aukcar.ac.uk
@AUKCAR
@AryellyR

PhD studentship at the
Asthma UK Centre for Applied Research (AUKCAR):

Are anonymised databases truly anonymous?

PhD Medical Informatics – Part time: 6 years

Ref: AUKCAR-17-01a

Evaluating Re-Identification Risks scores in Publicly Available Clinical Trial Datasets: Insights and Implications

07 June 2024

By Aryelly Rodriguez, Steff Lewis, Sandra Eldridge, Tracy Jackson, Chris Weir

Asthma UK Centre for Applied Research

ASTHMA+ LUNG UK

THE UNIVERSITY of EDINBURGH

This work is funded by CMVM/UoE. This work is carried out with the support of the Asthma UK Centre for Applied Research [AUK-AC-2012-01].

aukcar.ac.uk
@AUKCAR
@AryellyR

EDITORIALS

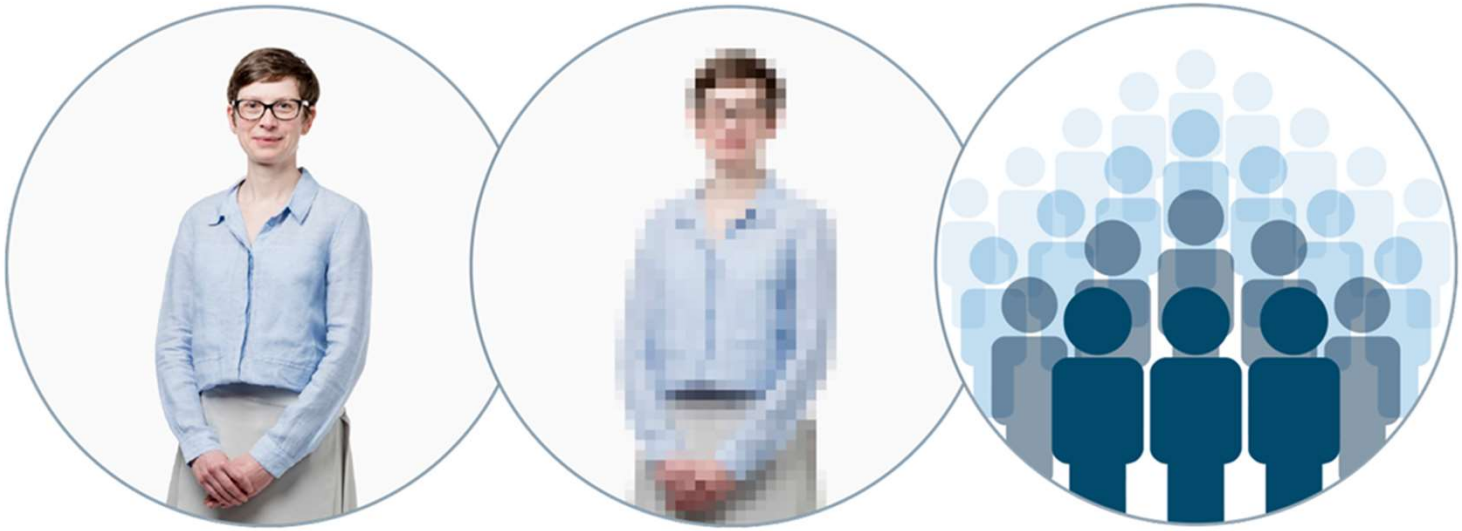
Mandatory data and code sharing for research published by *The BMJ*

New policy requires authors to share analytic codes from all studies and data from all trials

Elizabeth Loder, Helen Macdonald, Theodora Bloom, Kamran Abbasi

From 1 May 2024

Spectrum of identifiability



Identifiable De-identified Anonymous

More identifiable — Less identifiable

Do we know if these datasets pose a privacy risk for participants?

How anonymisation should be done?

