The ICD-10 Transition… Implications for Pragmatic Trials

NIH Health Care Systems Research Collaboratory Grand Rounds

August 14, 2015

Kin Wah Fung, MD, MS, MA

National Library of Medicine

Rachel Richesson, PhD, MPH

Duke University School of Nursing

Outline

- Pragmatic Trials and phenotypes
- ICD-10 background
- Mapping and tools
  - Preliminary translation of selected phenotype definitions using GEMS
- Implication for research
- Recommendations
- Discussion

Pragmatic Trials

- Studies sampling from and embedded within the context of healthcare delivery systems
- Use electronic health record systems and data
  - Cohort identification, sampling, recruitment
  - Randomization, workflow cues
  - Use of clinical data for study
  - De novo data collection

Clinical Phenotype Definitions

- Specifications for identifying patients or populations with a given characteristic or condition of interest from EHRs using data that are routinely collected in EHRs or ancillary data sources.

- Include widely adopted coding systems
  - ICD-9-CM
  - CPT
  - SNOMED CT
  - LOINC
  - RxNorm
  - NDC

Example phenotype definition

Diabetes defined as[1]:

ICD-9 codes

- one inpatient discharge diagnosis (ICD-9-CM 250.x, 357.2, 366.41, 362.01-362.07)

or any combination of two of the following events occurring within 24 months of each other:

Lab codes

- A1C ≥ 6.5% (48 mmol/mol)
- fasting plasma glucose ≥ 126 mg/dl (7.0 mmol/L)
- random plasma glucose ≥ 200 mg/dl (11.1 mmol/L)
- 2-h 75-g OGTT ≥ 200 mg/dl
- outpatient diagnosis code (same codes as inpatient)
- anti-hyperglycemic medication dispense (see details below)
- NDC in associated list

Medication codes

- **…etc., etc…**

1. Nichols GA, Desai J, Elston Lafata J, et al. Construction of a Multisite DataLink Using Electronic Health Records for the Identification, Surveillance, Prevention, and Management of Diabetes Mellitus: The SUPREME-DM Project. Prev Chronic Dis. 2012;9:110311.

Lots of phenotypes

- >75 phenotype/cohort definitions
  - 32 ICD-9 exclusive

- 30 public (92 private)
  - 79-96% phenotypes use ICD-9
  - Zero ICD-9 exclusive

NIH Collaboratory *Rethinking Clinical Trials®*

Health Care Systems Research Collaboratory

NIH Collaboratory | About Us | Demonstration Projects | Cores | News | Collaboration Spaces | The Living Te:

Demonstration Projects ▸ Home

https://www.nihcollaboratory.org/demonstration-projects/Pages/default.aspx

Demonstration Projects

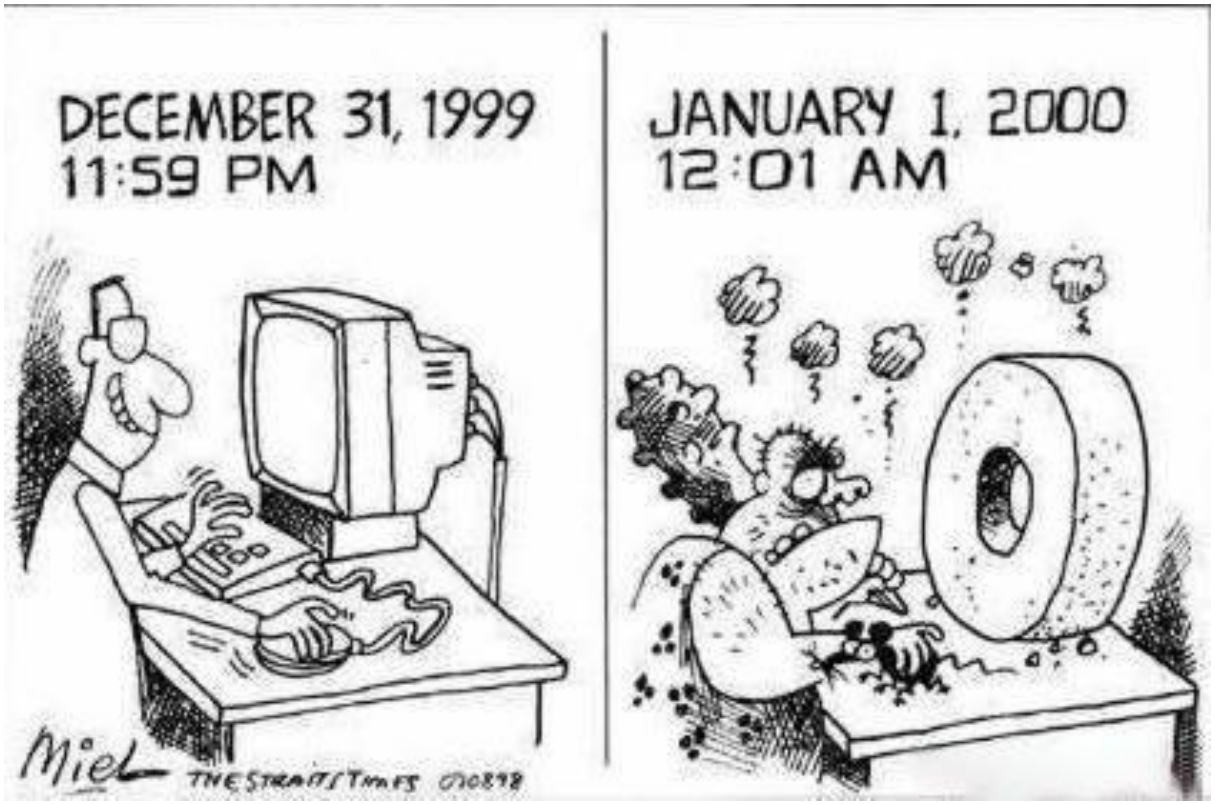
The Research Collaboratory is designed in part to support the design and rapid execution of several Pragmatic Clinical Trial Demonstration research partnerships. The data, tools, and resources produced by the Demonstration Projects will be made available to the greater resear supports the development of exploratory or innovative research activities, and a UH3 award provides support for the second phase of rese

Projects

Title	Investigator	Collaboratory Affiliation	Name
UH3 Project: Time to Reduce Mortality in End-Stage Renal Disease (TiME)	Dember, Laura	University of Pennsylvania	TiME
UH3 Project: Suicide Prevention Outreach Trial (SPOT)	Simon, Gregory	Group Health Cooperative; Group Health Research Institute	SPOT
UH3 Project: Strategies and Opportunities to Stop Colorectal Cancer (STOP CRC)	Coronado, Gloria	Kaiser Foundation Research Institute	STOP CRC
UH3 Project: Lumbar Image Reporting with Epidemiology (LIRE)	Jarvik, Jeffrey	University of Washington	LIRE
UH3 Project: Collaborative Care for Chronic Pain in Primary Care (PPACT)	DeBar, Lynn	Kaiser Foundation	PPACT
UH3 Project: Active Bathing to Eliminate (ABATE) Infection	Huang, Susan	University of California, Irvine	ABATE
UH2 Project: Pragmatic Trial of Video Education in Nursing Homes (PROVEN)	Mor, Vincent; Volandes, Angelo; Mitchell, Susan	Brown University School of Medicine	PROVEN
UH2 Project: Improving Chronic Disease Management with Pieces (ICD-Pieces)	Vazquez, Miguel	UT Southwestern Medical Center	ICD-Pieces
UH2 Project: A Policy-Relevant U.S. Trauma Care System Pragmatic Trial for PTSD and Comorbidity (Trauma Survivors Outcomes and Support [TSOS])	Zatzick, Douglas	University of Washington	TSOS
UH2 Project: A Blood Pressure Medication Timing Study (BPMedTime)	Rosenthal, Gary	University of Iowa	BPMedTime

