ETHICAL IMPLICATIONS OF USING THE IDI DATABASE FOR HEALTH RESEARCH

Noor Aljawahiri- MBCHB IV, under the supervision of Dr Hiran Thabrew from the Department of Psychological Medicine.

THE UNIVERSITY OF **AUCKLAND**
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

MEDICAL AND HEALTH SCIENCES

WHY DID WE START THIS PROJECT?

- Data from New Zealand national health databases held within the Integrated Data Infrastructure (IDI) have recently been combined, enabling insight into the prevalence and changes over time in health problems.

- Despite government sanction for the research use of this data, associated ethical issues have not been formally evaluated.

- This project aims to review information about the ethics of using 'big data' for research purposes and to develop initial guidance for NZ researchers wanting to use these datasets.

2

WHAT IS BIG DATA?

"Big data involves large sets of data with diverse levels of analysable structuration, coming from heterogeneous sources (online data, social media profiles, financial records, self-tracked parameters, etc.), produced with high frequency and which can be further processed and analysed using computational techniques." [1]

1-Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E (2018) Considerations for ethics review of big data health research: A scoping review. PLoS ONE 13(10): e0204937

Image:https://www.youtube.com/watch?v=TzxmjbL-i4Y

BIG DATA IN THE HEALTHCARE SECTOR –
THE 21st CENTURY REVOLUTION

"Health-related big data is the umbrella term used to describe extremely large and heterogeneous data sets that may be analysed computationally to reveal patterns, trends, and correlations, that have relevance for human health." [1]
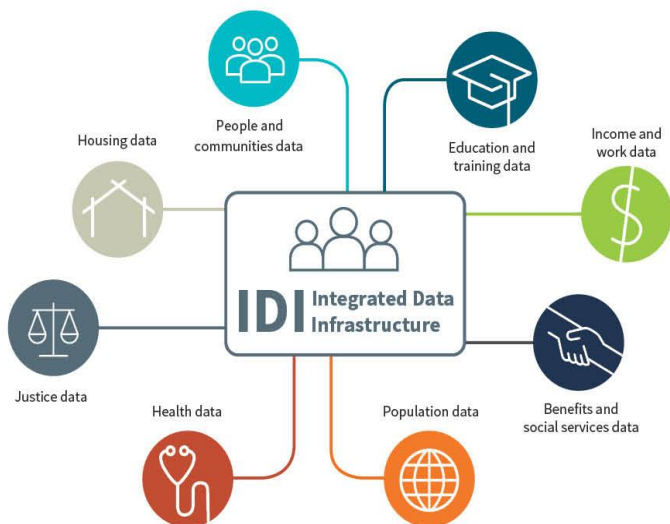
- Electronic health records (EHRs)
- Data from mobile health (mHealth) applications
- Medical Blogs
- Web networks
- Healthcare robotics
- Direct-to-consumer genetic and screening tests
- Online personal dietary programs
- Fitness club memberships
- Twitter hashtags

4

Image: https://www.simplilearn.com/big-data-in-healthcare-sector-rar359-article

WHAT IS THE IDI?

Data in the IDI September 2018

Stats NZ
Tatauranga Aotearoa

Stats NZ's Integrated Data Infrastructure (IDI) is a large research database containing de-identified microdata about people and households.

Housing data

People and communities data

Education and training data

Income and work data

IDI Integrated Data Infrastructure

Justice data

Health data

Population data

Benefits and social services data

The IDI contains person-centred microdata from a range of government agencies, Stats NZ surveys including the 2013 Census, and non-government organisations. For more information about data in the IDI, see
www.stats.govt.nz/integrated-data/integrated-data-infrastructure

The Longitudinal Business Database (LBD) complements the IDI with microdata about businesses. For more information about data in the LBD, see
www.stats.govt.nz/integrated-data/longitudinal-business-database