Clinical Trials

SCT
SOCIETY FOR CLINICAL TRIALS

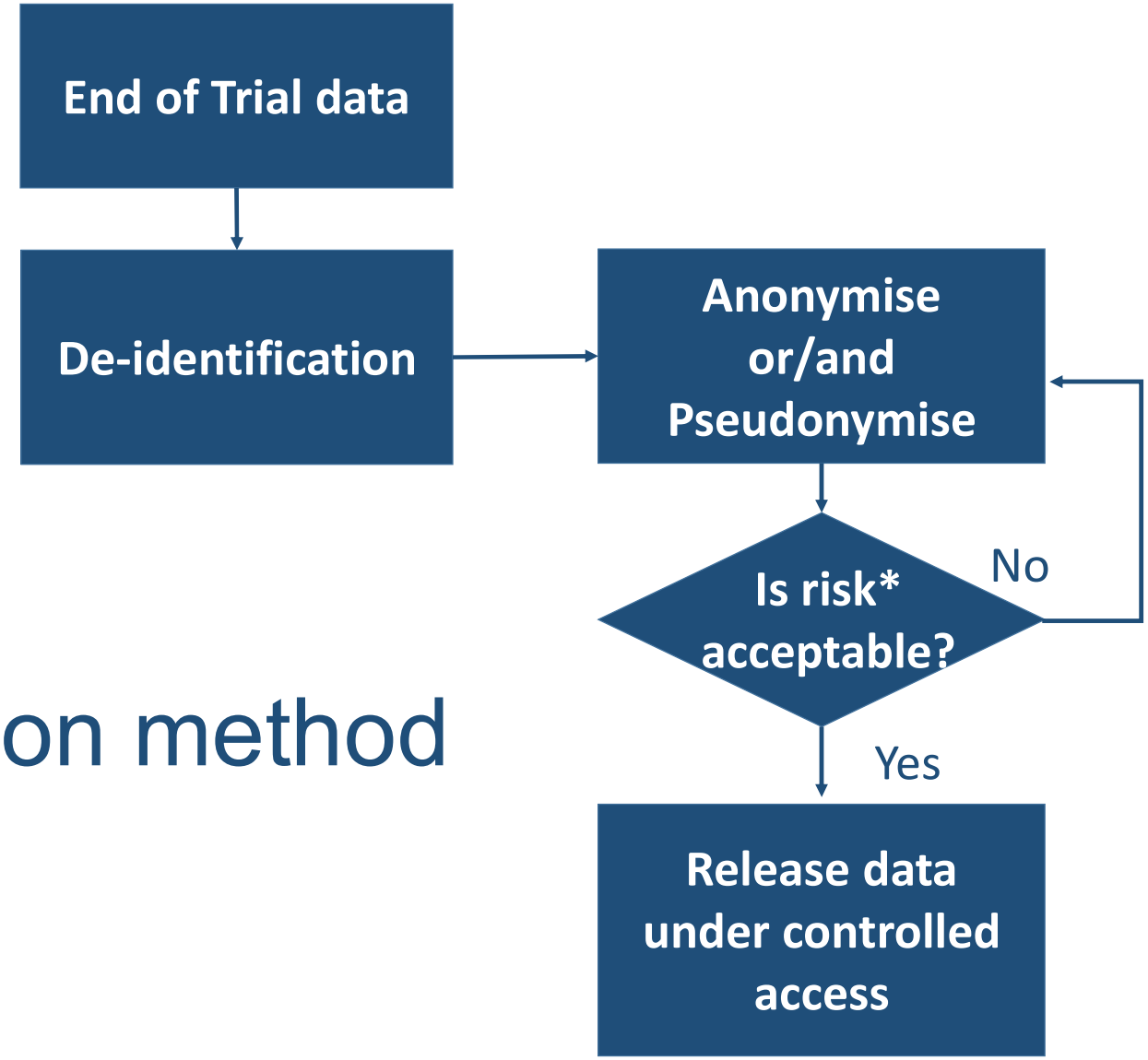
2.599 Impact Factor
5-Year Impact Factor 2.989
Journal Indexing & Metrics »

Current recommendations/practices for anonymising data from clinical trials in order to make it available for sharing: A scoping review

First published online June 22, 2022

Aryelly Rodriguez, Christopher Tuck, [...], and Christopher J Weir +5 View all authors and affiliations

Most common method



Is risk* acceptable?