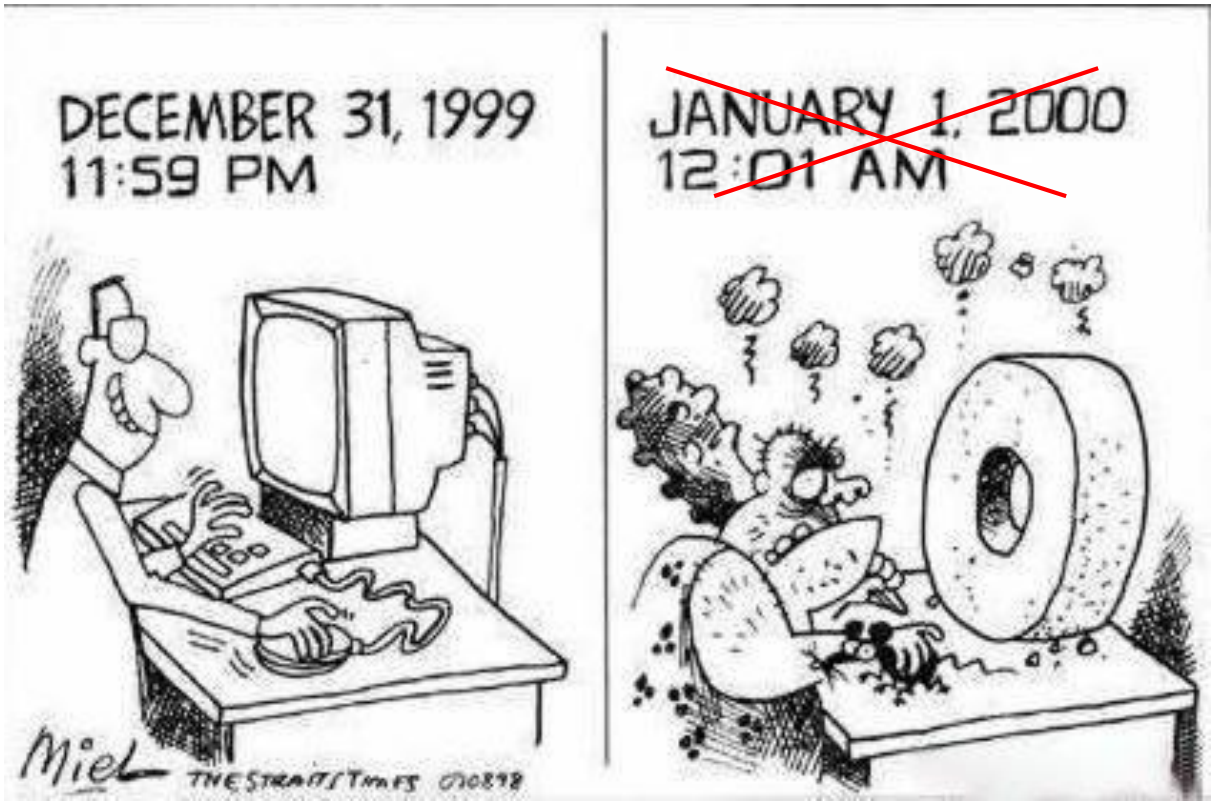
Use Cases for Clinical Phenotypes

- Estimating numbers of patients potentially eligible for a proposed trial (study feasibility).
- Identifying patients for recruitment into prospective trials.
- Describing patient cohorts for analysis of existing data for comparative effectiveness or health services research.
- Presenting baseline characteristics or conditions to describe research populations.
- Presenting primary outcomes to test the trial hypothesis.
- The implementation of supportive tools for providers that are embedded within EHR systems and clinical workflows.



Source: http://blog.ivman.com/y2k-bug-in-retrospect/

October 1, 2015



Source: http://blog.ivman.com/y2k-bug-in-retrospect/

Use of ICD-9-CM in the US

- CM – “clinical modification” based on the international ICD to give more detailed codes
- ICD-9-CM has been used in the US since 1979 (4 years after the international version) for:
  - Classification of morbidity and mortality (mortality reporting changed to ICD-10 since 1999)
  - Reimbursement (since 1983)
  - Analysis of healthcare delivery and cost
  - Epidemiological and clinical research

