Realizing the Potential of Secondary Uses of Clinical Data

Kent A. Spackman, MD PhD

Professor, Oregon Health & Science University

Chair, SNOMED International Editorial Board

OHSU
OREGON HEALTH & SCIENCE UNIVERSITY

SNOMED® INTERNATIONAL

Overview

- Definitions
- Successful example of collecting data for secondary uses
  - CORI-NED
- Standards required:
  - Reference terminologies
  - Data collection protocols and checklists
- Dilemmas & challenges

Definitions

- Primary use = “direct patient care”
- Secondary use = anything beyond primary

- but …

Why do clinicians record patient data?

- To aid their memory
- To legally document what they saw & did (and sometimes why)
- To communicate to other members of a team
- To support and justify reimbursement

- To satisfy requirements of protocols & systems
  - Research protocols
  - Minimum data sets
  - Professional guidelines
  - (? And some incidental constraints imposed by software)

Slightly different view of “secondary”

- Instead of focusing on “direct patient care”
- Secondary uses of clinical data are any uses other than the primary purpose(s) for which the data is recorded
  - ICD-9-CM coding for reimbursement can be derived from the dictated discharge summary, where the primary purpose may be documentation +/- communication, (not reimbursement).
  - Communicable disease reports to the health department can be derived from routine lab culture reports, where the primary purpose is communication to the ordering physician, (not epidemic detection).

The ideal

- Record clinical data once
  - with fidelity to the clinical situation
- Allow systems to derive needed data from that single instance of recording

The reality

- Clinicians find themselves entering the same basic clinical facts multiple times from slightly different perspectives for different purposes
- Reimbursement coding skews clinical data
  - The level of detail is tuned to optimize reimbursement
  - Sometimes the clinical reality is obscured by lack of fidelity in the coding options available

Examples where the ideal is beginning to work

- Microbiology laboratories
  - positive Salmonella culture
    - Reports go to the physician(s) caring for the patient
    - Reports also go to local/state reportable disease registries

Example of successful collection and use of secondary data: CORI - NED

- The Clinical Outcomes Research Initiative (CORI) was founded in 1995 by the ASGE as a national data repository for endoscopic research.
- The shared repository (National Endoscopic Database – NED) is designed to promote endoscopic research among GI physicians.
- CORI is a leading source of GI research information, receiving more than 20,000 clinical procedure reports from more than 750 physicians nationwide each month.
- CORI research data have been used to support more than 50 major research initiatives to date.
- CORI operates as a not-for-profit organization under the auspices of the ASGE. For more information, visit CORI's web site: www.cori.org

Clinical Outcomes Research Initiative (CORI) – National Endoscopic Database (NED)

- Free software for producing procedure reports
  - Starting with the CORI software, physicians can enter their procedure notes with point and click ease and produce a medical report in minutes.
- Local databases
  - Data is saved to provide the physician access to historical records and aggregate data provides new approaches to implement powerful quality assurance methods not known in the dictation-based practice.
- Expert technical services and support are provided at no charge to members of the consortium.
- Members participate in a growing consortium from around the world contributing to endoscopic research.
  - CORI participants agree to send procedure data to the NED weekly via an automated send function. No patient or endoscopist identifiers are transmitted. The data from CORI's participating sites are tested for quality and merged into the NED. The aggregate data are used for a variety of research purposes.

CORI - NED

- Since the 1995 introduction of the CORI software and National Endoscopic Database, over 750 endoscopists at 114 sites across the United States have a combined dataset consisting of 1,092,045 procedures.
- Colonoscopies, EGD's and flexible sigmoidoscopies account for approximately 95% of all procedures in the NED, while ERCP, EUS, motility, bronchoscopies and capsule endoscopies make up the remaining 5%.