Health data

- B4 School Checks – from 2011
- Cancer registrations – from 1995
- Chronic conditions – from 2007
- General medical services claims – from 2002
- Health tracker – 2006-13
- Laboratory claims – from 2003
- Mortality – from 1988
- Immunisation – from 2006
- National non-admitted patient collection – from 2007
- Pharmaceuticals – from 2005
- PHO enrolments – from 2003
- Population cohort demographics and addresses – from 2004
- Mental health and addiction – from 2008
- Publicly funded hospital discharges – from 1988
- National Needs Assessment and Service Coordination Information System (SOCRATES)
- Maternity – from 2003

Education and training data

- Early childhood education participation – from 2008
- Primary education – from 2007
- Secondary education – from 2004
- Tertiary education – from 1994
- Industry training – from 2001
- Targeted training – from 2001
- Adult competency assessments – from 2014

Benefits and social services data

- Benefits – from 1990
- Youth services – from 2004
- Children's Action Plan – from 1996
- Working for Families – from 2003
- Child, Youth, and Family – from 1991
- Student loans and allowances – from 1992
- ACC injury claims – from 1994
- Family Start – from 2008

Justice data

- Recorded crime: offenders – from 2009
- Recorded crime: victims – from 2014
- Court charges – from 1992
- Sentencing and remand – from 1998

People and communities data

- Auckland City Mission – from 1996
- Migrant Survey – from 2012
- Driver licence and motor vehicle registers
- Longitudinal Immigration Survey of NZ – 2005-09
- General Social Survey – 2008-2016
- Disability Survey – 2013
- Te Kupenga – 2013

Population data

- Border movements – from 1997
- Visa applications – from 1997
- Departure and arrival cards – from 1997
- 2013 Census
- Births, deaths, marriages, and civil unions – from 1840

Income and work data

- Tax and income – from 1999
- NZ Income Survey – from 2006
- Household Labour Force Survey – from 2006
- Survey of Family, Income, and Employment – 2002-10
- Household Economic Survey – from 2006

Housing data

- Tenancy – from 2000
- Social housing – from 1980

WHAT QUESTIONS ARE WE TRYING TO ANSWER?

1)What ethical issues associated with the use of national datasets and other forms of 'big data' have other international researchers found?

2)What do NZ people think about the use of IDI data in health research?

3)What do specific groups such as Maori and 16-18-year-olds think of the use of IDI data? How does this differ from non-Maori or older participants?

4)Are the ethical issues raised from our study similar/different to those found by international researchers?

WHAT DOES OUR PROJECT INVOLVE?

- A scoping review of the available literature on people's views regarding the use of their health-related data and national datasets or 'big data' for research

- A qualitative analysis of NZ people's views on the use of their IDI-linked health data for research. Specific sampling will be undertaken to ensure adults, Maori participants and young people (aged 16-18yrs) are included.

- The scoping review will provide insight into the patient views and ethical issues raised by international researchers. The qualitative study will enable us to compare these views to the those of New Zealanders and to identify any culturally-relevant issues (especially those that are relevant to Maori).

METHODS:

1- Interviews: A minimum of 40 interviews will be undertaken in 2019, with participants recruited from ADHB outpatient clinic (10 of those participants being Maori patients, 10 being 16-18 year olds, and 20 being any gender/ethnicity.) These interviews will be 30-60 min long, involving questions to gauge the views of NZ people regarding the use of IDI data in research.

2- Scoping review: A review of international research into the ethical issues associated with the use of national datasets such as the IDI and other forms of 'big data' for research purposes.

3- Results from the interviews will be compared to the thematic analysis from the scoping review.