Long road to change

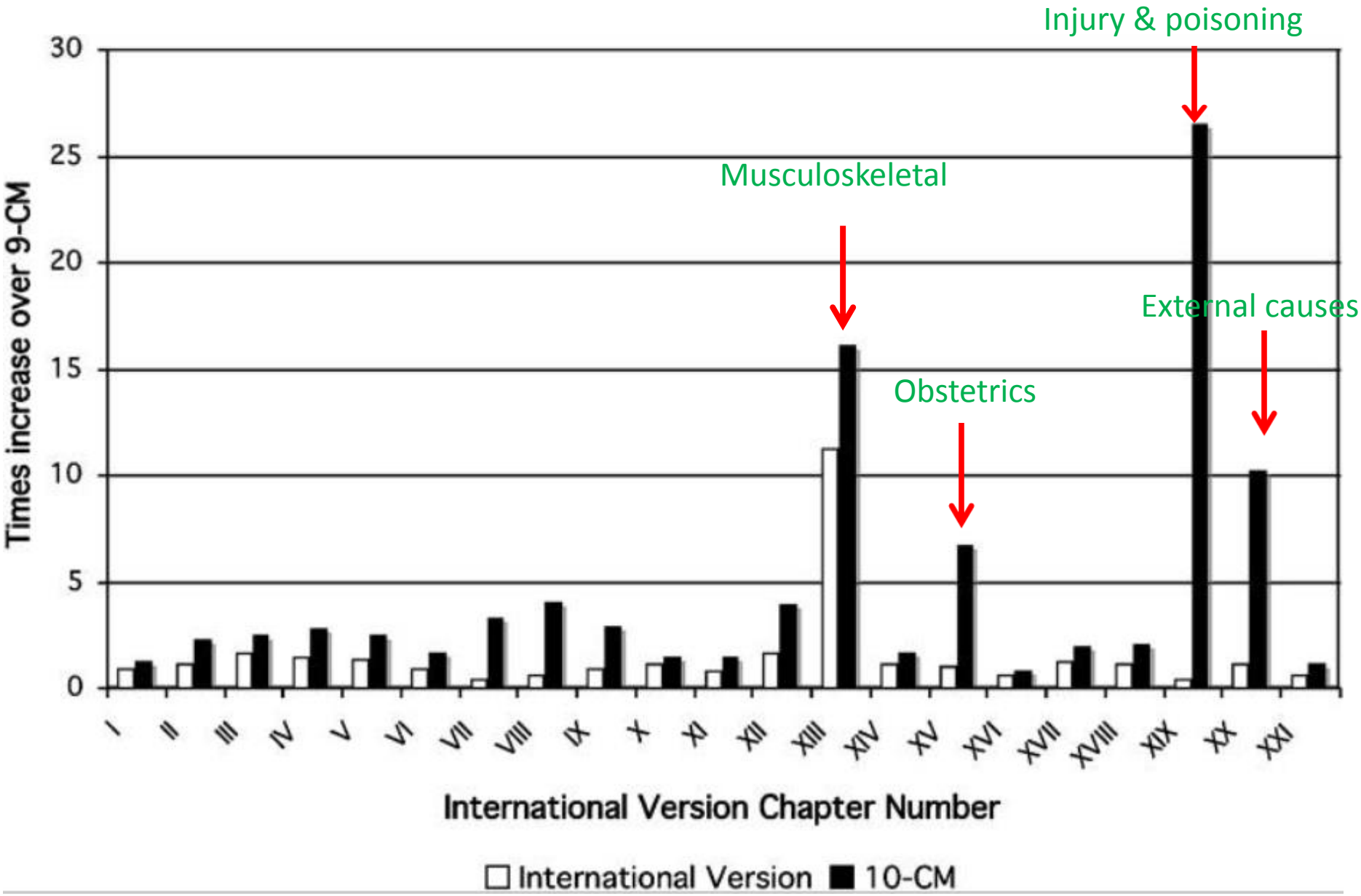
- ICD-9-CM became more and more out-dated, and there is no way to add new codes because of its rigid code structure
- 2008 – CMS issued NPRM proposing 2011 date
- 2009 – CMS final rule with deadline Oct 2013
- 2012 – postponed to Oct 2014
- 2014 – Congress passed law to delay ICD-10-CM for at least one year, CMS set new date to Oct 2015

I9-10 differences - Codes

- Codes look different:
  - Lymphocytopenia
    - ICD-9-CM: 288.51
    - ICD-10-CM: D72.810
  - First digit of an ICD-10-CM code is always alphabetic
    - Some ICD-9-CM codes also start with a letter (E and V codes)
    - No ‘code collision’ – no code is valid in both I9 and I10
  - A valid ICD-10-CM code (leaf code) has between 3 to 7 digits
    - 3-digit (< 1%)
    - 4-digit (8%)
    - 5-digit (9%)
    - 6-digit (13%)
    - 7-digit (70%)

I9-10 differences - Size

- Total number of valid codes:
  - ICD-9-CM: 14,567
  - ICD-10-CM: 69,823
- The jump in size is not uniform across chapters



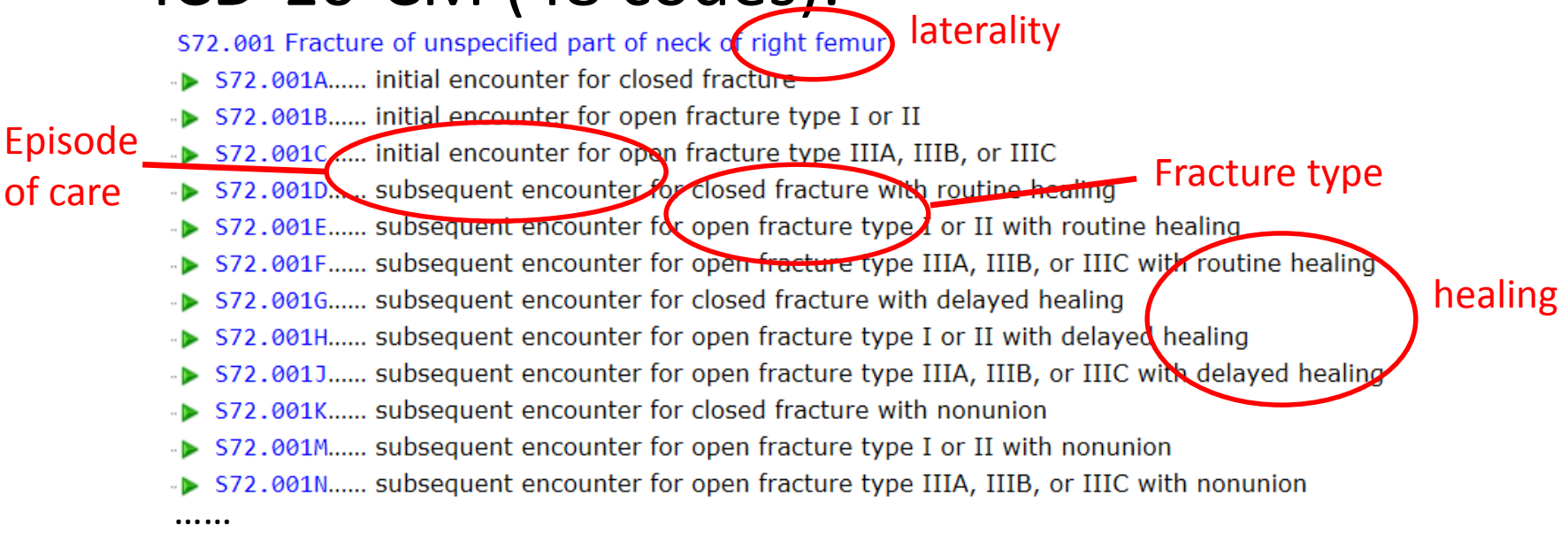
Steindel S. International classification of diseases, 10th edition, clinical modification and procedure coding system: descriptive overview of the next generation HIPAA code sets. J Am Med Inform Assoc. 2010 May-Jun;17(3):274-82.

Reasons for increase in size

- New codes for
  - New diseases
  - Uncommon diseases
  - Subtypes of diseases
- Additional details (combinatorial explosion) e.g. Fractures:
  - Laterality: left, right, unspecified, bilateral
  - Episode of care: initial encounter, subsequent encounter, sequela
  - Type of fracture: closed, open (Gustilo classification type I, II, IIIA, IIIB, IIIC)
  - Healing status: routine, delayed, nonunion, malunion

Fracture neck of femur

- ICD-9-CM:
  - 820.8 Unspecified part of neck of femur, closed
  - 820.9 Unspecified part of neck of femur, open
- ICD-10-CM (48 codes):



Chorioamnionitis

- ICD-9-CM: 762.7 Chorioamnionitis
- ICD-10-CM: (28 codes)

041.121 Chorioamnionitis, first trimester
- 041.1210 ..... not applicable or unspecified
- 041.1211 ..... fetus 1
- 041.1212 ..... fetus 2
- 041.1213 ..... fetus 3
- 041.1214 ..... fetus 4
- 041.1215 ..... fetus 5
- 041.1219 ..... other fetus

041.122 Chorioamnionitis, second trimester
- 041.1220 ..... not applicable or unspecified
- 041.1221 ..... fetus 1
- 041.1222 ..... fetus 2

.....

