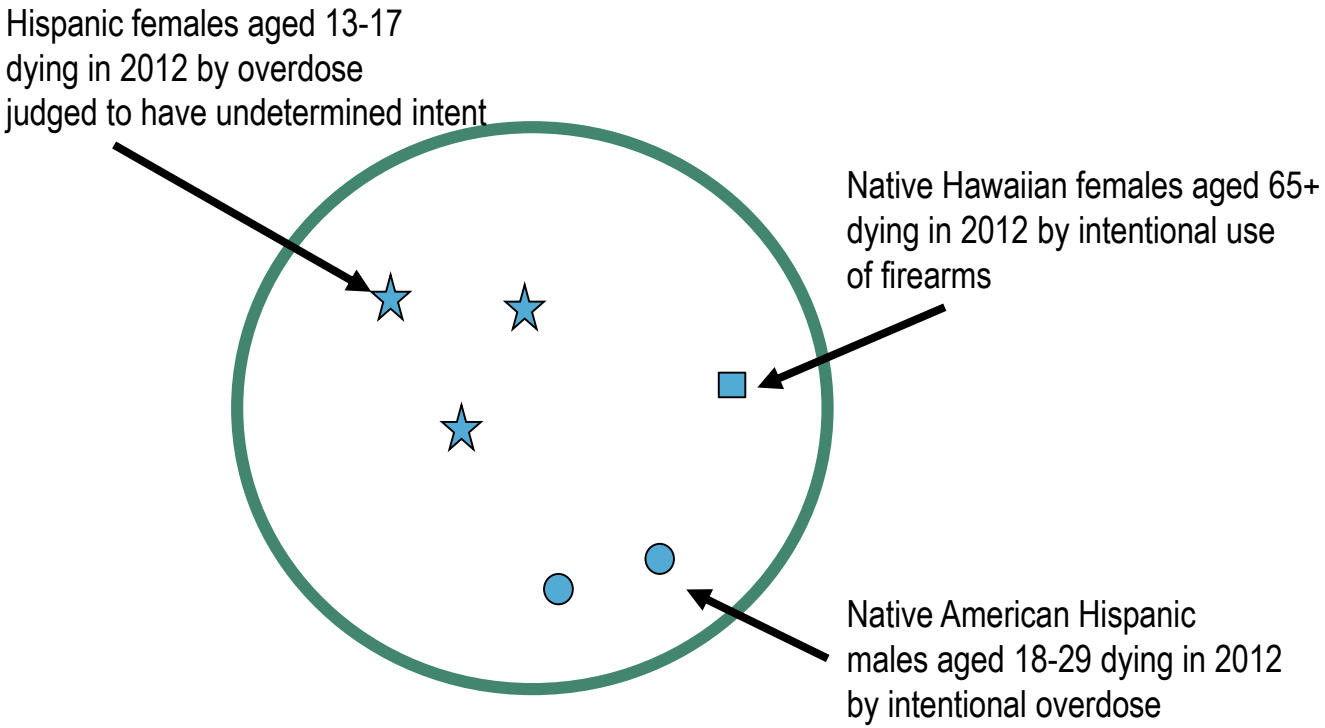
- Our data set includes only 3 people with recent diagnoses of PTSD and Asthma dying by suicide in 2012
- Our data set includes only 3 Hispanic females aged 13-17 in Washington in 2012 with recent diagnoses of schizoaffective disorder
- Our dataset includes only 3 Hispanic females aged 13-17 dying in 2012 in Washington state by overdose judged to have undetermined intent

Which of these "small cell sizes" would create risk of re-identification state mortality data?

- Our data set includes only 3 people with recent diagnoses of PTSD and Asthma dying by suicide in 2012
- Our data set includes only 3 Hispanic females aged 13-17 in Washington in 2012 with recent diagnoses of schizoaffective disorder
- Our dataset includes only 3 Hispanic females aged 13-17 dying in 2012 in Washington state by overdose judged to have undetermined intent

KAISER PERMANENTE®

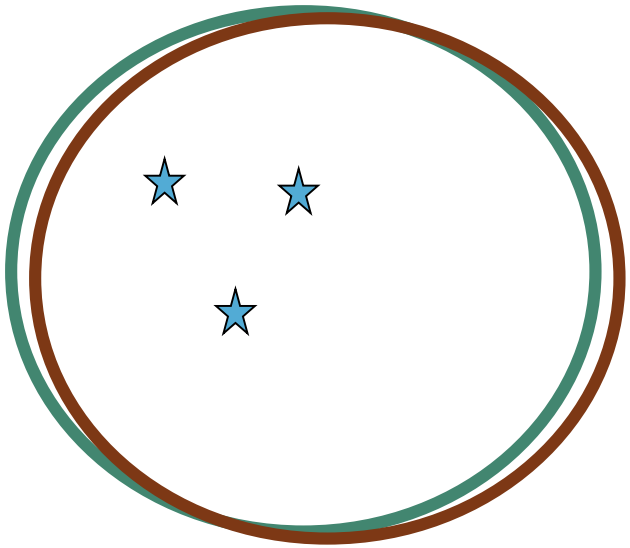
Small cell sizes or risky records in our Washington state sample