CORI-NED: Lessons for Secondary Data

- Clinicians will accept some constraints and requirements on data collection if they get something in return. In this case:
  - free software
    - automated printed reports that
      - Can be sent to the referring physician sooner, relative to transcription of dictations
      - Do not require duplicate effort for billing / reimbursement and medicolegal documentation
  - contribution to a specialty-focused database
    - altruism/advancement of the profession
    - availability of pooled data for research

CORI-NED: Lessons for Secondary Data (2)

- Successful research data analysis and aggregation depends on the quality of data collection
  - Some fields, such as ethnicity, had to be made a required field to get adequate completeness
- The data collection process needs to be crafted and tuned for the individual practice case
  - Endoscopists designed and refined the menus/terminology choices based on their own reporting experience, with research analysis in mind

CORI-NED: Lessons for Secondary Data (3)

- Unfortunately, NED data is a “silo”
- Endoscopic data collected using other software, or on another terminologic basis, is not directly interoperable

- The terminology should be mapped to a common reference terminology
  - Such a mapping could provide interoperability with other “silos” of data

Terminology standards

- How far does a reference terminology take us towards being able to use secondary data?
  - Permits common reference points for meaning
  - With appropriate history mechanism, sustains the value of previously recorded data
  - Does not (independently) solve the problem of data collection / data entry

SNOMED Clinical Terms:

- Directly supports
  - Representation & queries based on meaning
  - Computable tracking of historical relationships of retired codes
- Indirectly supports
  - Specification of user interface
  - Definition of minimum data sets, checklists, and data collection standards

How does a reference terminology facilitate data re-use?

- Provides a common representation
- Independent of how the data was recorded:
  - natural language
    - e.g. English, French, Spanish, ...
  - terms
    - e.g. craniopharyngioma, Erdheim tumor, pituitary adamantinoma, Rathke's pouch tumor
  - data sets
  - information system interface
  - implementation details
  - type of site
  - type of user

Data re-use

- Requires the ability to:
  - query databases
  - systematically retrieve patients based on general criteria
  - aggregate data in ways not directly encoded
  - E.g.
    - If the patient has had an MI but has no CHF, AV block, asthma, peripheral vascular disease, or Type 1 diabetes mellitus, and is not taking a beta blocker, you need to consider adding beta blocker therapy ...

Reference Properties of Terminology

- Explicit representation of
  - formal semantic definitions
  - relationships between codes (is-a, finding-site, causative-agent, associated-morphology, …)

Is tuberculous ascites a kind of bacterial effusion?

- tuberculous ascites
  - Finding-site: peritoneal cavity
  - Associated-morphology: serous effusion
  - Causative-agent: Mycobacterium tuberculosis
- M.tuberculosis is-a Mycobacterium is-a bacterium
- Serous effusion is-a effusion