SCOPING REVIEW

Databases used: Medline, EMBASE, and Philosopher's Index (articles limited to 2009-2019, English only and Humans only)
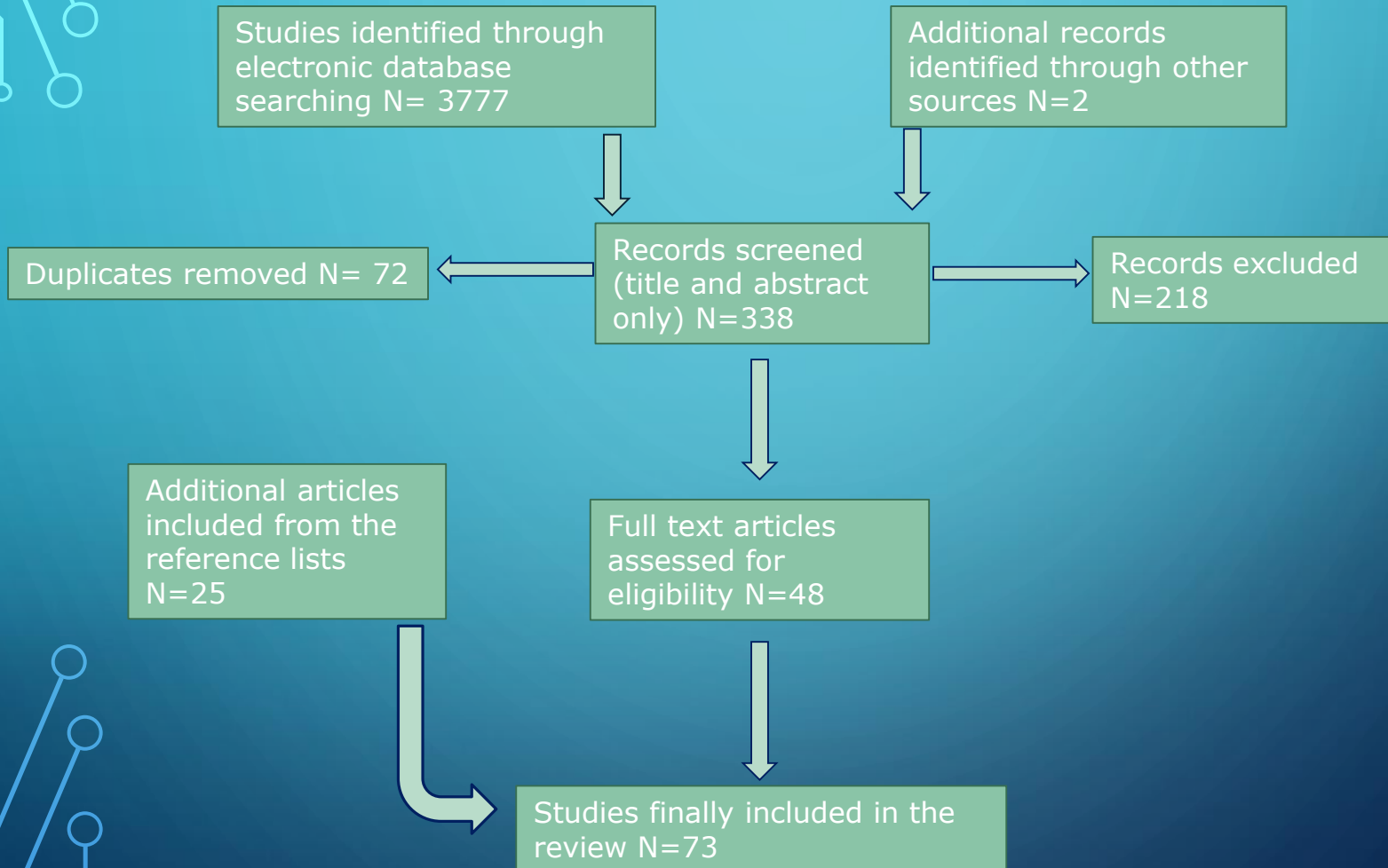
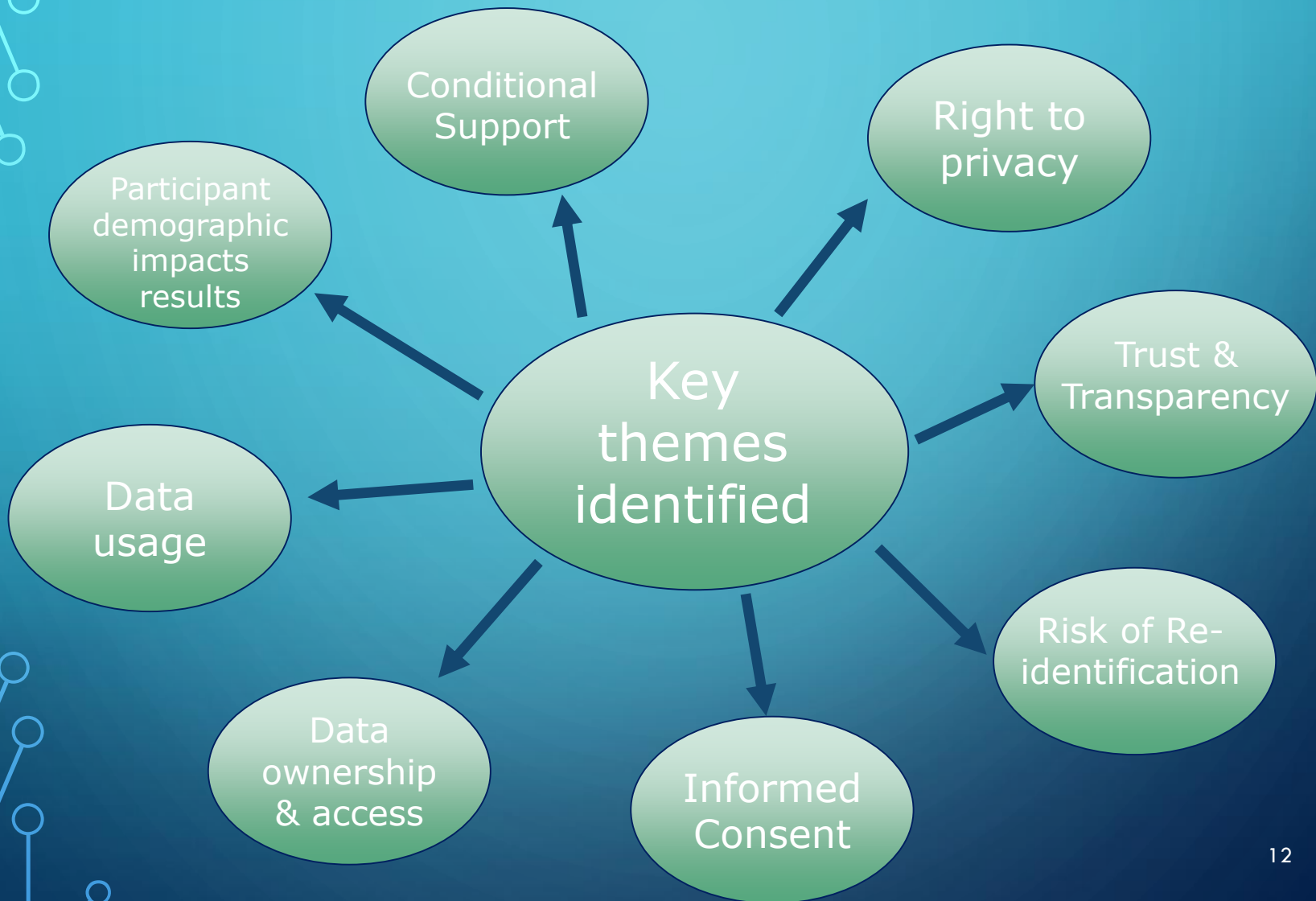| Key search terms: |
| --- |
| (Ethics/ OR moral*mp. OR exp Ethics,Research/ OR Ethics,Medical/ OR Policy/ OR governance.mp. OR Ethics Committees/es[Ethics] OR Confidentiality OR Informed Consent/es [Ethics] OR Privacy/ OR bioethic*.mp. |
| AND |
| bigdata.mp. OR data science.mp. OR datasets as topic/ OR Data Collection/ OR digital data.mp. OR Artificial Intelligence OR Medical Records Systems, Computerized/ OR Electronic Health Records/ OR Medical Records/ OR health records.mp. OR data set.mp. OR Dataset/ OR dataset.mp. OR national data.mp. OR integrated data infrastructure.mp. OR primary data.mp. |