KAISER PERMANENTE®

Risk depends on overlap of two populations

Complete overlap: 3 names matched to 3 sets of health records



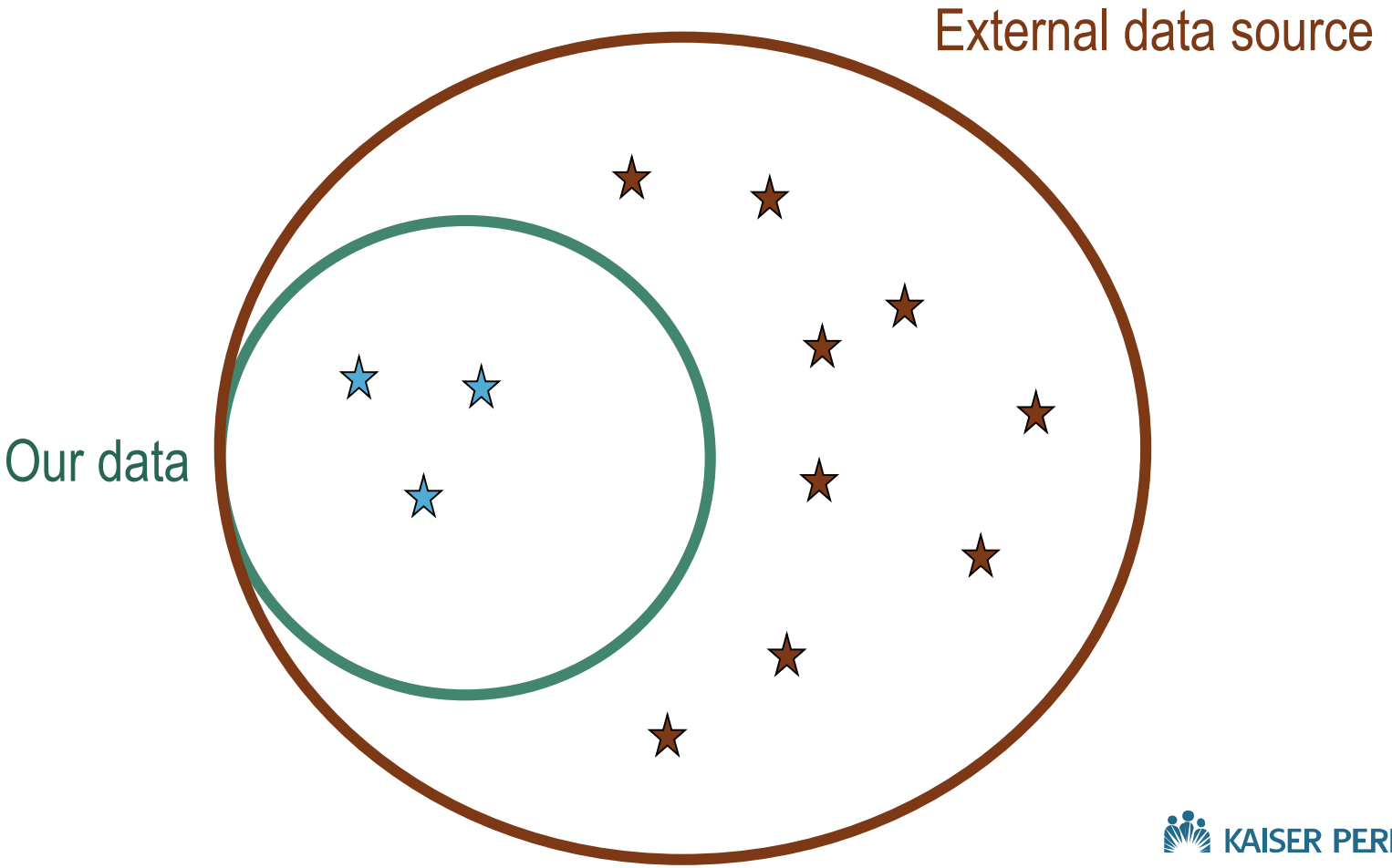
External data source

Our data

KAISER PERMANENTE®

Risk depends on overlap of two populations

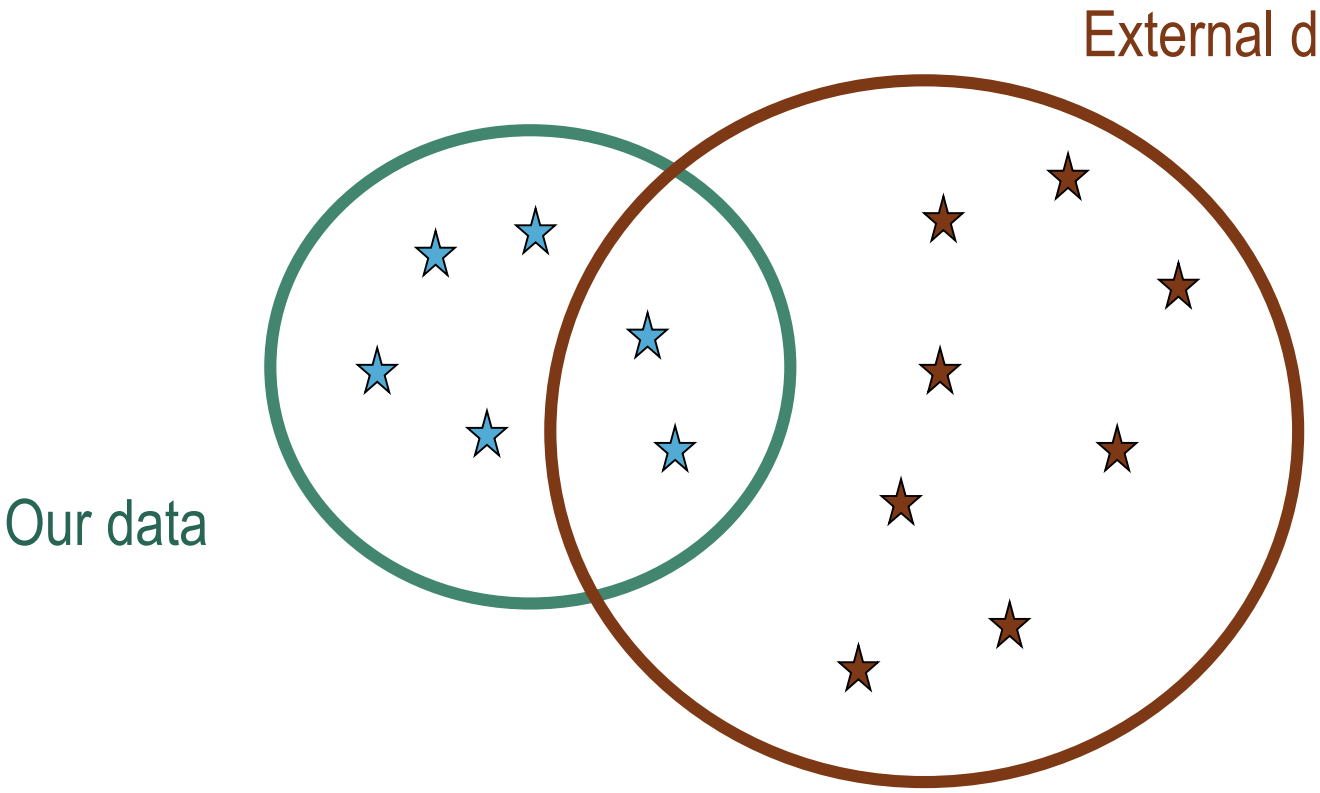
Our population is a subset: 15 names matched to 3 sets of health records