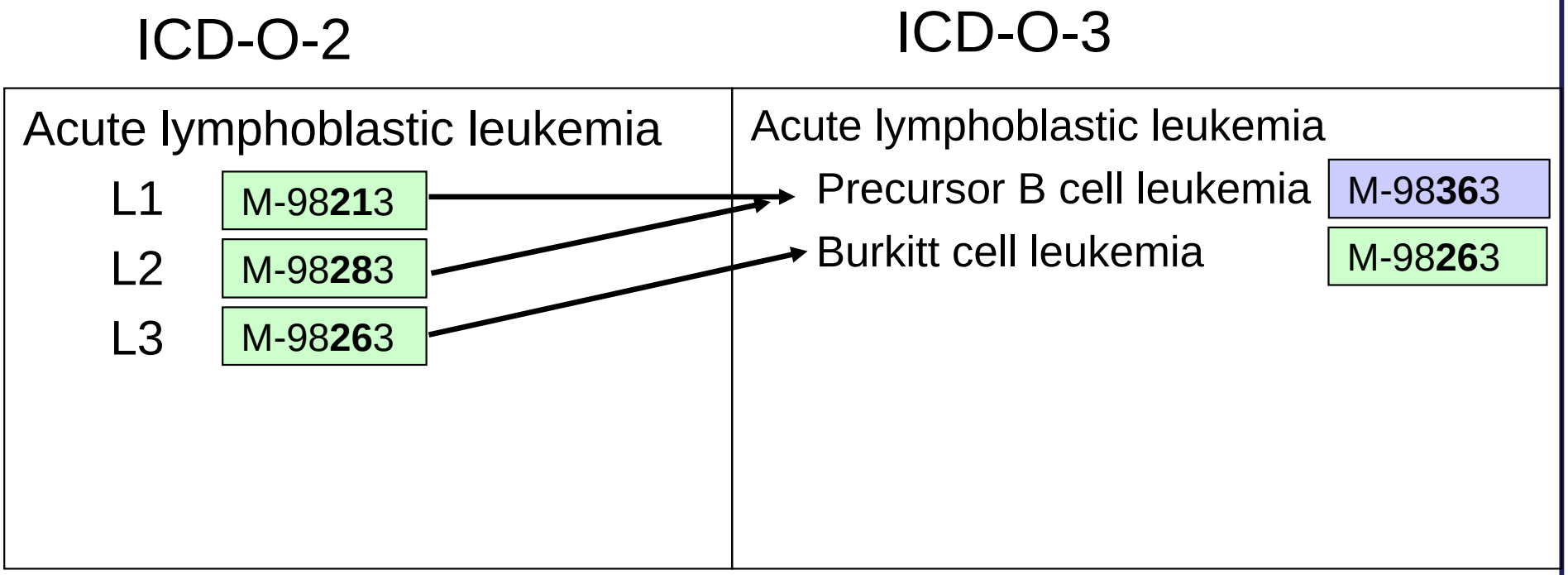
First Rule of Coding

- Yesterday’s data should be usable tomorrow

Secondary Use: Cancer Registry

- ICD-O version 2 was replaced in 2001 with ICD-O version 3.
- Example: Acute lymphoblastic leukemia
  - French-American-British classification had three subtypes: L1, L2, and L3
  - ICD-O-2 codes were:
    - L1: 9821/3
    - L2: 9828/3
    - L3: 9826/3

Acute lymphoblastic leukemia



SNOMED history table contains two rows:

M-98213 REPLACED-BY M-98363

M-98283 REPLACED-BY M-98363

First Rule of Data Quality

- The quality of the data collected is directly proportional to the care with which options are presented to the user

Data collection specifications

Two basic parts to the specifications:

1. Required elements
   - Specifies data elements that should be collected
   - E.g. “minimum data sets” specify these
2. How the elements should be described or coded

Illustration: CAP Cancer Protocols

- Non-Hodgkin lymphoma:
- Specification part 1: required elements
  - Specimen type
  - Tumor site
  - Histologic type
  - Extent of pathologically examined tumor
  - Phenotyping

Specification Part 2

- Non-Hodgkin's lymphoma histologic type
  - Classification has been changing
  - More than 25 different classifications have been published since 1925
  - Major classifications in past 30 yrs:
    - Rappaport
    - Working Formulation
    - Kiel
    - REAL
    - WHO

A portion of the SNOMED CT hierarchy

Bold=WHO
Underline=REAL

Both
neither

B-cell neoplasm
- **precursor B-cell neoplasm**
  - **precursor B-lymphoblastic leukemia/lymphoblastic lymphoma**
    - precursor B-cell lymphoblastic leukemia
    - precursor B-cell lymphoblastic lymphoma
- **mature (peripheral) B-cell neoplasm**
  - **chronic lymphocytic leukemia**
  - **prolymphocytic leukemia, B-cell type**
  - **malignant lymphoma, lymphoplasmacytic**
  - **mantle cell lymphoma**
  - **follicular lymphoma**
    - follicular lymphoma, cutaneous follicle center sub-type
  - follicular lymphoma, grade 1
    - follicular lymphoma, diffuse follicle center sub-type, grade 1
  - follicular lymphoma, grade 2
    - follicular lymphoma, diffuse follicle center cell sub-type, grade 2
  - follicular lymphoma, grade 3
  - diffuse predominantly small cell lymphoma
  - marginal zone B-cell lymphoma
    - **extranodal marginal zone B-cell lymphoma, mucosa-associated lymphoid tissue**
    - **nodal marginal zone B-cell lymphoma**
  - **splenic marginal zone B-cell lymphoma**
  - **hairy cell leukemia**
  - **diffuse large B-cell lymphoma - category**
    - **mediastinal large B-cell lymphoma**
    - malignant lymphoma, large B-cell, diffuse
      - **intravascular large B-cell lymphoma**
    - **primary effusion lymphoma**
  - **Burkitt lymphoma/leukemia**
    - Burkitt lymphoma
    - Burkitt cell leukemia
    - **endemic Burkitt's lymphoma**
    - **sporadic Burkitt's lymphoma**
    - **immunodeficiency associated Burkitt's lymphoma**
    - **atypical Burkitt's lymphoma**
  - high grade B-cell lymphoma, Burkitt-like
  - plasma cell myeloma/plasmacytoma
    - plasmacytoma - category
      - **plasmacytoma, extramedullary (not occurring in bone)**
      - plasmacytoma, bone
    - **plasma cell myeloma**
      - multiple myeloma
      - plasma cell leukemia