I9-10 differences: Organization

- Chapter structure largely preserved
- Sense organs separated from nervous system disorders, creating 2 new chapters:
  - Eye and Adnexa
  - Ear and Mastoid Process
- Major reorganization of some chapters e.g.
  - Mental and Behavioral Disorders
  - Diseases of the Skin and Subcutaneous Tissues
- Some diseases moved chapters e.g.
  - Gout moved from Endocrine, Nutritional and Metabolic Diseases to Musculoskeletal Diseases

I9-10 differences: Semantic

- More subtle changes e.g.,
  - Acute myocardial infarction
    - ICD-9-CM: within 8 weeks of onset
    - ICD-10-CM: within 4 weeks of onset
  - Cutoff for abortion vs. fetal death
    - ICD-9-CM: 22 weeks
    - ICD-10-CM: 20 weeks
  - Tuberculosis
    - Method of diagnosis (bacteriological or histological) no longer specified
  - Diabetes
    - No longer distinguished as controlled or uncontrolled
  - Asthma
    - No longer classified as intrinsic or extrinsic

I9-10 code sets transition

- Cohort definitions coded in ICD-9-CM will have to be transitioned to ICD-10-CM
- Resources to ease the burden
  - General Equivalence Maps (GEM)
  - Quality measure value sets
  - SNOMED CT

General Equivalence Maps

- Published by CMS and CDC
- Provide linkages between
  - ICD-9-CM and ICD-10-CM
  - ICD-9-CM volume III (procedures) and ICD-10-PCS
- Forward (9 to 10) and backward (10 to 9) maps
  - Independent maps, not mirror images
  - Different coverage of ICD-9-CM and ICD-10-CM
  - Partial overlap in the mappings

	Forward GEM	Backward GEM	Common to both GEMs
Unique ICD-9-CM codes* (% of ICD-9-CM)	13,409 (92.0%)	10,949 (75.0%)	10,880 (74.7%)
Unique ICD-10-CM codes* (% of ICD-10-CM)	16,614 (23.8%)	69,154 (99.0%)	16,614 (23.8%)
Unique ICD-9-CM/ICD-10-CM code pairs	23,330	78,034	18,484

* *Not including codes with no maps*

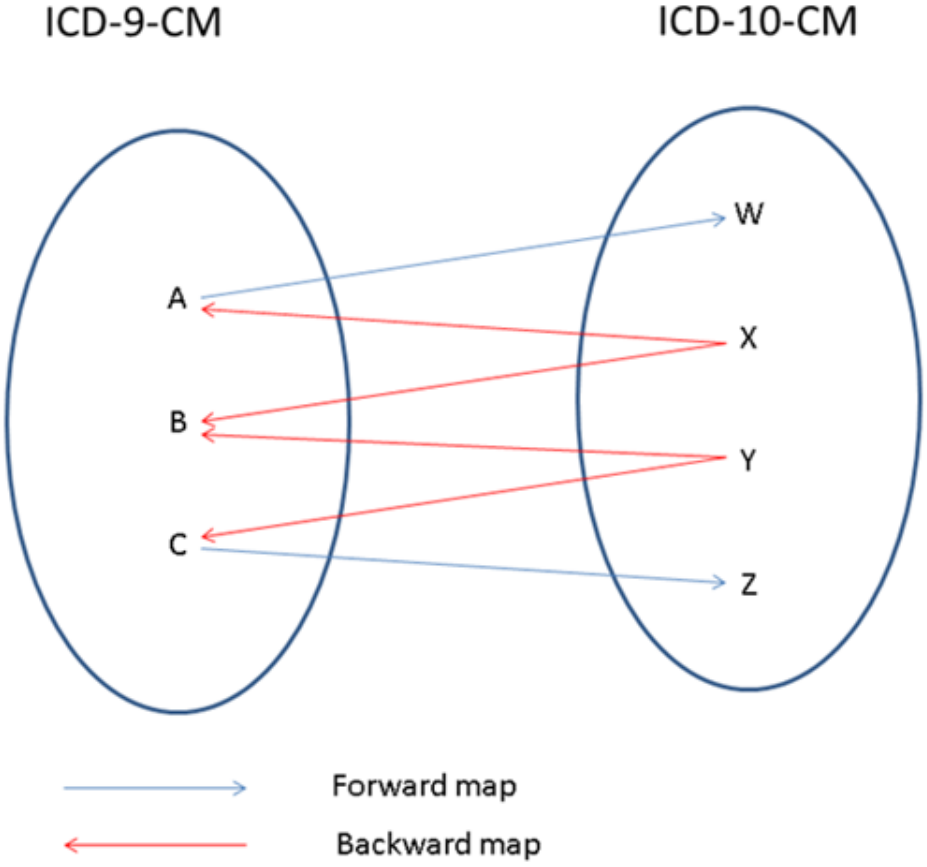
Comparison of forward and backward GEMs

Using the GEMs in code set translation

- Study using 32 ICD-9-CM code sets (clinical phenotypes) from 3 pragmatic trials:
  - Collaborative Care for Chronic Pain in Primary Care (PPACT)
  - Strategies and Opportunities to Stop Colorectal Cancer in Priority Populations (STOP CRC)
  - A Pragmatic Trial of Population-Based Programs to Prevent Suicide Attempt
- Code sets with 3 – 161 (median=4) ICD-9-CM codes, altogether 536 unique codes
- Compared 4 mapping methods using the GEMs

The mapping methods

- 4 progressively more aggressive methods to identify ICD-10-CM targets for an ICD-9-CM code:
  1. Simple forward map – forward GEM only
  2. Forward backward map – 1. + backward GEM
  3. Secondary map – 2. + map targets identified by secondary ICD-9-CM codes
  4. Tertiary map – 3. + map targets identified by tertiary ICD-9-CM codes