10

PRISMA

Studies identified through electronic database searching N= 3777

Additional records identified through other sources N=2

Duplicates removed N= 72

Records screened (title and abstract only) N=338

Records excluded N=218

Additional articles included from the reference lists N=25

Full text articles assessed for eligibility N=48

Studies finally included in the review N=73

11

1- CONDITIONAL SUPPORT

All studies identified a general, widespread support for the use of health data in research, so long as this research was for the greater good, and serves public interest. However, most participants preferred this data to be anonymized, with appropriate consent measures taken before it's use.

2- RIGHT TO PRIVACY

The studies reviewed identified patient concerns about their right to privacy when their data is being accessed for secondary purposes. Some participants would only share their data if privacy was assured. This was found to be more evident when the data is considered more sensitive such as mental health, sexual health sexuality and religion.

13

3- TRUST & TRANSPARENCY

The level of trust individuals have for the researchers using their data, determines their willingness to share their information. Not surprisingly, insurance and for-profit industries were the least trusted, whereas doctors and hospitals were the most trusted. Transparency in data usage is also valued amongst participants, and is seen as crucial to develop and secure patient trust.

4- RISK OF RE-IDENTIFICATION

Some studies identified fear of participants being re-identified despite encryption of data in these databases. These participants indicated the necessity for companies/governments to ensure safety protocols are in place to protect their data, before they would consent to its use.

14

5- INFORMED CONSENT

All studies reviewed, indicated that patients preferred to be consented for the use of their data. Some preferred 'broad consent' to be obtained (this was also the preferred method for researchers themselves), however, the majority of patients preferred to be consented every time their data is used, this was seen as a means of providing individual control over the data. There seems to be public preference for flexible/varied consent models to allow patients to choose their preference. Degree of consent is affected by the degree of trust in the researchers using the data.

6-DATA OWNERSHIP & ACCESS

The studies reviewed identified that most patients thought that they owned their data, and were unaware of these databases storing their information, and being used for research. A majority showed interest in sharing ownership of this data, and preferred it to be owned by trusted individuals like DHBs rather than for-profit companies like insurance/pharmaceutical companies.

7- DATA USAGE

All studies identified a general support for data usage in research benefitting the public, as they see value in contributing to future advancements in healthcare; referred to as "research altruism". A fear of data being misused for commercial purposes such as insurance or pharmaceutical companies was evident.

8- PARTICIPANT DEMOGRAPHIC IMPACTS RESULTS

The studies identified that older people, people with a university degree, healthy individuals, and men are more willing to share their data than younger people, those without a university degree, women, and individuals with stigmatizing health conditions. When patients were interviewed/surveyed their main concerns were privacy, trust, and misuse of data, whereas, other stakeholder's interviewed (like researchers and data-owners) were more concerned about the legal implications of data-ownership along with the risk of depriving original investigators from their deserved credit and compensation.

MOVING FORWARD...

- Based on our findings, there is a clear support of the general public for the secondary use of their health data, as long as the research is for the greater good.

- There is a concerning difference between the opinions of the public/patients and researchers/data-owners.

- Some members of society seem to be misinformed about data ownership, consent processes, and protection protocols of this data.

- These are issues that should be addressed in the future when we want to consider facilitating secondary use of data.

- We will be conducting interviews with NZ patients in the near future and will compare the interview results to the results of our scoping review.

- This information will hopefully support the development of ethical guidelines for researchers who consider using this data and will guide future consent processed that need to be considered before 'Big Data' like IDI data is used in research.

Acknowledgements

I would sincerely like to thank the Health Research Council of NZ for funding my summer studentship.

Also special thank you to Dr Hiran Thabrew, Dr Vanessa Jordan, Dr Sarah Hetrick, Dr Megan Prictor, Dr Toni Shepherd, Dr Josefein Breedvelt and Mrs Anne Wilson for providing their guidance, support, and expertise throughout the research process.

Thank you for listening!