The REAL Classification of Lymphoid Malignancies

B-cell neoplasm

- **precursor B-lymphoblastic leukemia/lymphoblastic lymphoma**
- **chronic lymphocytic leukemia**
- **prolymphocytic leukemia, B-cell type**
- **malignant lymphoma, lymphoplasmacytic**
- **mantle cell lymphoma**
- **follicular lymphoma**
  - follicular lymphoma, grade 1
  - follicular lymphoma, grade 2
  - follicular lymphoma, grade 3
  - diffuse predominantly small cell lymphoma
  - marginal zone B-cell lymphoma
    - **nodal marginal zone B-cell lymphoma**
- **splenic marginal zone B-cell lymphoma**
- **hairy cell leukemia**
- **mediastinal large B-cell lymphoma**
- malignant lymphoma, large B-cell, diffuse
- Burkitt lymphoma
- high grade B-cell lymphoma, Burkitt-like
- plasma cell myeloma/plasmacytoma

Underlined entries are also in the WHO classification

The WHO Classification of Lymphoid Malignancies

B-cell neoplasm

- **precursor B-cell neoplasm**
  - **precursor B-lymphoblastic leukemia/lymphoblastic lymphoma**
- **mature (peripheral) B-cell neoplasm**
  - **chronic lymphocytic leukemia**
  - **prolymphocytic leukemia, B-cell type**
  - **malignant lymphoma, lymphoplasmacytic**
  - **mantle cell lymphoma**
  - **follicular lymphoma**
  - **extranodal marginal zone B-cell lymphoma, MALT type**
  - **nodal marginal zone B-cell lymphoma**
  - **splenic marginal zone B-cell lymphoma**
  - **hairy cell leukemia**
  - **diffuse large B-cell lymphoma**
    - **mediastinal large B-cell lymphoma**
    - **intravascular large B-cell lymphoma**
    - **primary effusion lymphoma**
  - **Burkitt lymphoma/leukemia**
    - **endemic Burkitt's lymphoma**
    - **sporadic Burkitt's lymphoma**
    - **immunodeficiency associated Burkitt's lymphoma**
    - **atypical Burkitt's lymphoma**
  - **plasmacytoma, extramedullary (not occurring in bone)**
  - **plasma cell myeloma**

Underlined entries are also in the REAL classification

Full hierarchy vs ICD-O vs REAL Classification vs WHO Classification

- Some distinctions are now less important clinically
  - Lymphoma/leukemia distinction
    - Consensus is that these are different phases of the same disease, esp. for B-cell
  - "Non-Hodgkin's lymphoma" category
- Some distinctions are more important
  - B cell vs T/NK cell origin

Dermatology Classification (EORTC) of primary cutaneous lymphomas

- Cutaneous T-cell lymphoma (CTCL)
  - **Mycosis fungoides**
  - Follicular mycosis fungoides
  - Pagetoid reticulosis
  - **CTCL, large cell, CD 30-positive**
  - Lymphomatoid papulosis
  - **Sezary syndrome**
  - CTCL, large cell, CD 30-negative
  - CTCL, pleomorphic, small/medium-sized
  - **Subcutaneous panniculitis-like T-cell lymphoma**
- Cutaneous B-cell lymphoma (CBCL)
  - Primary cutaneous immunocytoma/marginal zone B-cell lymphoma
  - Primary cutaneous follicle center cell lymphoma
  - Primary cutaneous large B-cell lymphoma of the leg
  - Primary cutaneous plasmacytoma
  - **Intravascular large B-cell lymphoma**