Targets identified for each source code by the 4 mapping methods

Source code	SFM	FBM	SM	TM
A	W	W, X	W, X, Y	W, X, Y, Z
B	-	X, Y	W, X, Y, Z	W, X, Y, Z
C	Z	Y, Z	X, Y, Z	W, X, Y, Z

Study methodology

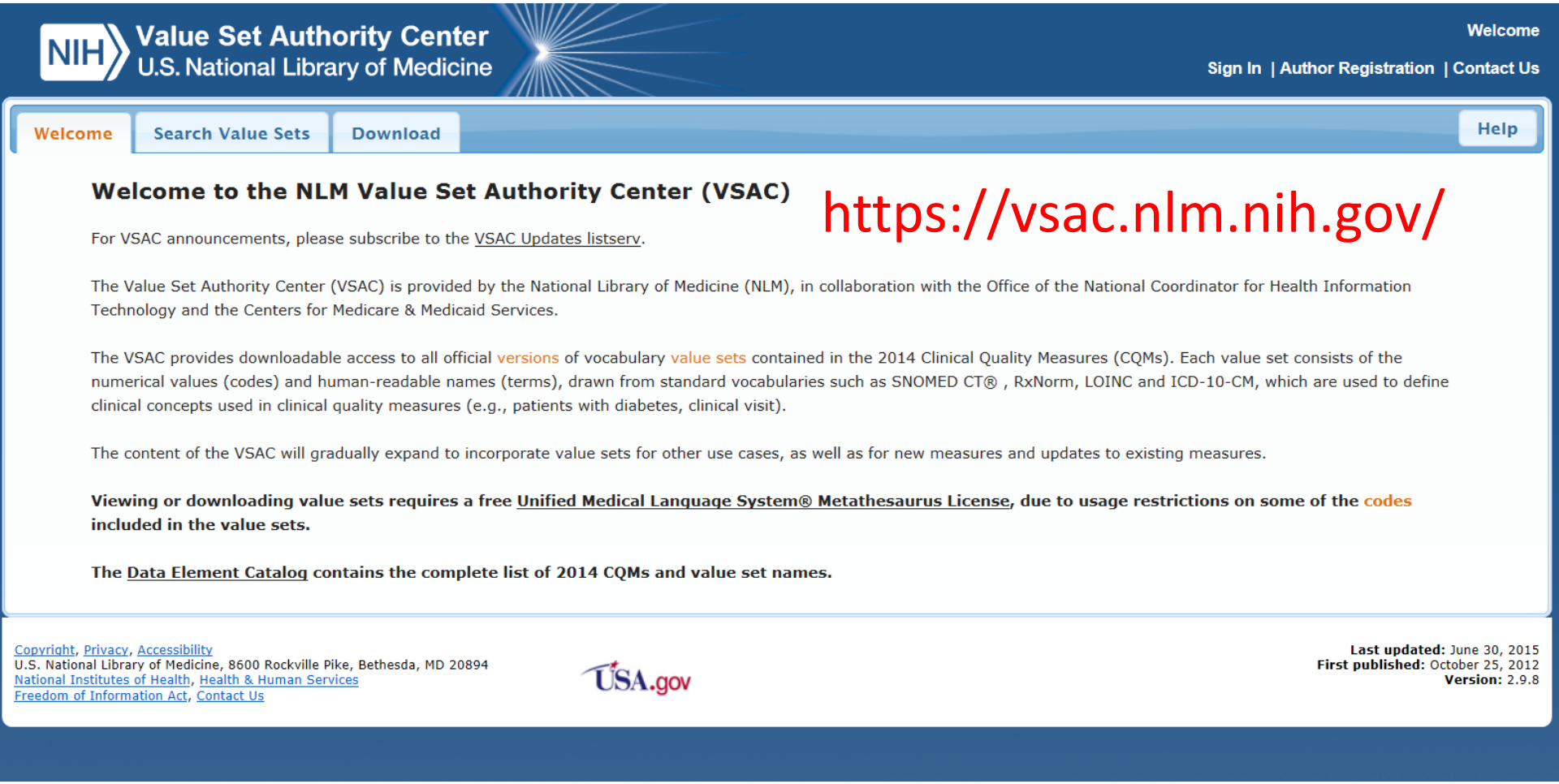
- Generate ICD-10-CM code sets for each of the 32 ICD-9-CM code sets using each mapping method
- Generated code sets reviewed by clinical experts for validity
- Recall, precision and F-score for each mapping method

Summary of results

- Must use both forward and backward GEMs
- More aggressive methods can identify valid ICD-10-CM targets that are indirectly related to an ICD-9-CM code, but precision is reduced. Choice of method will depend on use case.
- Works better for well-defined conditions (e.g. colorectal cancer) than vaguely-defined conditions (e.g. chronic pain)
- Not fully-automated translation – manual validation still required

Electronic clinical quality measurement

- Meaningful Use requires EHRs to demonstrate electronic submission of data for some clinical quality measures
- Quality measure value sets:
  - Code sets from standard terminologies used to identify patients with certain characteristics
  - Very similar in function to cohort definition code sets in clinical studies
  - Available from NLM's VSAC website


Value Set Authority Center
U.S. National Library of Medicine
Welcome
Sign In | Author Registration | Contact Us
Welcome
Search Value Sets
Download
Help
Welcome to the NLM Value Set Authority Center (VSAC)
https://vsac.nlm.nih.gov/
For VSAC announcements, please subscribe to the VSAC Updates listserv.
The Value Set Authority Center (VSAC) is provided by the National Library of Medicine (NLM), in collaboration with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.
The VSAC provides downloadable access to all official versions of vocabulary value sets contained in the 2014 Clinical Quality Measures (CQMs). Each value set consists of the numerical values (codes) and human-readable names (terms), drawn from standard vocabularies such as SNOMED CT® , RxNorm, LOINC and ICD-10-CM, which are used to define clinical concepts used in clinical quality measures (e.g., patients with diabetes, clinical visit).
The content of the VSAC will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.
Viewing or downloading value sets requires a free Unified Medical Language System® Metathesaurus License, due to usage restrictions on some of the codes included in the value sets.
The Data Element Catalog contains the complete list of 2014 CQMs and value set names.
Copyright, Privacy, Accessibility
U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894
National Institutes of Health, Health & Human Services
Freedom of Information Act, Contact Us
USA.gov
Last updated: June 30, 2015
First published: October 25, 2012
Version: 2.9.8

Apply Filters | Clear Filters

Narrow search results by selecting from pull-down menus below:

CMS eMeasure (NQF Number)
Select

Quality Data Model Category
Select

Steward
Select

Meaningful Use Measures
Select

Code System
Select

Search the NLM Value Set Repository

Query: diabetes | Search

Search Results | Value Set Details

Export Search Results (Excel)

Matched Value Sets

Download | View | Toggle | Clear | Page 2 of 6 | 20 | View 21 - 40 of 116

Name	Type	Code System	Steward	OID
Complications of Pregnancy, Childbirth and the Pue	Extensional	SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.111.11.1023
Complications of Pregnancy, Childbirth and the Pue	Extensional	ICD10CM	NCQA	2.16.840.1.113883.3.464.1003.111.11.1022
Conditions Possibly Justifying Elective Delivery Pri	Grouping	ICD10CM ICD9CM SNOMEDCT	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.286
Conditions Possibly Justifying Elective Delivery Pri	Extensional	ICD9CM	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.394
Conditions Possibly Justifying Elective Delivery Pri	Extensional	ICD10CM	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.393
Conditions Possibly Justifying Elective Delivery Pri	Extensional	SNOMEDCT	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.395
Degeneration of Macula and Posterior Pole	Grouping	ICD10CM ICD9CM SNOMEDCT	AMA-PCPI	2.16.840.1.113883.3.526.3.1453
Degeneration of Macula and Posterior Pole	Extensional	SNOMEDCT	AMA-PCPI	2.16.840.1.113883.3.526.2.1643
Delivery	Extensional	ICD10CM	Optum	2.16.840.1.113762.1.4.1078.3
Delivery - Diagnosis	Grouping	ICD10CM ICD9CM SNOMEDCT	Optum	2.16.840.1.113883.3.67.1.101.1.278
Delivery ICD9Dx	Extensional	ICD9CM	Optum	2.16.840.1.113883.3.67.1.101.1.83
Diabetes	Extensional	ICD10CM	NCQA	2.16.840.1.113883.3.464.1003.103.11.1002
Diabetes	Grouping	ICD10CM ICD9CM SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.103.12.1001
Diabetes	Extensional	ICD9CM	NCQA	2.16.840.1.113883.3.464.1003.103.11.1001
Diabetes	Extensional	SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.103.11.1003
Diabetes Visit		CPT		