No

Yes

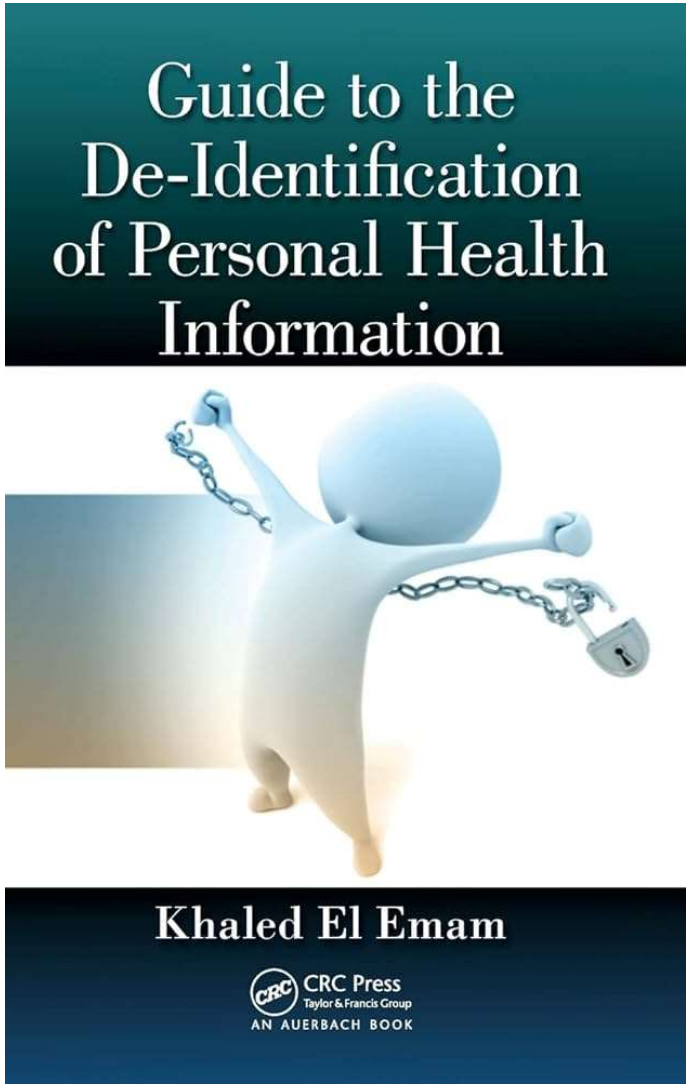
But what are those risks, and how do we assess them?

Datasets Variables Assessment

aukcar.ac.uk
@AUKCAR
@AryellyR

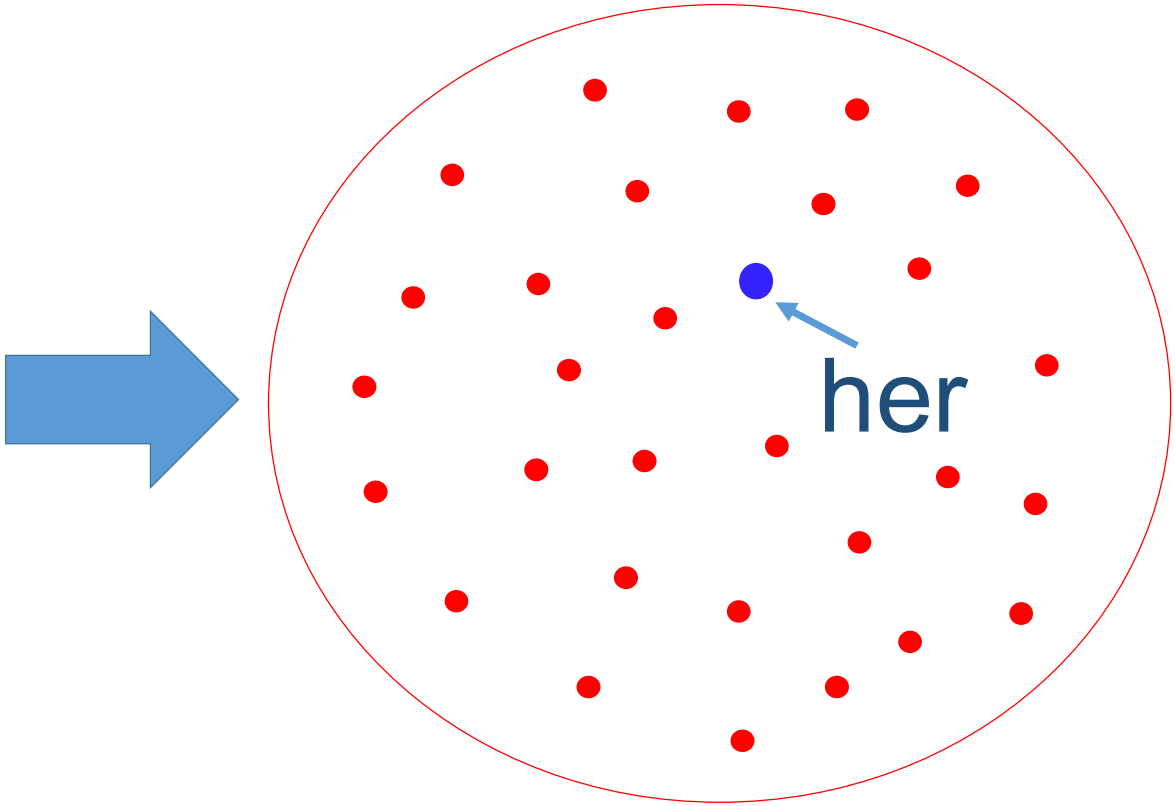
Hrynaszkiewicz et al. 28 items of personal and clinical information (UK and Europe)	
Direct Identifier - Remove	**Indirect Identifier – Carefully check**
• Name • Initials • Date of birth	• Sex • Place of birth • Occupation • Place of work • Ethnicity

