An introduction to case finding and outcomes

Dr Harshana Liyanage
Department of Clinical & Experimental Medicine
University of Surrey
Objectives

- Problems with routine data
- Brief introduction to
  - Coded data
  - Ontologies
- Code list approach vs Ontological approach
- Three step ontological approach for identifying case
Problems with routine data

• Routinely collected datasets are:
  • Incomplete
  • Inconsistent

• Routine data often has errors:
  • Miscoding
  • Misclassification
  • Misdiagnosis

• Sub optimal data undermines:
  • measures of the quality of care and research
  • Psychological and financial implications
Approximately 40% of patients identified by computer searches (5.8% of people with diabetes) had errors; misdiagnosis is commonest.
Ontological approaches to the rescue...

Ontologies!!

Ontologies enable us to derive better outputs from sub-optimal data!
An **ontology** is a formal naming and definition of the types, properties, and interrelationships of the entities within a particular domain.
• “Weekly Bill of Mortality” was established in 1634
• WHO developed the International Classification of Diseases (ICD) in 1948
• Primary care in UK uses Read codes: developed by James Read (1980s)
  • Read ver2 (5 Byte)
  • CTV3 (5 Byte)
• Secondary care in UK:
  • ICD (disease classification)
  • OPCS-4 (procedure classification)
• SNOMED CT
  • Primary & secondary care (2020)
• Examine code dictionary/ code book for codes of interest

• Prepare lists of codes

• Perform data extraction (database administrator/developer)

• Clean dataset

• Analyse dataset
Problems with the code list approach

- Code hierarchy may change (newer versions)
- Utilisation of codes in code list cannot be guaranteed
- Data extraction specification is coding system dependent
- Limited scope for validating coding lists (dependency on code list creator/reviewer)
- Coding systems have inherent issues
Clinical Terminology Browser

C Endocrine or metabolic disease
C1 Other endocrine disease
C10 Diabetes mellitus
Clinical Terminology Browser

- Diabetes mellitus: autocecal dominant type 2
- Type 1 diabetes mellitus
- Type 2 diabetes mellitus
  - Type 2 diabetes mellitus with renal complications
  - Type 2 diabetes mellitus with ophthalmic complications
  - Type 2 diabetes mellitus with neurological complications
  - Type 2 diabetes mellitus with multiple complications
  - Type 2 diabetes mellitus with ulcers
  - Type 2 diabetes mellitus with gangrene
  - Type 2 diabetes mellitus with retinopathy
  - Type 2 diabetes mellitus - poor control
  - Rosent's syndrome
  - Type 2 diabetes mellitus without complication
  - Type 2 diabetes mellitus with mononeuropathy
  - Type 2 diabetes mellitus with polyneuropathy
  - Type 2 diabetes mellitus with nephropathy
  - Type 2 diabetes mellitus with hypoglyceemic coma
  - Type 2 diabetes mellitus with diabetic cardiomyopathy
  - Type 2 diabetes mellitus with peripheral neuropathy
  - Type 2 diabetes mellitus with arthropathy
  - Type 2 diabetes mellitus with neuropathic arthropathy
  - Insulin treated type 2 diabetes mellitus
  - Hyperosmolar non-ketotic state in type 2 diabetes mellitus
  - Type 2 diabetes mellitus with persistent proteinuria
  - Type 2 diabetes mellitus with persistent microalbuminuria
  - Type 2 diabetes mellitus with ketoacidosis
  - Type 2 diabetes mellitus with ketoacidotic coma
  - Type 2 diabetes mellitus with exudative maculopathy
  - Type 2 diabetes mellitus with gastroparesis
  - Maternally inherited diabetes mellitus
- Secondary pancreatic diabetes mellitus
- Diabetes mellitus induced by non-steroid drugs
Ontological process

- Ontological approach starts with concepts NOT codes
  - The ontological process should enable code lists used in research based on routine data to be constructed in a logical and open way.

- Interoperable
  - Can be used to handle data coded in other coding systems

- Reusable
  - This process will enable others to use the ontology and as is, update or modify it, or apply it to other coding systems.