Apply Filters | Clear Filters

Narrow search results by selecting from pull-down menus below:

CMS eMeasure (NQF Number)
Select

Quality Data Model Category
Select

Steward
Select

Meaningful Use Measures
Select

Code System
Select

Search the NLM Value Set Repository

Query: diabetes | Search

Search Results | Value Set Details

Export Search Results (Excel)

Matched Value Sets

Download | View | Toggle | Clear | Page 2 of 6 | 20 | View 21 - 40 of 116

Name	Type	Code System	Steward	OID
Complications of Pregnancy, Childbirth and the Pue	Extensional	SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.111.11.1023
Complications of Pregnancy, Childbirth and the Pue	Extensional	ICD10CM	NCQA	2.16.840.1.113883.3.464.1003.111.11.1022
Conditions Possibly Justifying Elective Delivery Pri	Grouping	ICD10CM ICD9CM SNOMEDCT	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.286
Conditions Possibly Justifying Elective Delivery Pri	Extensional	ICD9CM	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.394
Conditions Possibly Justifying Elective Delivery Pri	Extensional	ICD10CM	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.393
Conditions Possibly Justifying Elective Delivery Pri	Extensional	SNOMEDCT	The Joint Commissior	2.16.840.1.113883.3.117.1.7.1.395
Degeneration of Macula and Posterior Pole	Grouping	ICD10CM ICD9CM SNOMEDCT	AMA-PCPI	2.16.840.1.113883.3.526.3.1453
Degeneration of Macula and Posterior Pole	Extensional	SNOMEDCT	AMA-PCPI	2.16.840.1.113883.3.526.2.1643
Delivery	Extensional	ICD10CM	Optum	2.16.840.1.113762.1.4.1078.3
Delivery - Diagnosis	Grouping	ICD10CM ICD9CM SNOMEDCT	Optum	2.16.840.1.113883.3.67.1.101.1.278
Delivery ICD9Dx	Extensional	ICD9CM	Optum	2.16.840.1.113883.3.67.1.101.1.83
Diabetes	Extensional	ICD10CM	NCQA	2.16.840.1.113883.3.464.1003.103.11.1002
Diabetes	Grouping	ICD10CM ICD9CM SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.103.12.1001
Diabetes	Extensional	ICD9CM	NCQA	2.16.840.1.113883.3.464.1003.103.11.1001
Diabetes	Extensional	SNOMEDCT	NCQA	2.16.840.1.113883.3.464.1003.103.11.1003
[illegible]		CPT		

Re-using quality measure value sets

- Many value sets are already defined in multiple terminologies
  - 267 value sets with ICD-9-CM code sets
  - 259 (97%) also have ICD-10-CM code sets
  - 253 (95%) also have ICD-10-CM and SNOMED CT code sets
- Some value sets are exact matches of cohort definitions (e.g., “malignant neoplasm of colon” value set and “colon cancer” phenotype code set)
- Finding matching value sets
  - May be difficult to browse through 800+ value sets
  - Can compute some similarity score between the ICD-9-CM code sets for phenotype definition and quality measurement e.g., Jaccard similarity coefficient = size of intersection / size of union

SNOMED CT

- Most comprehensive, multilingual clinical terminology in the world
- Used in > 50 countries
- Meaningful Use requires use of SNOMED CT in the EHR for problem lists, procedures etc.
- SNOMED CT is better than ICD for clinical data capture because:
  - Better content coverage
  - Clinically oriented
  - Flexible data entry and retrieval

	SNOMED CT	ICD-9-CM	ICD-10-CM
Congenital skin anomalies	205573006 Focal dermal hypoplasia 79468000 Familial benign pemphigus 5132005 Keratosis pilaris *… (total 21 codes)*	757.39 Other specified congenital anomalies of skin	Q82.8 Other specified congenital malformations of skin
Acidosis	59455009 Metabolic acidosis 12326000 Respiratory acidosis 91273001 Lactic acidosis *… (total 60 codes)*	276.2 Acidosis	E87.2 Acidosis
Brachial plexus disorders	72893007 Brachial neuritis 278065000 Pancoast's syndrome 78141002 Erb-Duchenne paralysis *... (total 33 codes)*	353.0 Brachial plexus lesions	G54.0 Brachial plexus disorders

