



Big Data and Medical Registries

Dr. Adrian Spörri, PhD MPH

SwissRDL, Medical Registries and Data Linkage

Institute of Social and Preventive Medicine

University of Bern, Switzerland

SwissRDL – Medical Registries and Data Linkage



- Center of excellence for medical registries
- >15 medical registries, from small local disease registries to national and international registries
- Plans, implements and maintains registries
- Develops and maintains its own registry software
- Local IT infrastructure at the University, with more than 50 virtual servers
- Team with epidemiologists, statisticians, monitors, data managers, programmers
- Linkage projects (Swiss National Cohort, South Africa Match Study)

What is Big Data?

A. Big numbers

- B. Big numbers
- C. Big amount of data

- D. Just a vogue word
- F. I don't know



193'954'946'875'345





Definition

Big data is

- huge amount of data
- unstructured
- variety of types
- from different resources
- complex data

Four V's

- Volume
- Velocity
- Variety
- Veracity



Four V's

- Volume: increases exponentially
- Velocity: real time generation
- Variety: text, audio, video, images, emails, messages
- Veracity: unreliable data quality, uncontrolled environments



Can Big Data be used for our research?

- A. Yes, the more data the better
- B. Yes, the truth lies in unstructured data
- C. No, as it doesn't follow clearly defined protocols
- D. No, the risk of loss of privacy is too big
- F. I don't know



Big Data in health care research

- Health insurances
- Data warehouse in University hospitals
- Internet searches, e.g. seasonal influenza
- Daily movement data, smart phones



How to analyze?

- temporal, spatial analyses
- pattern recognition
- correlations
- neural networks
- parallel processing



I.	PatID	Α	В	С	D
	100	34	yes	1020	good
	101	55	yes	5847	bad
	102	12	no	253	unknown

I think that



- A. Big Data replaces traditional hypothesis driven research
- B. Big Data has no potential for quality assessment
- C. Registries allow long term outcome research
- D. All of above
- F. I don't know



Epidemiological work depends to a great deal on routinely collected data

- e.g. Census data
- mortality data
- morbidity data
- and... registries

This report contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the World Health Organization.

WORLD HEALTH ORGANIZATION TECHNICAL REPORT SERIES

No. 365

EPIDEMIOLOGICAL METHODS IN THE STUDY OF CHRONIC DISEASES

Eleventh Report of the WHO Expert Committee on Health Statistics

WHO, Geneva 1967

Routinely collected data

- Census data, e.g. Switzerland 1850 2000
- Registrar General Act 1836 in England and Wales: data on death, births and marriages
- Population data often called vital statistics, census data, administrative data or population registry
- Needed for the ennumeration of the population at risk
- e.g. in CH 2015, all cancers: 164 deaths per 100'000 men and 106 deaths per 100'000 women

Definition of medical registry

"...defined as a data base of identifiable persons containing a clearly defined set of health and demographic data collected for a specific public health purpose."

(Solomon, D J et al. 1991)

Description of medical registries

3. DISEASE REGISTERS

Registration of individuals with certain chronic diseases can be an important epidemiological tool, provided that the registration covers a defined region (e.g., a province, state or city) and that the majority of cases in the region are registered. The term "registration" implies something more than "notification". A register requires that a permanent record be established, that the cases be followed up, and that basic statistical tabulations be prepared both on frequency and on survival. In addition, the patients on a register should frequently be the subjects of special studies. Even in the absence of a defined population base, useful information may be obtained from registers on the natural course of disease in different parts of the world.



What do we need registries for ...

... if we have RCTs?!

- RCTs are the gold standard in study designs and generate high level of evidence
- RCTs eliminate (patient) selection bias (randomization), and several other biases (blinding): high internal validity

... but they only answer the following question:

What do RCTs tell us?

CAN IT WORK? -> efficacy

- RCTs are experiments done in highly controlled artificial settings
- Low external validity: findings cannot easily be generalized to real life clinical settings

They cannot answer the following question:

The role of registries

DOES IT WORK? -> effectiveness

- Registries allow health technology assessments, reveal results of therapies applied in day-to-day routine clinical setting
- Not interventional, but reflect all facets and problems of clinical reality
- Have a limited internal validity...
- ...but high external, i.e generalizability

Registries can be used for

- Quality assurance: outcome, competence on level surgeon, therapist, hospital
- Outcome research, comparative effectiveness: e.g. surgical vs non-surgical
- Post market surveillance: e.g. efficiency or safety (revisions) of implants

Quality assessment

Indication:

Do we do the right thing?