- Use standards and tools to validate
Identify criteria for defining a Type 2 diabetes patient from routine data

10 minutes
Ontological approach for case finding
Identifying a case from routine computer data

- **Ontological layer**
  - Diagnostic criteria
  - Symptom & examination findings
  - Pathology and other test results
  - Therapies and other treatments

- **Coding layer**
  - Created in relevant coding systems

- **Logical data extract model**
  - Test extract
  - Results feedback into ontological layer

- **Final schema for identifying cases**
Step 1: Constructing the ontology

• The ontological layer defines the relevant concepts.

• For an ontology that defines a diagnosis, this might include aetiology, diagnosis and other clinical features of the condition and its therapy.

• Reflects the requirements and purpose of the investigation.

• An ontology for diabetes would explicitly set out the criteria used in a study so that it is possible to understand how a particular prevalence might be defined.

• Ontological case definition might be restricted to one or more categories of data or require a combination

• E.g. a case of Type 1 diabetes must have a Type 1 diabetes diagnostic code AND currently prescribed insulin.
Aetiology:

- Criteria that enable the validity of case identified in a population to be validated. The prevalence of most conditions is known.

- For example, Type 2 diabetes is rare in people under 30 years old, more common with increasing age and in men compared with women.

Diagnostic criteria:

- Recording of a diagnostic code for diabetes, or we might stipulate classification as either Type 1 or Type 2 diabetes.
- People with Type 1 diabetes mellitus require insulin for survival, whereas people with Type 2 have altered glucose metabolism and may or may not require insulin.

Symptom codes:

- Thirst, polydipsia, polyuria, and describing weight loss might be diagnostic of diabetes.
- The World Health Organisation (WHO) criteria for diagnosis of diabetes include abnormal blood glucose plus symptoms of diabetes; however the latter are rarely looked for in database studies.

Examination findings compatible with the diagnosis:

• Measured weight loss
• Smelling ketotic

Pathology test criteria:

• Fasting or random blood test results showing a raised glucose meeting the diagnostic criteria set out by the WHO
• Glycated haemoglobin (HbA1c) levels compatible with diabetes
• Urine tests positive for glucose

Medication and prescriptions:

- There are some medication and other prescribed items that imply a diagnosis of diabetes; others make the diagnosis unlikely.

- Some medicines, such as insulin, and some injectable and oral anti-diabetes drugs are used only in diabetes, whereas metformin is a medicine generally prescribed in diabetes but also used in other conditions.

- Prescriptions for testing for blood or urinary glucose or ketones make a diagnosis of diabetes more likely but not definite.

- For example, they may be prescribed in pregnancy or where there is impaired fasting glucose

Process of care codes:

- There are a number of codes associated with the process of delivery of care, remuneration and administration of care which imply **but do not make the diagnosis certain**.

- ‘Seen in diabetes clinic’ and ‘Attending diabetes clinic’. Most people with these codes in their records will have diabetes, but some people with gestational diabetes or impaired fasting glucose may also attend.

- **Excepted from diabetes quality indicators: informed dissent** – declines to attend for review (removes them from the practice pay-for-performance target payment).

- **DNA – Did not attend diabetic clinic** – is an example of an administrative code.
Treatment or procedure codes:

• There are very rare operations or other procedure codes related to diabetes.

• Surgery for very rare tumours can cure diabetes.
  • Glucagonoma
  • Phaeochromocytoma

(This heading is included for ontological completeness)
Each of the types of information included in the ontology should be included in the coding list.

If you restrict your ontology to one or more categories of information (e.g. simply to diagnosis), then the same will apply to the coding list.
Step 2: Coding layer – creating a coding list from the ontology

![Diagram showing the flow from Diabetes Mellitus Ontology to Code Lists (Read Ver 2, CTV 3, ICD, SNOMED).]
Step 3: Logical data extract model

- Check that it is possible to extract the data you anticipate.

- Sometimes codes do not have sufficient granularity.

- Just because a code exists within a terminology, do not expect that clinicians or those involved in data entry will necessarily use it!

- Literature reviews, pilot searches of data sources and speaking to practitioners in the field about their data recording.
Advantages of the ontological approach

- The ontological process should enable code lists used in research based on routine data to be constructed in a logical and open way.