KAISER PERMANENTE®

Risk depends on overlap of two populations

Partial overlap: 10 names matched to 2 (out of 6) sets of health records



KAISER PERMANENTE®

If we can directly examine external data:

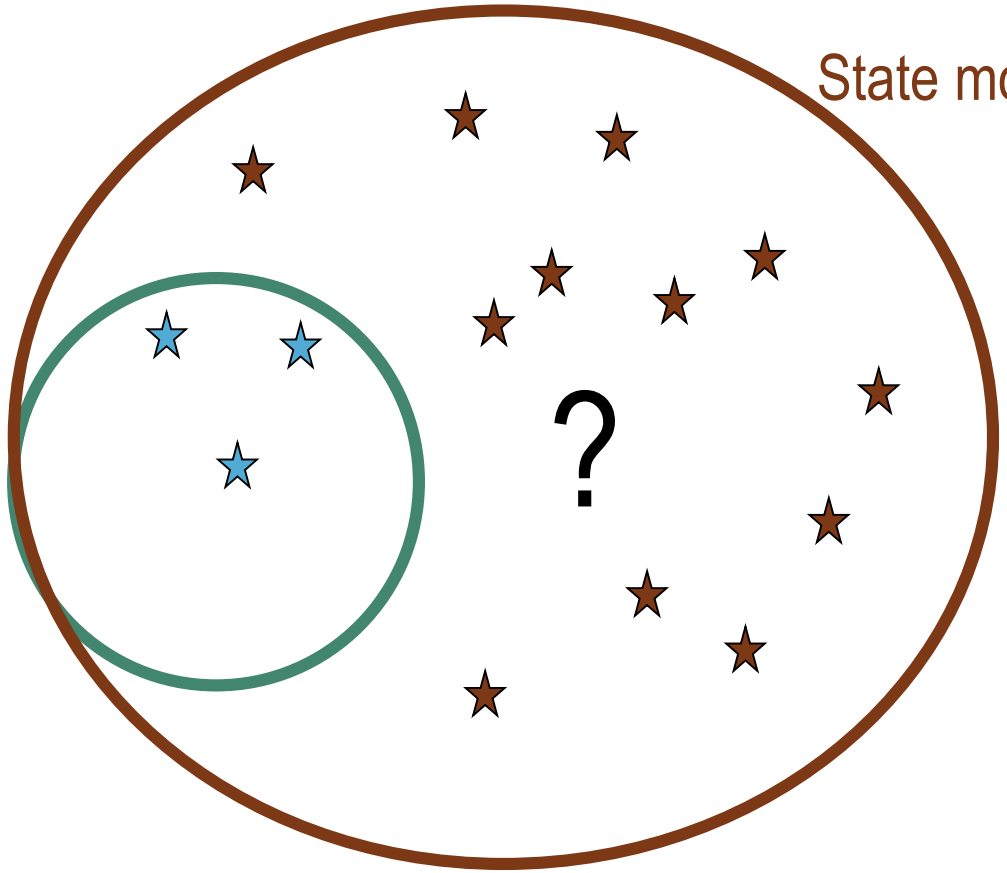
- We can precisely identify unique or nearly unique matches
- We don't need to estimate number of matches in non-overlapping portion of external data
- We can directly address re-identification risk at the record level by:
  - Modifying individual records
  - Removing individual records

But examining external data source(s) is usually not possible

So we usually need to estimate risk

Our actual situation in any state

20% subset: 15 names matched to 3 sets of health records
If you were one of those 3, you might think that risk is too high



KAISER PERMANENTE®

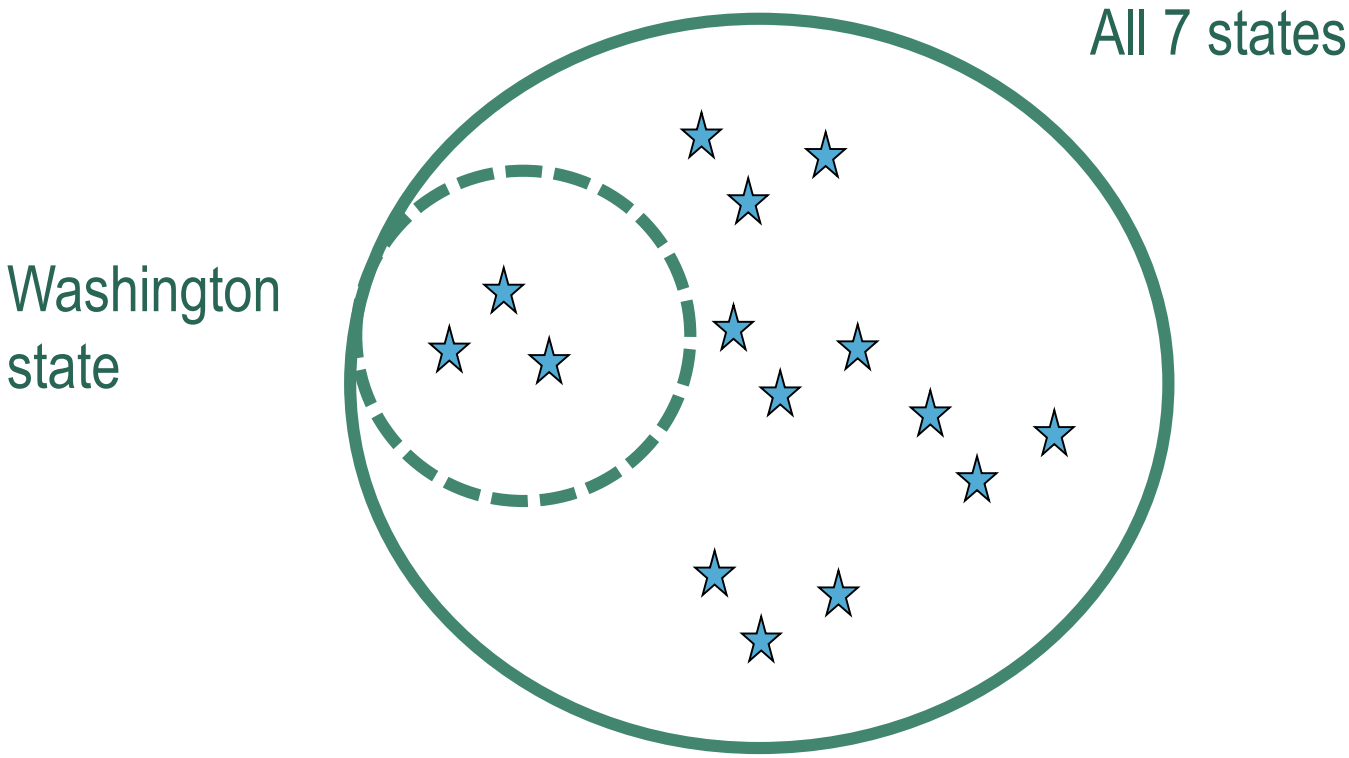
Proposal:

- Remove state (i.e. health system) variable
- Leave everything else intact

State	Year	Age	Sex	Race	Hisp	Suicidal Behavior					Mental Health Diagnoses					General Medial Diagnoses				
	2012	13-17	M	WH	Y	1	0	0	0	…	1	0	0	0	…	0	0	0	1	…
	2011	65+	F	AS	N	0	0	0	0	…	1	0	0	1	…	0	0	0	0	…
	2015	30-44	F	WH	N	0	0	0	0	…	0	0	0	0	…	0	0	0	0	…
	2010	18-29	M	AS	N	0	0	0	0	…	1	1	0	0	…	0	0	1	0	…
	2014	13-17	F	BL	Y	0	0	0	1	…	1	0	1	0	…	0	1	1	1	…
	2009	45-64	M	WH	N	0	0	0	0	…	1	0	0	0	…	0	0	1	0	…
	2011	13-17	F	BL	N	0	0	0	0	…	1	0	1	0	…	0	0	O	1	…
	2015	45-64	M	HPI	N	0	0	1	0	…	0	0	0	0	…	0	1	1	0	…
	2010	65+	M	WH	N	0	0	0	0	…	1	0	0	1	…	0	0	1	0	…
	2009	18-29	F	BL	Y	1	0	0	0	…	0	1	0	1	…	1	0	0	0	…
	2012	45-64	F	WH	N	0	0	0	0	…	0	0	0	1	…	0	0	0	0	…
	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

KAISER PERMANENTE®

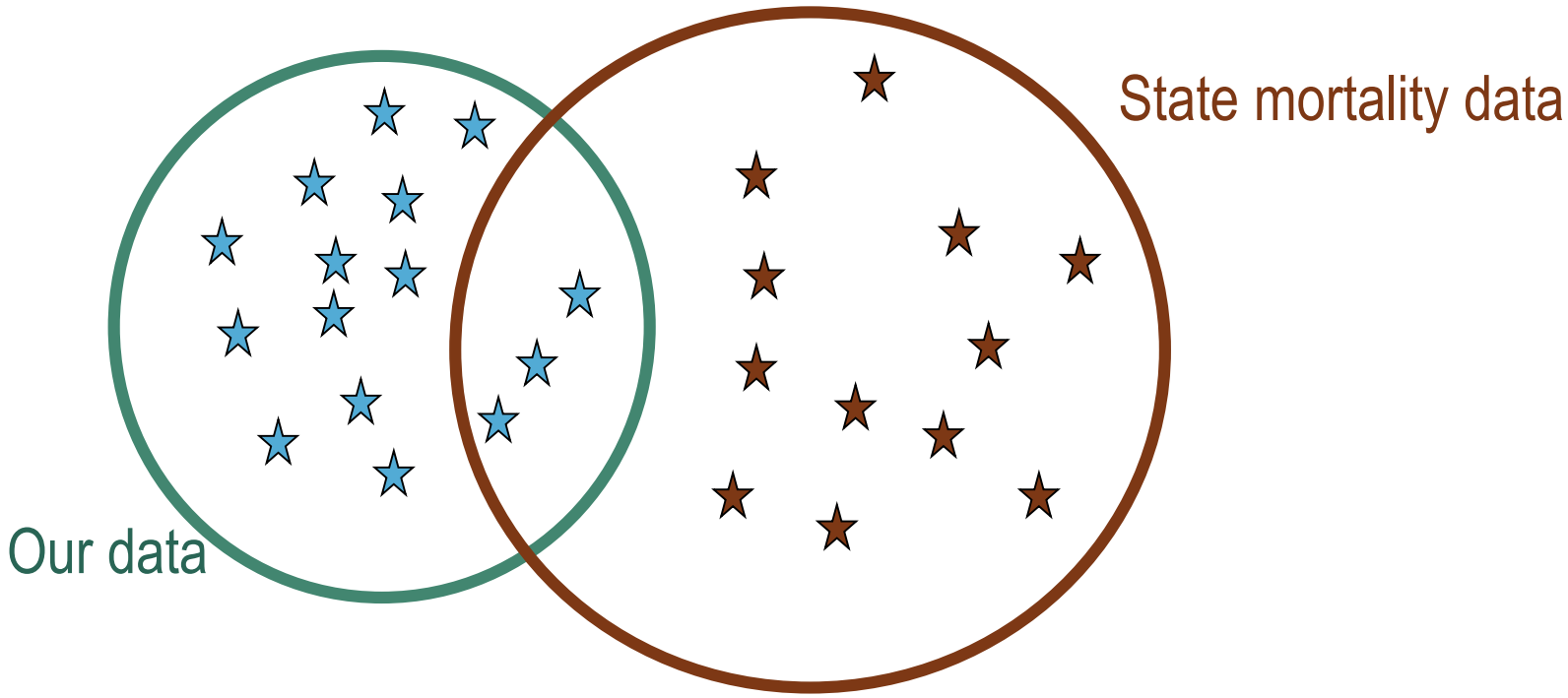
Deleting state variable increases size of smallest cells or classes



KAISER PERMANENTE®

Our new situation (with state data)

Washington state accounts for 20% of our data
Our data account for only 20% of Washington state
Partial overlap scenario: lower risk



KAISER PERMANENTE®

Questions:

- Will removing the site (state) variable adequately address risk of re-identification using state or national mortality data?
- Given elements in our dataset, what other external data sources should we consider?
- Anything else that could cost Greg his job?