Different views

- Primary cutaneous lymphomas vs lymphoid neoplasms in general
  - EORTC ignores non-cutaneous forms
  - EORTC lumps some primary cutaneous forms that are split in the WHO classification
    - E.g. primary cutaneous follicle center cell lymphoma includes the primary cutaneous forms of:
      - Extranodal marginal zone B-cell lymphoma
      - Follicular lymphoma unclassified
      - Diffuse large B-cell lymphoma

“real disease entities”

- Many classifications (notably the EORTC and WHO Lymphoid Neoplasms) strive to represent only “real disease entities”
- They reject differences based on
  - incidental manifestation patterns
  - phase of disease progression
- Goal is to group people according to the etiologic and prognostic features, in order to better select therapeutic approach
- SNOMED’s URU criteria allow a broader definition of usefulness
  - Therefore it contains codes for entities that aren’t “real diseases”
  - Subsetting for data entry should take this into account

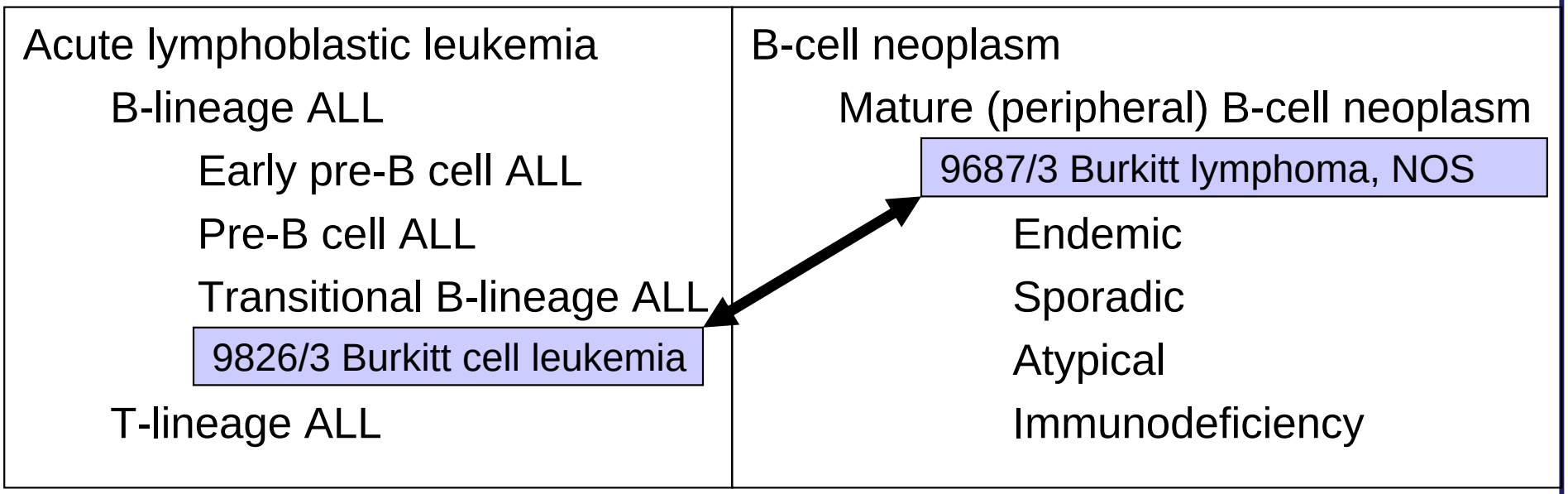
Example: Burkitt's lymphoma/leukemia

Acute lymphoblastic leukemia B-lineage ALL Early pre-B cell ALL Pre-B cell ALL Transitional B-lineage ALL Mature B cell ALL T-lineage ALL	B-cell neoplasm Mature (peripheral) B-cell neoplasm Burkitt lymphoma/leukemia Endemic Sporadic Atypical Immunodeficiency

Example: Burkitt's lymphoma/leukemia

Acute lymphoblastic leukemia B-lineage ALL Early pre-B cell ALL Pre-B cell ALL Transitional B-lineage ALL 9826/3 Burkitt cell leukemia T-lineage ALL	B-cell neoplasm Mature (peripheral) B-cell neoplasm 9687/3 Burkitt lymphoma, NOS Endemic Sporadic Atypical Immunodeficiency

Example: Burkitt's lymphoma/leukemia



Recent WHO Classification published a consensus that Burkitt cell leukemia is just a leukemic phase of the same disease as Burkitt lymphoma

Example: Burkitt's lymphoma/leukemia

Acute lymphoblastic leukemia B-lineage ALL Early pre-B cell ALL Pre-B cell ALL Transitional B-lineage ALL Mature B cell ALL T-lineage ALL	B-cell neoplasm Mature (peripheral) B-cell neoplasm Burkitt lymphoma/leukemia Endemic Sporadic Atypical Immunodeficiency assoc.