Re-identification risks scores Calculation



Prosecutor Scenario

aukcar.ac.uk
@AUKCAR
@AryellyR



Prosecutor Scenario - Example

aukcar.ac.uk
@AUKCAR
@AryellyR

3 indirect identifiers			
Race	Age	Sex	
		Male	Female
White	>=18 and <70 years	95	28
	>+70	233	137
Other	>=18 and <70 years	19	4
	>+70	13	8

Description	Value
All Patients	537
All unique groups	8
Maximum Risk	25%
Average Risk	7%
Groups above 20%	1
People on groups above 20%	4
Above 20% Threshold Risk	0.7%

Re-identification risks scores – Results

aukcar.ac.uk
@AUKCAR
@AryellyR

Re-identification risks scores – Results – All 70 Datasets

aukcar.ac.uk
@AUKCAR
@AryellyR

	Mean	Median	Min	Max
No. of participants	4404	355	>10	>25000
No. of variables	968	237	>9	>20000
No. of indirect identifiers	4.2	4	0	>8
Above 20% Threshold Risk	79%	100%	0%	100%
Maximum Risk	91%	100%	0%	100%
Average Risk	74%	95%	0%	100%

- Re-identification risk scores:
  - Feasible and high in magnitude
  - Not affected by type access
  - Do not describe the actual risk
  - Could help inform the data sharing process

Risk Calculation – Next step

Is this going to help someone, or is it already in use?

aukcar.ac.uk
@AUKCAR
@AryellyR

Online Survey

UK researchers' views regarding their experiences with anonymisation

Asthma UK Centre for Applied Research

THE UNIVERSITY of EDINBURGH

29% knew about re-identification risk scores, but they did not use them

There are gaps in guidance and training

Data sharing consumes a lot of resources

Image by redgreystock on Freepik.com

aukcar.ac.uk
@AUKCAR
@AryellyR

"Not sharing" is no longer an option

Available data does not means free (at not cost)

Re-identification risk scores are feasible and useful

Still gaps in training and resources

aukcar.ac.uk
@AUKCAR
@AryellyR

Many thanks to my funders

Asthma UK Centre for Applied Research

ASTHMA+ LUNG UK

THE UNIVERSITY of EDINBURGH

my supervisors:

Prof S Lewis (University of Edinburgh) steff.lewis@ed.ac.uk
Prof C Weir (University of Edinburgh) christopher.Weir@ed.ac.uk
Prof S Eldridge (Queen Mary University of London) s.eldridge@qmul.ac.uk
Dr T Jackson (University of Edinburgh) Tracy.Jackson@ed.ac.uk

and my husband and children

Asthma UK Centre for Applied Research
ASTHMA+ LUNG UK
THE UNIVERSITY of EDINBURGH

aukcar.ac.uk
@AUKCAR
@AryellyR

Thank you!

Any questions?

Asthma UK Centre for Applied Research

ASTHMA+ LUNG UK

THE UNIVERSITY of EDINBURGH