Risk-based De-identification

Khaled El Emam

Pseudonymous Data

Examples of direct identifiers: Name, address, telephone number, fax number, MRN, health card number, health plan beneficiary number, VID, license plate number, email address, photograph, biometrics, SSN, SIN, device number, clinical trial record number

X

Examples of quasi-identifiers: sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, total years of schooling, marital status, criminal history, total income, visible minority status, profession, event dates, number of children, high level diagnoses and procedures

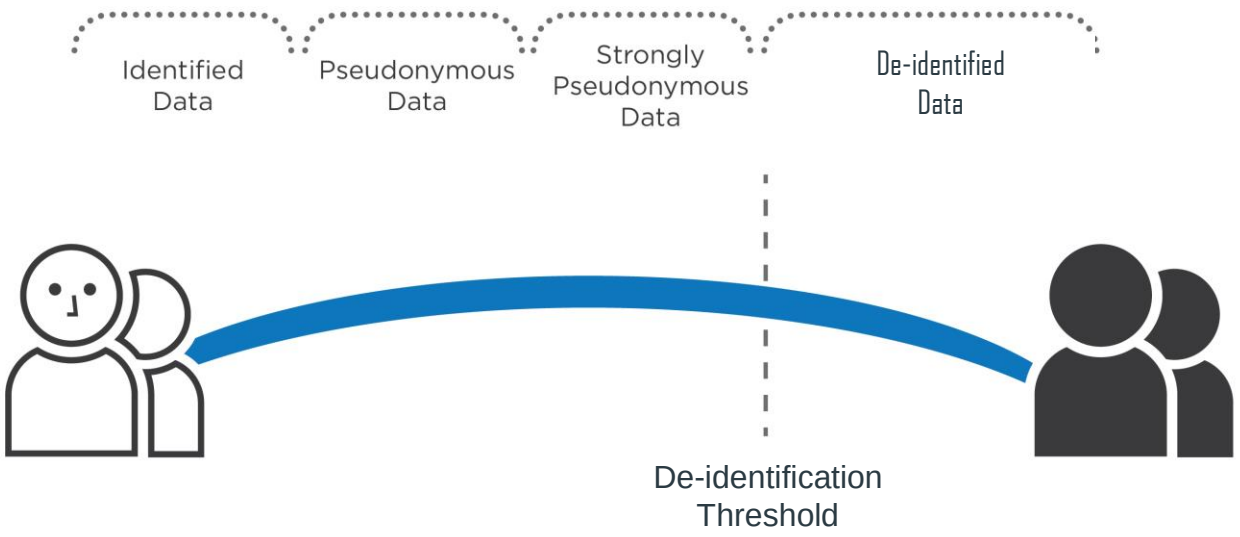
Pseudonymous Data

Examples of direct identifiers: Name, address, telephone number, fax number, MRN, health card number, health plan beneficiary number, VID, license plate number, email address, photograph, biometrics, SSN, SIN, device number, clinical trial record number

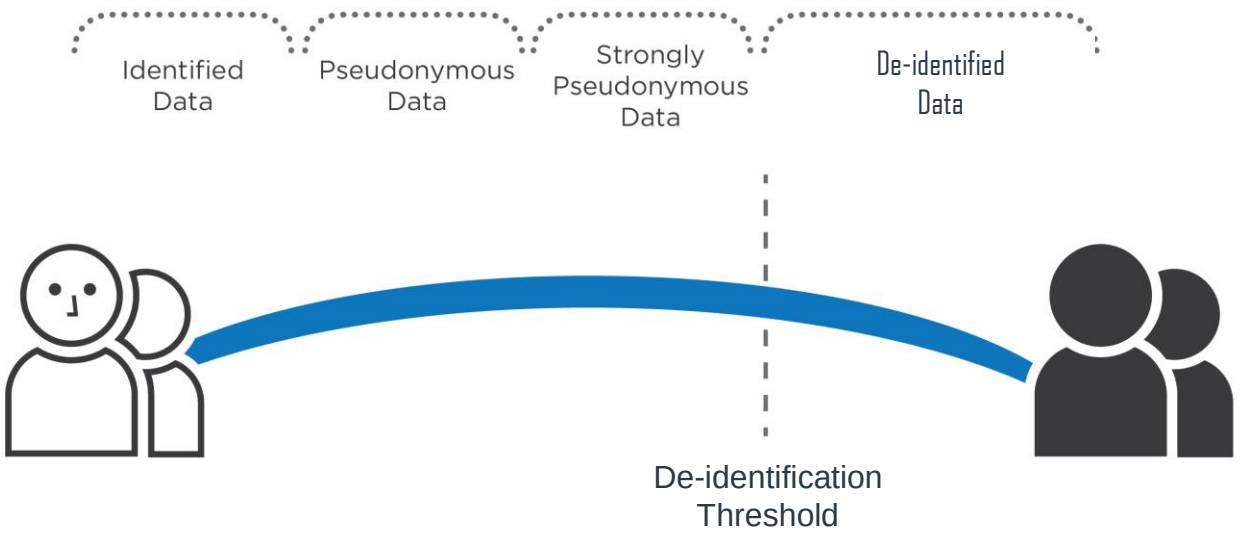
X

Examples of quasi-identifiers: sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, total years of schooling, marital status, criminal history, total income, visible minority status, profession, event dates, number of children, high level diagnoses and procedures