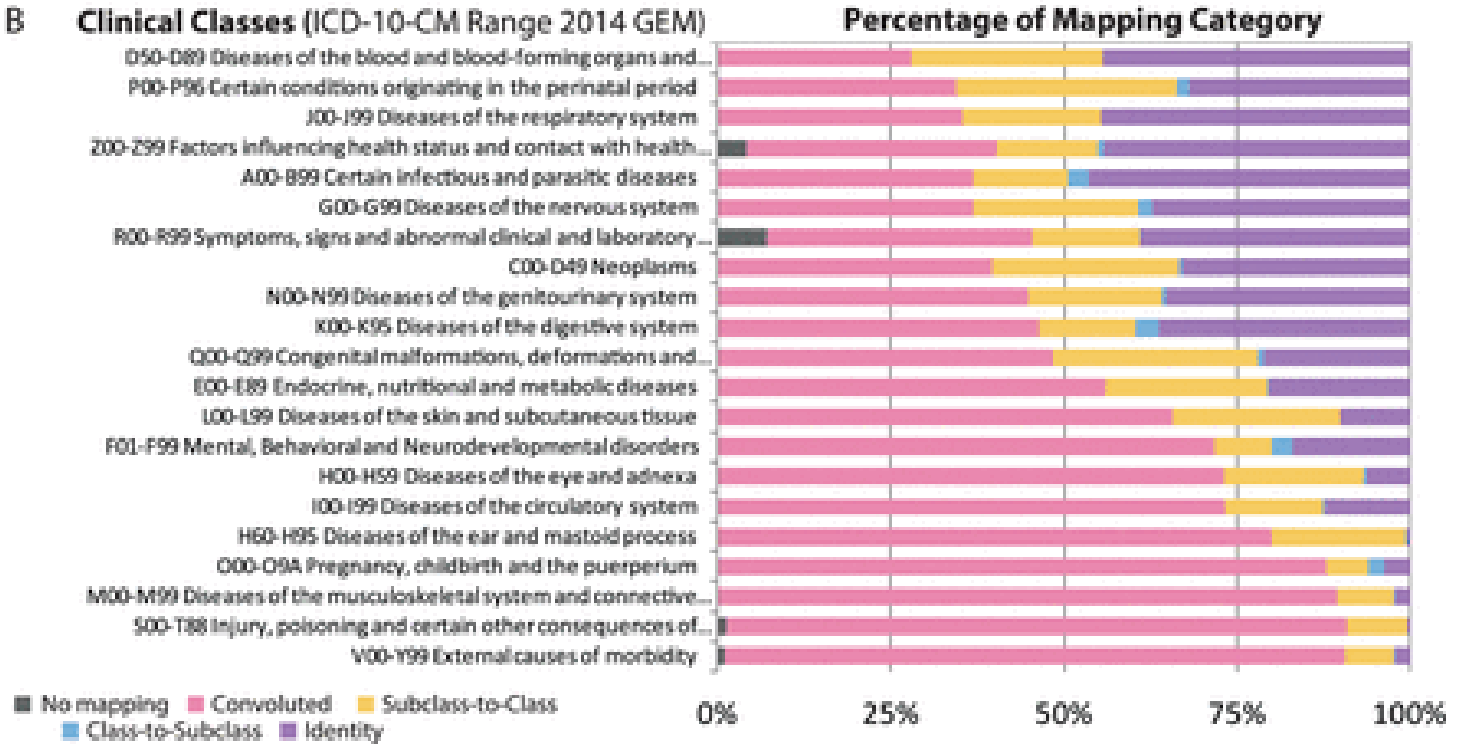
Role of SNOMED CT

- Can help to map from ICD-9-CM to ICD-10-CM
  - 2 maps available:
    - SNOMED CT to ICD-9-CM (IHTSDO)
    - SNOMED CT to ICD-10-CM (NLM)
  - Possible to do sequential mapping from ICD-9-CM to ICD-10-CM through SNOMED CT
- Use SNOMED CT directly in cohort definitions
  - SNOMED CT codes will become more ubiquitous in EHR
  - More granular concepts → fine-tuning of definitions
  - Many quality measure value sets are already defined in SNOMED CT

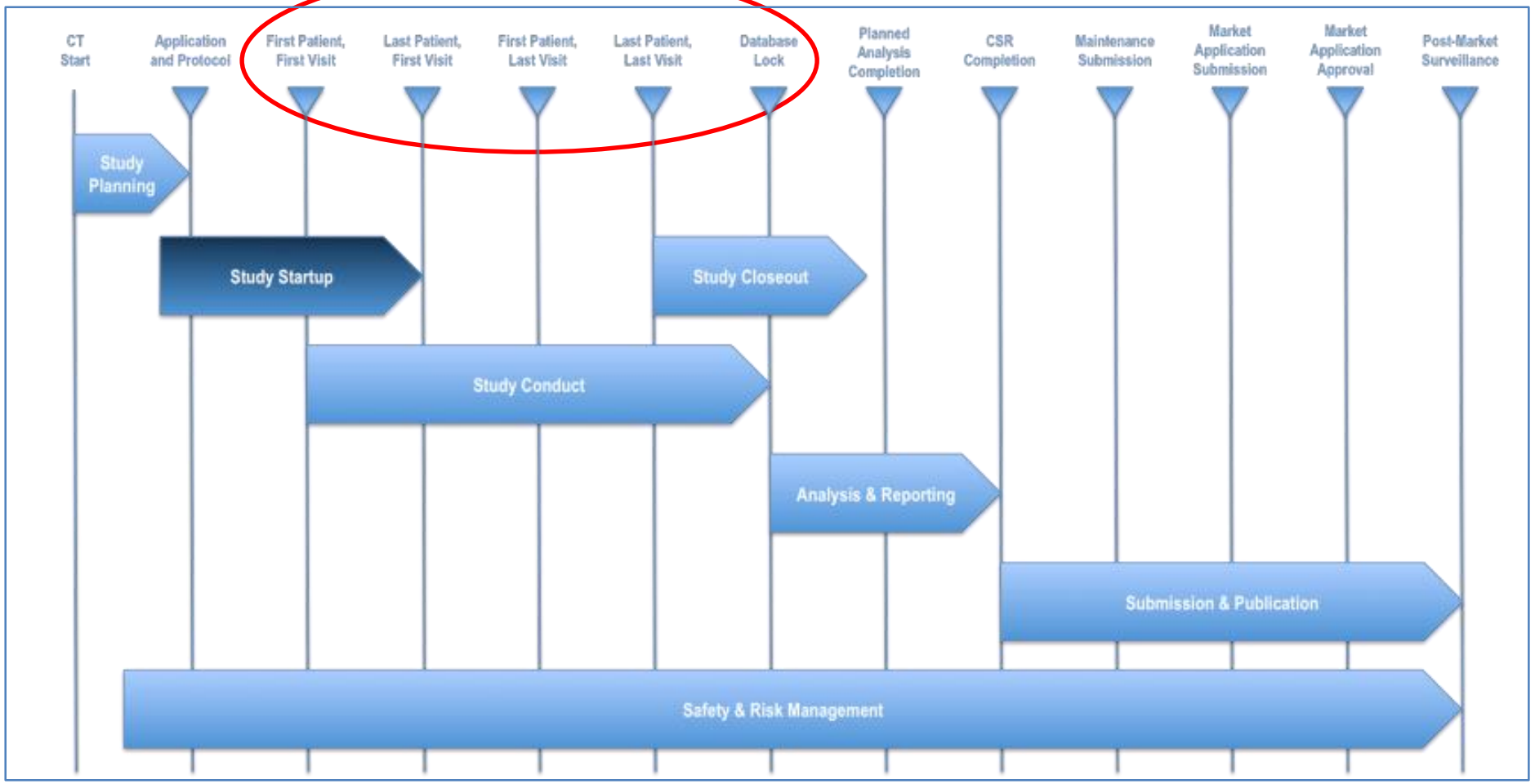
Implications for Pragmatic Trials

- The same system, organizational, and cultural changes that drive variation of ICD-9 coding will also impact ICD-10 coding
- There are various tools and approaches to mapping between ICD-9 and ICD-10
- There can be variation by organization and system on how these maps are used
- More problematic in different medical specialties
- Many “convoluted” relationships (Boyd et al., 2015)

Andrew Boyd et al. "Metrics and tools for consistent cohort discovery and financial analyses post-transition to ICD-10-CM." JAMIA 2015.