- This process will enable others to use the ontology and as is, update or modify it, or apply it to other coding systems.

- Make use of a wide range of tools to building, manage and query ontologies
Thank you for listening!

Questions?
Clinical Ontologies

Dr Harshana Liyanage

Department of Clinical & Experimental Medicine
University of Surrey
Objectives

• Explain the basic idea of an ontology

• Identify problems that can be resolved using ontologies

• Explain applications of ontologies

• Identify potential use cases for ontologies

• Ontology standards and tools
Ontology:

onto- “being; that which is”
logia - "science, study, theory"
“An *ontology* is a formal, explicit specification of a shared conceptualization”

T. Gruber

What are the characteristics of a movie in a typical movie database?
### Movie Database - A

<table>
<thead>
<tr>
<th>Movie Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>Genre</td>
</tr>
<tr>
<td>List of actors</td>
</tr>
<tr>
<td>Director</td>
</tr>
<tr>
<td>Producer</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>Studio</td>
</tr>
<tr>
<td>Age Rating</td>
</tr>
<tr>
<td>Music Producer</td>
</tr>
<tr>
<td>Special FX Producer</td>
</tr>
<tr>
<td>Distributor</td>
</tr>
</tbody>
</table>

### Movie Database - B

<table>
<thead>
<tr>
<th>Movie Title</th>
<th>Year of release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genre</td>
<td>Cast</td>
</tr>
<tr>
<td>Director</td>
<td>Producer</td>
</tr>
<tr>
<td>Screen Writer</td>
<td>Age Rating</td>
</tr>
<tr>
<td>Official Website</td>
<td>Runtime</td>
</tr>
<tr>
<td>Budget</td>
<td>Gross Income</td>
</tr>
</tbody>
</table>
Alice’s movie database

Alice’s movie website

Bob’s movie Database

Bob’s movie website
## Movie Ontology

<table>
<thead>
<tr>
<th>Movie Title</th>
<th>Year of release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genre</td>
<td>Cast</td>
</tr>
<tr>
<td>Director</td>
<td>Producer</td>
</tr>
<tr>
<td>Music Producer</td>
<td>Screen Writer</td>
</tr>
<tr>
<td>Special FX Producer</td>
<td>Studio</td>
</tr>
<tr>
<td>Country</td>
<td>Official Website</td>
</tr>
<tr>
<td>Age Rating</td>
<td>Runtime</td>
</tr>
<tr>
<td>Budget</td>
<td>Gross Income</td>
</tr>
<tr>
<td>Distributor</td>
<td></td>
</tr>
</tbody>
</table>
Ontologies are expressed in computational logic-based languages so that knowledge can be exploited by computer programs.
Building Ontologies

Formal, Explicit Specification

Human readable version

Machine readable

Wednesday, 25 January 2017
"An ontology is a formal, explicit specification of a shared conceptualization"

T. Gruber

Gruber TR. The role of common ontology in achieving sharable, reusable knowledge bases. KR. 1991 Apr 22;91:601-2.
Data sharing problems

Possible use of ontologies

Research Dataset - A

Researcher - A

Research Dataset - B

Researcher - B
Gene Ontology

Initiative to unify the representation of gene and gene product attributes across all species
Interoperability Problem

Ideal Situation

- Hospital Database
- GP Database
- Pharmacy Database
Interoperability Problem

Actual Situation

Hospital Database

Pharmacy Database

GP Database
• Systematized Nomenclature of Medicine

• Will be introduced to NHS UK in 2017 / 18

• Facilitate integrated care across all health services in UK (GP, Hospitals etc)
Ontology Web Language – OWL

Notation for building Ontologies

(Machines can read OWL)
Tools for developing Ontologies

ONTOLOGY AUTHORING TOOLS

Protégé - Stanford Center for Biomedical Informatics Research
SPARQL

Language for querying Ontologies

(Questions can be asked across multiple ontologies)
Identify potential uses of ontologies in the domain you work in?
Summary

- Explain the basic idea of an ontology
- Identify problem that can be resolved using ontologies
- Explain applications of ontologies
- Identify potential uses of ontologies in other domains
Thank you for listening!

Questions??

Wednesday, 25 January 2017