The Identifiability Spectrum



The Identifiability Spectrum

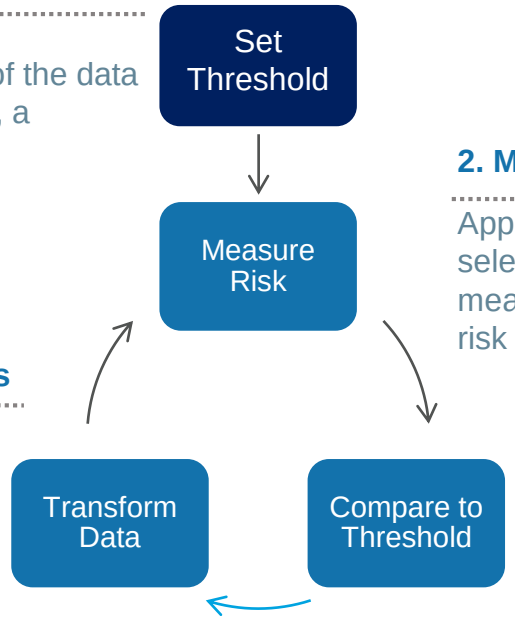


De-identification Guidelines

De-identification Cycle

1. Set Risk Threshold

Based on the characteristics of the data and precedents, a quantitative risk threshold is set.



2. Measure Risk

Appropriate metrics are selected and used to measure re-identification risk from the data.

3. Evaluate Risk

Compare the measured risk against the threshold to determine if it is above or below it.

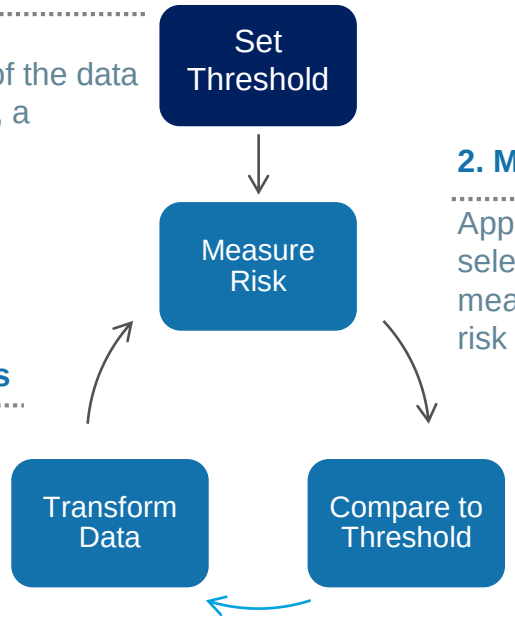
4. Apply Transformations

If the measured risk does not meet the threshold, specific transformations are applied to reduce the risk.

De-identification Cycle

1. Set Risk Threshold

Based on the characteristics of the data and precedents, a quantitative risk threshold is set.



2. Measure Risk

Appropriate metrics are selected and used to measure re-identification risk from the data.

3. Evaluate Risk

Compare the measured risk against the threshold to determine if it is above or below it.

4. Apply Transformations

If the measured risk does not meet the threshold, specific transformations are applied to reduce the risk.

Measuring Data Risk

DIRECT IDENTIFIERS			QUASI-IDENTIFIERS		OTHER VARIABLES		
ID	Name	Telephone No.	Sex	Year of Birth	Lab Test	Lab Result	Pay Delay
1	~~John Smith~~	~~(412) 668-5468~~	M	1959	Albumin, Serum	4.8	37
2	~~Alan Smith~~	~~(413) 822-5074~~	M	1969	Creatine Kinase	86	36
3	~~Alice Brown~~	~~(416) 886-5314~~	F	1955	Alkaline Phosph		52
4	~~Hercules Green~~	~~(613)763-5254~~	M	1959	Bilirubin		36
5	~~Alicia Freds~~	~~(613) 586-6222~~	F	1942	BUN/Creatinine		82
6	~~Gill Stringer~~	~~(954) 699-5423~~	F	1975	Calcium, Seru		34
7	~~Marie Kirkpatrick~~	~~(416) 786-6212~~	F	1966	Free Thyroxine Index	2.7	23
8	~~Leslie Hall~~	~~(905) 668-6581~~	F	1987	Globulin, Total	3.5	9
9	~~Douglas Henry~~	~~(416) 423-5965~~	M	1959	B-type Natriuretic peptide	134	38
10	~~Fred Thompson~~	~~(416) 421-7719~~	M	1967	Creatine Kinase	80	21

