Opening Access to Patient Data – Finding a Balance

David Buckeridge, MD PhD FRCPC
Professor | School of Population and Global Health, McGill University
CIHR Chair | E-Health Interventions
Medical Director | RI-MUHC Data Warehouse
david.buckeridge@mcgill.ca

McGill

Open Science in Action
Panel on Open Science and Patient Contributions
November 18, 2019, Montreal, Canada

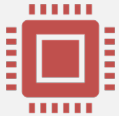
Privacy and Patient Data

A fundamental expectation

Harms are possible if privacy is not protected

Has tended to be a predominant consideration

Access to Patient Data for Research



Supports innovation, data are the ‘new oil’ for a learning health system

Social and economic implications are large

Has tended to be a secondary consideration


theglobeandmail.com
THE GLOBE AND MAIL
CANADA
MEMBER BENEFITS
SURVEILLANCE.LAB
Ontario's Health Ministry mulls sharing health data with researchers, certain third parties as part of privacy update
JOSH O'KANE
PUBLISHED NOVEMBER 13, 2019
UPDATED NOVEMBER 14, 2019
FOR SUBSCRIBERS
23 COMMENTS
SHARE

Increasing Demands for Access

Big data are needed for precision medicine

AI requires big computing next to data

Biased or missing data skews analyses and reinforces inequalities

Two Strategies for Opening Access and Protecting Privacy

Anonymization

Access control

Anonymization

Removes personally identifiable information (e.g., names, dates, locations, unique variables)

Trades off information loss against risk of re-identification

There are limits in theory and practice

nytimes.com

The New York Times

PLAY THE CROSSWORD

Google and the University of Chicago Are Sued Over Data Sharing

Sundar Pichai, Google's chief executive, testifying at a House of Representatives hearing last year. Google is building artificial intelligence that can diagnose medical conditions, which is raising privacy concerns. Sarah Silbiger/The New York Times

By Daisuke Wakabayashi

June 26, 2019

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-10933-3 OPEN

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher[1,2,3], Julien M. Hendrickx[1] & Yves-Alexandre de Montjoye[2,3]

While rich medical, behavioral, and socio-demographic data are key to modern data-driven research, their collection and use raise legitimate privacy concerns. Anonymizing datasets through de-identification and sampling before sharing them has been the main tool used to address those concerns. We here propose a generative copula-based method that can accurately estimate the likelihood of a specific person to be correctly re-identified, even in a heavily incomplete dataset. On 210 populations, our method obtains AUC scores for predicting individual uniqueness ranging from 0.84 to 0.97, with low false-discovery rate. Using our model, we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR and seriously challenge the practical and legal adequacy of the de-identification release-and-forget model.

"... we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes."

Access Control

Requires researchers to access data in secure location

Data cannot be removed, only approved results

Limits linkage, sophisticated analysis

Example of MUHC Data Warehouse

Centre universitaire de santé McGill
McGill University Health Centre

Requests for access directed by default towards anonymized data

Encrypted data made available in secure cloud environment

Investigators sign data use agreement and additional approvals for external collaborations

Closing Reflections

Placing more weight on anonymization reinforces focus on certain types of data and health contexts

Value of data for health system improvement is blurring the line between research and business of healthcare

GREGORY BARBER MEGAN MOLTENI BUSINESS 11.11.2019 09:27 PM

Google Is Slurping Up Health Data—and It Looks Totally Legal

Tech giants can access all of your personal medical details under existing health privacy laws. The question is how else that data might get used.

PHOTOGRAPH: ALBERTO PEZZALI/GETTY IMAGES

Last week, when Google gobbled up Fitbit in a $2.1 billion acquisition, the talk was mostly about what the company would do with all that