SNOMED:

Burkitt leukemia/lymphoma	R-1009B	
Burkitt lymphoma	M-96873	9687/3 Burkitt lymphoma, NOS
Burkitt cell leukemia	M-98263	9826/3 Burkitt cell leukemia
Endemic		
Sporadic		
Atypical		
Immunodeficiency associated		

ICD-O

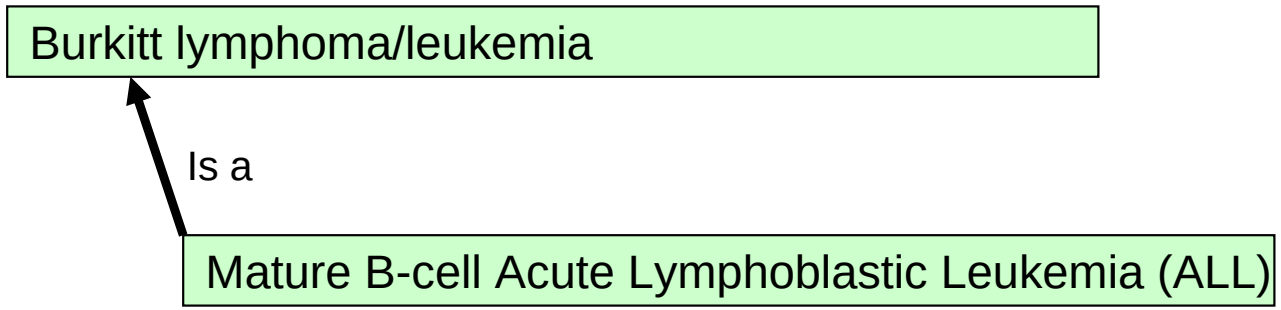
Example: Burkitt's lymphoma/leukemia

Acute lymphoblastic leukemia B-lineage ALL Early pre-B cell ALL Pre-B cell ALL Transitional B-lineage ALL Mature B cell ALL T-lineage ALL	B-cell neoplasm Mature (peripheral) B-cell neoplasm Burkitt lymphoma/leukemia Endemic Sporadic Atypical Immunodeficiency assoc.

Burkitt leukemia/lymphoma	R-1009B	
Burkitt lymphoma	M-96873	9687/3 Burkitt lymphoma, NOS
Burkitt cell leukemia	M-98263	9826/3 Burkitt cell leukemia
Endemic		
Sporadic		
Atypical		
Immunodeficiency associated		

Data re-use and the value of a reference terminology

- A well-defined reference terminology permits the maximum re-use of data collected according to different views
  - This is one of the main purposes for a reference terminology such as SNOMED CT



What does this mean for data collection ?

- Need to define and present to the users a subset of codes/names that are in a coherent “view”
- Ideally, this view should be **current** and **specific** to specialty and context
- Presenting the entire SNOMED hierarchy is almost certain to cause confusion

A little reminder

- There have been 25 lymphoma classifications over the past 75 years
- Stability should not be expected
  - Molecular research is revolutionizing our understanding of these diseases
- But yesterday's data should be re-usable tomorrow (as much as possible)

Dilemmas requiring attention

- The value of secondary data accrues (mainly) to parties other than those who collect it
- The value of secondary data depends on its quality, while the quality of data is directly proportional to the care with which it is collected

Data quality questions needing attention

- Who is responsible for defining professional standards of data quality?
  - Will professional specialty organizations step up to the challenge, as CAP has done?
- What clinical data is essential?
  - Can HHS help coordinate data needs, so clinicians are not overburdened?
- How can support and incentives be provided to clinicians?

Questions?