3

Two quasi-identifiers matching in three cells within a data set

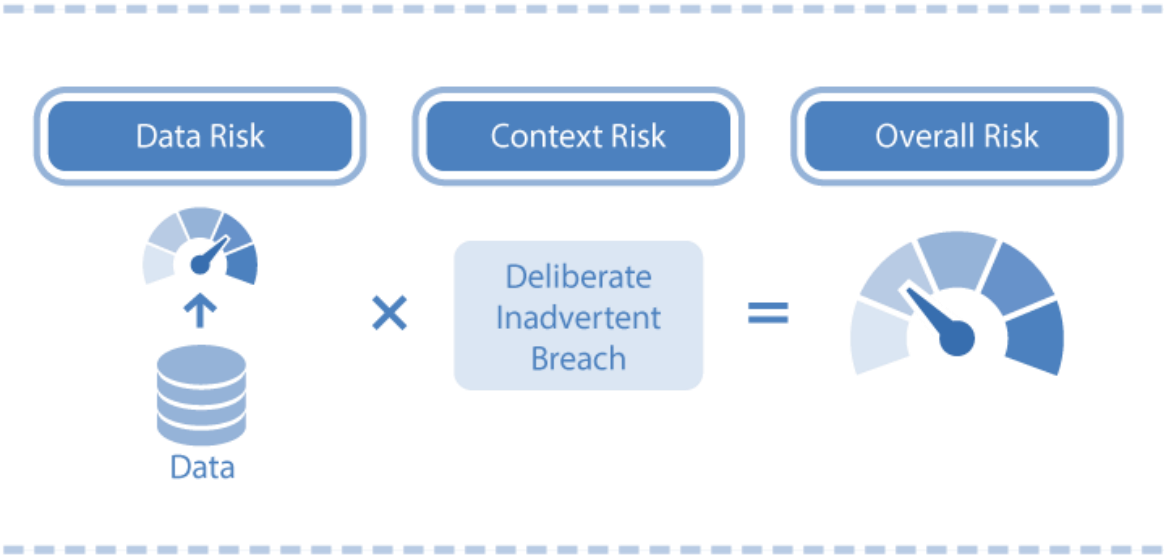
Measuring Data Risk

DIRECT IDENTIFIERS			QUASI-IDENTIFIERS		OTHER VARIABLES		
ID	Name	Telephone No.	Sex	Year of Birth	Lab Test	Lab Result	Pay Delay
1	~~John Smith~~	~~(412) 668-5468~~	M	1959	Albumin, Serum	4.8	37
2	~~Alan Smith~~	~~(413) 822-5074~~	M	1969	Creatine Kinase	86	36
3	~~Alice Brown~~	~~(416) 886-5314~~	F	1955	Alkaline Phosph		52
4	~~Hercules Green~~	~~(613)763-5254~~	M	1959	Bilirubin		36
5	~~Alicia Freds~~	~~(613) 586-6222~~	F	1942	BUN/Creatinine		82
6	~~Gill Stringer~~	~~(954) 699-5423~~	F	1975	Calcium, Seru		34
7	~~Marie Kirkpatrick~~	~~(416) 786-6212~~	F	1966	Free Thyroxine Index	2.7	23
8	~~Leslie Hall~~	~~(905) 668-6581~~	F	1987	Globulin, Total	3.5	9
9	~~Douglas Henry~~	~~(416) 423-5965~~	M	1959	B-type Natriuretic peptide	134	38
10	~~Fred Thompson~~	~~(416) 421-7719~~	M	1967	Creatine Kinase	80	21