Quality assessment

Therapy: conservative, operation Is it done correctly?





Quality assessment

Outcome: objective, subjective What is the result?





Patient recorded outcome measures

- Quality of treatment used to be evaluated mainly by medical personnel and clinical parameters
- Increasingly, the patients view is used to evaluated the quality
- E.g. Core outcome measures index, COMI (Mannion, 2009), EuroQol EQ5 (euroqol.org)
- Domains: pain, function, symptom specific well-being, general quality of life, social disability, work disability
- Registries: nested study, e.g. COMI Basel

Quality criteria for registries

- Completeness case ascertainment
 - compare numbers e.g. with hospital data, industry data, insurance data
 - => coverage
- Completeness missing data
- Validity accuracy of data entry
 - monitoring and audits
- Reporting correct analysis, reporting and interpretation
 - data management, analysis plans, reviews

Data quality

Measures

- before data collection
- during data entry
 - validations, type, range, warnings
- after data entry
 - data monitoring, e.g. change in frequencies, inconsistencies



Most important measures...

- prevent double data entry import existing data
- connect registry to clinic information systems
- reduce workload for medical personnel

Monitoring and audits

- Continuous monitoring in hospitals
- Simplify procedures, training of personnel
- Audits: Random sample of patient records are checked, using internal hospital data, e.g. operation reports
- Comparison of hospital data and registry data
- Quality report for each hospital

Experiences from running registries

- Sparse resources for monitoring and audits
- Trust in technical and automated checks and validations only
- Unrealistic expectation of fancy statistical procedures to resolve data issues
- => Promotion of quality measures needed

Registry is no Big Data

- Collect what is needed to answer most important questions
- Envision other existing data
- ⇒ make registries linkable (e.g. SSN, names, unique ID's)

- Informed consent
- Privacy preserving measures
- Legal aspects

Would I enter data into a registry?

- A. Yes, registries are an important source of quality assessment
- B. Yes, registry data are a great source for research
- C. No, me and my patient don't profit from registries
- D. No, I don't have time for additional administrative work
- F. I already enter data into a registry regularly



vote at srdl.participoll.com

Examples of registries

• For overview of medical registries, see for example FMH:

http://www.fmh.ch/saqm/ service/forum medizinische register.cfm

- Swiss Childhood Cancer registry SCCR
- Cantonal Cancer registries
- Mortality registry, Federal Statistical Office
- SIRIS: hip and knee implant registry
- see <u>www.SwissRDL.unibe.ch</u>

Why do we need registries?

- Self-evaluation: changes over time, improvements, changes in revision rates
- Benchmarking: comparison of own results with hospital and/or whole registry
- Benchmarking: outcome of hospitals (e.g. nosocomial infections, revision rates)
- New technologies: post marketing surveillance, serious adverse events
- Research: long-term effectiveness of new procedures, implants, therapies

Results of medical registries are used

- Health technology assessment
- Rapid development of new technologies and procedures
- Evaluation of effectiveness needed
- Demographic change: who needs what?
- Better value for money
- Sustainable financing of public health
- Internal and external quality management
- Marketing
- Regulations (e.g. new EU regulation for medical devices and implants)

Advantages of medical registries

- Multi-centric or even national
- High coverage if mandatory
- Quality assessment, locally at clinic level, regional or national
- Large numbers, if participation high
- Large numbers can reveal rare events (complications)
- Not time limited => reveal long term results of therapies or interventions

Disadvantages of medical registries

- Often voluntary
- Prone to selection bias
- Lack of monitoring or audits
- Heterogenious patient population
- Underfounded

Number of publications with medical registries pubmed.com



Outlook

- Increasing number of publications based on registries
- Medical association aware of the need for quality measures and registries
- National authorities push registries, e.g. national cancer registry, evaluation registries mandated by Federal Office of public health
- Supplier of medical device need registry data due to new regulations

Thank you for your attention



b UNIVERSITÄT BERN

Polling

- This is how big data starts: unstructured collection of data
- No protocol
- Not systematic
- What about privacy protection?
- What about question E?

Thank you for your attention



b UNIVERSITÄT BERN