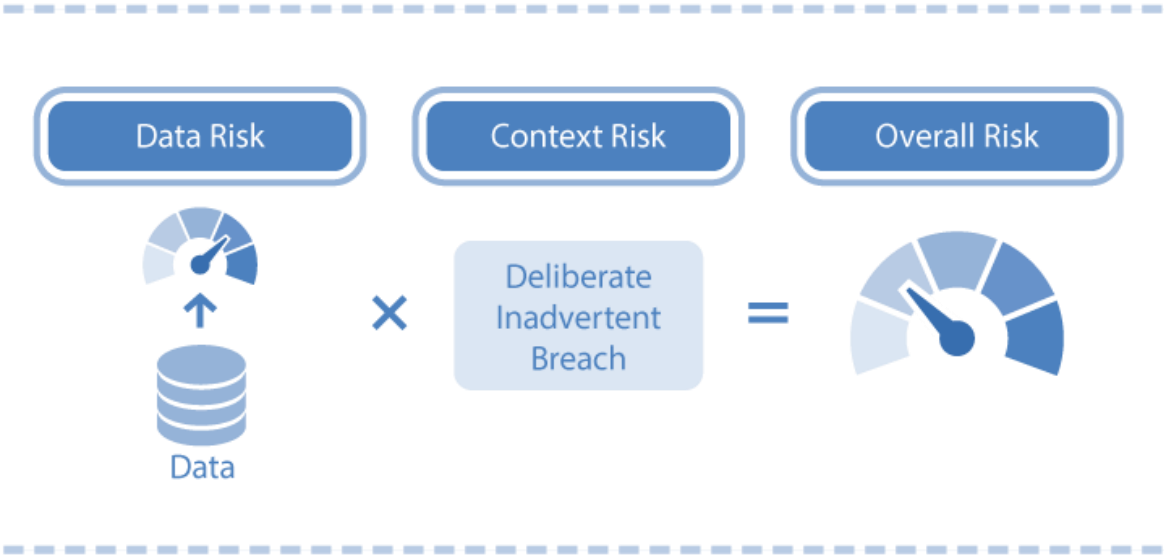
3

Two quasi-identifiers matching in three cells within a data set

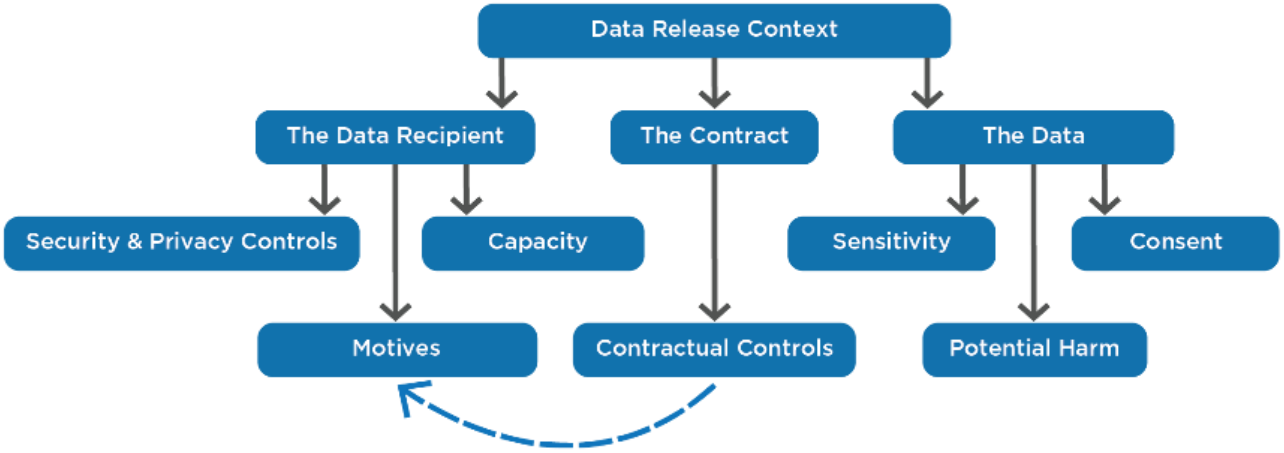
Overall Risk



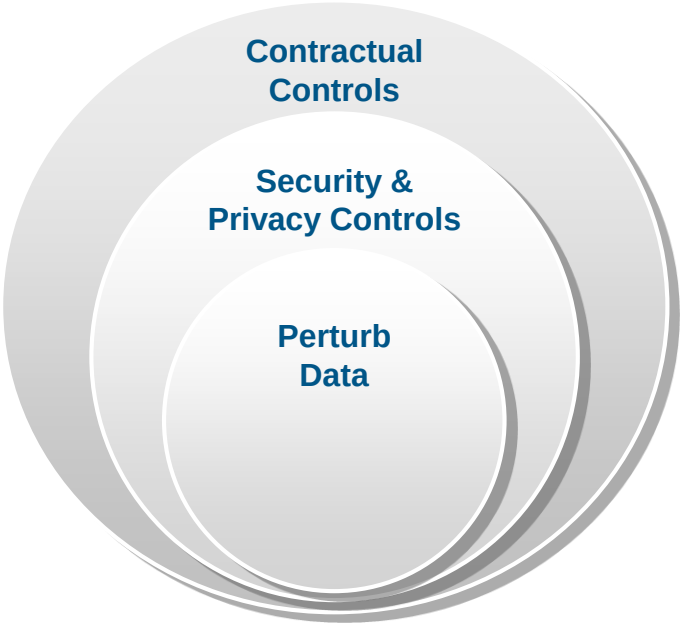
Overall Risk



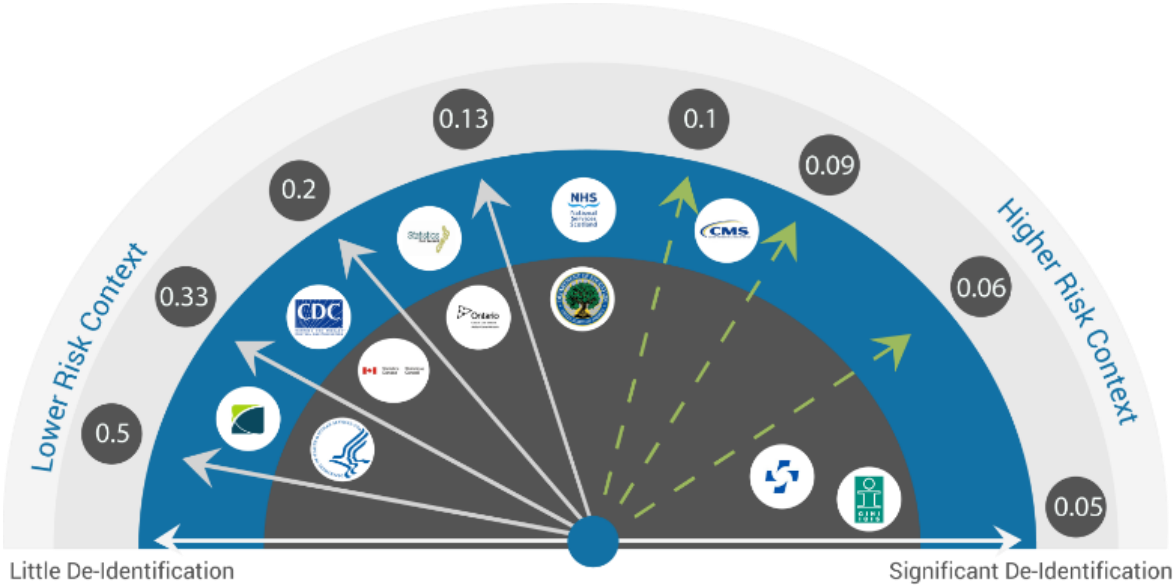
Context of Data Sharing



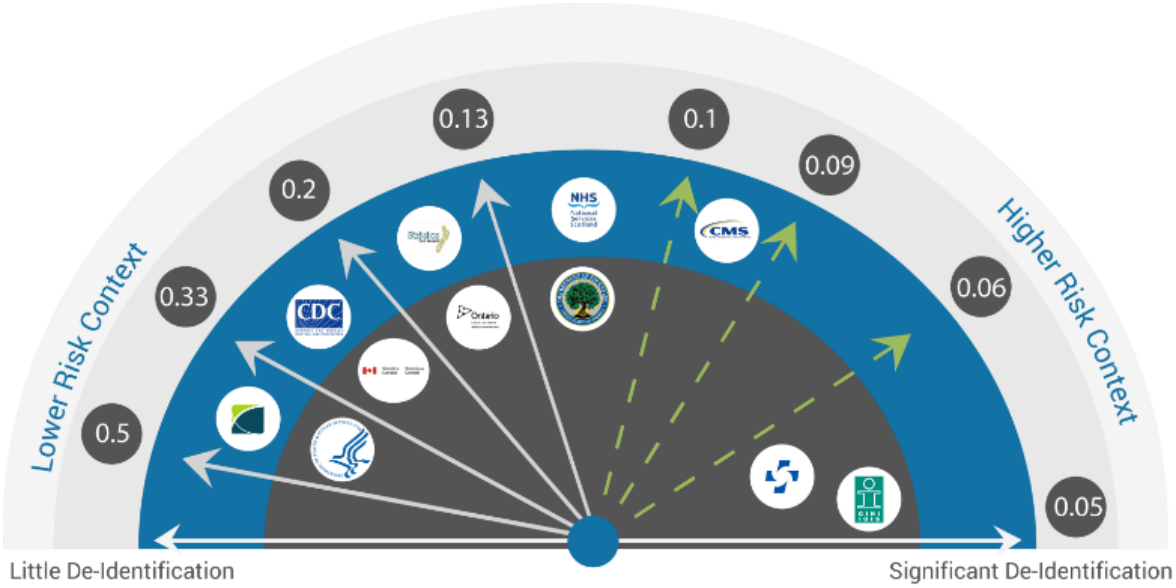
Layers of Protection



Precedents for Thresholds



Precedents for Thresholds



The HITRUST De-ID Framework

- After reviewing multiple De-ID programs and methods, HITRUST believes no one method is appropriate for all organizations
- Instead, HITRUST has identified twelve criteria for a successful De-ID program and methodology that can be scaled for use with any organization
- These twelve characteristics are divided into two general areas:
  - De-ID Program
  - De-ID Methodology

The HITRUST De-ID Framework

- After reviewing multiple De-ID programs and methods, HITRUST believes no one method is appropriate for all organizations
- Instead, HITRUST has identified twelve criteria for a successful De-ID program and methodology that can be scaled for use with any organization
- These twelve characteristics are divided into two general areas:
  - De-ID Program
  - De-ID Methodology