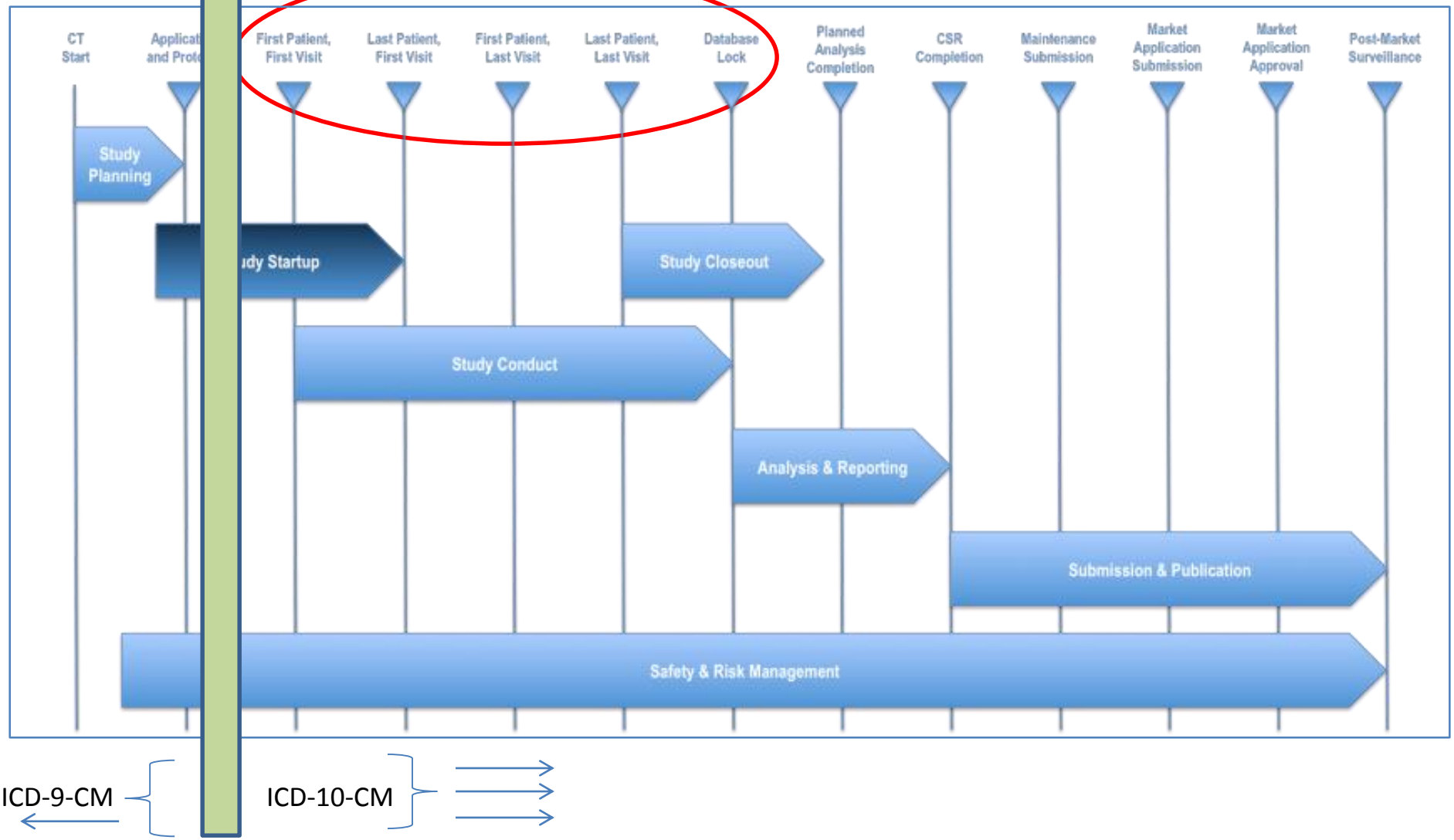


planning *data collection* *analysis & dissemination*

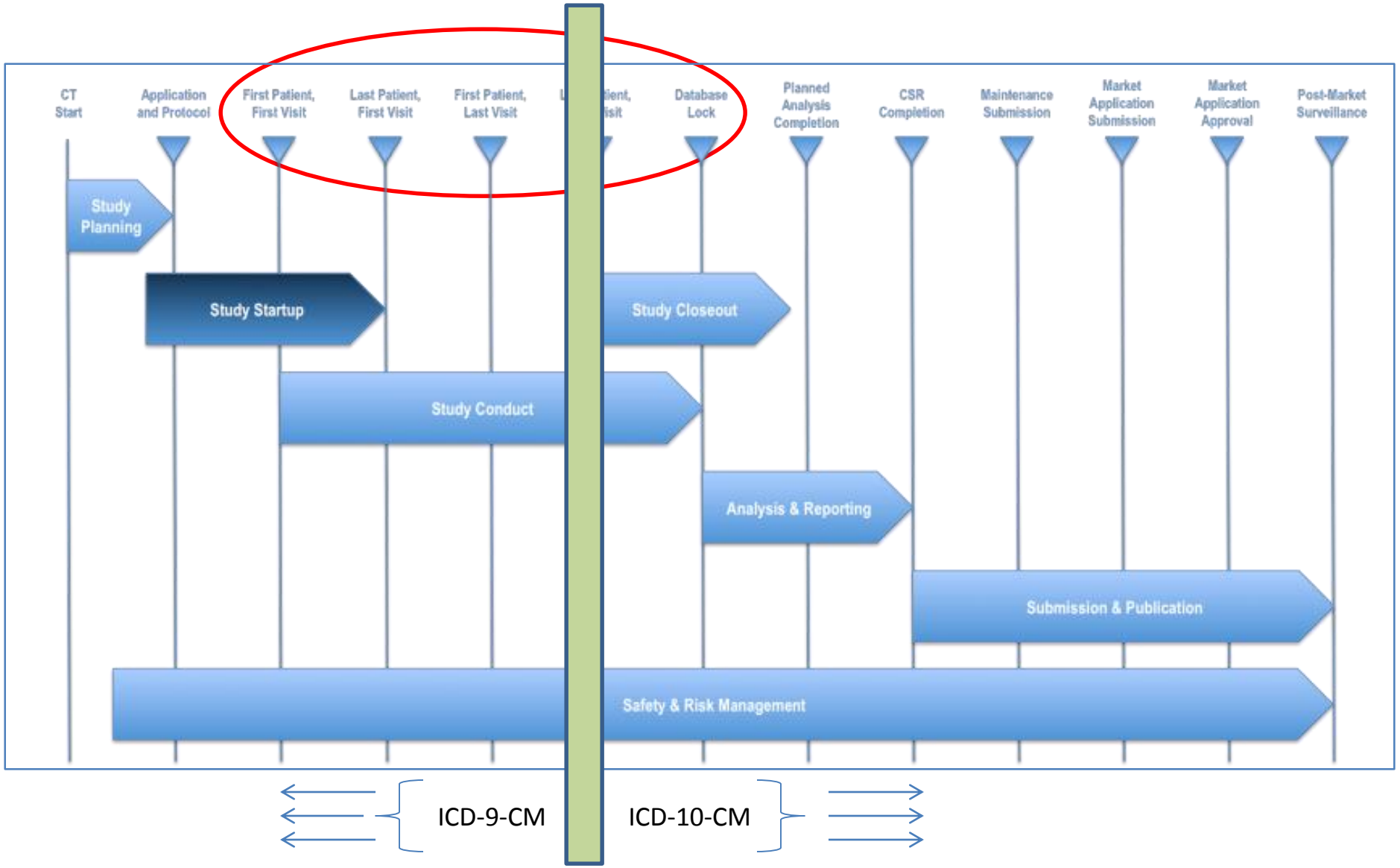


Graphic from: Lagrangian Points Blog: http://lagrangianpoints.com/wp-content/uploads/ct1.png

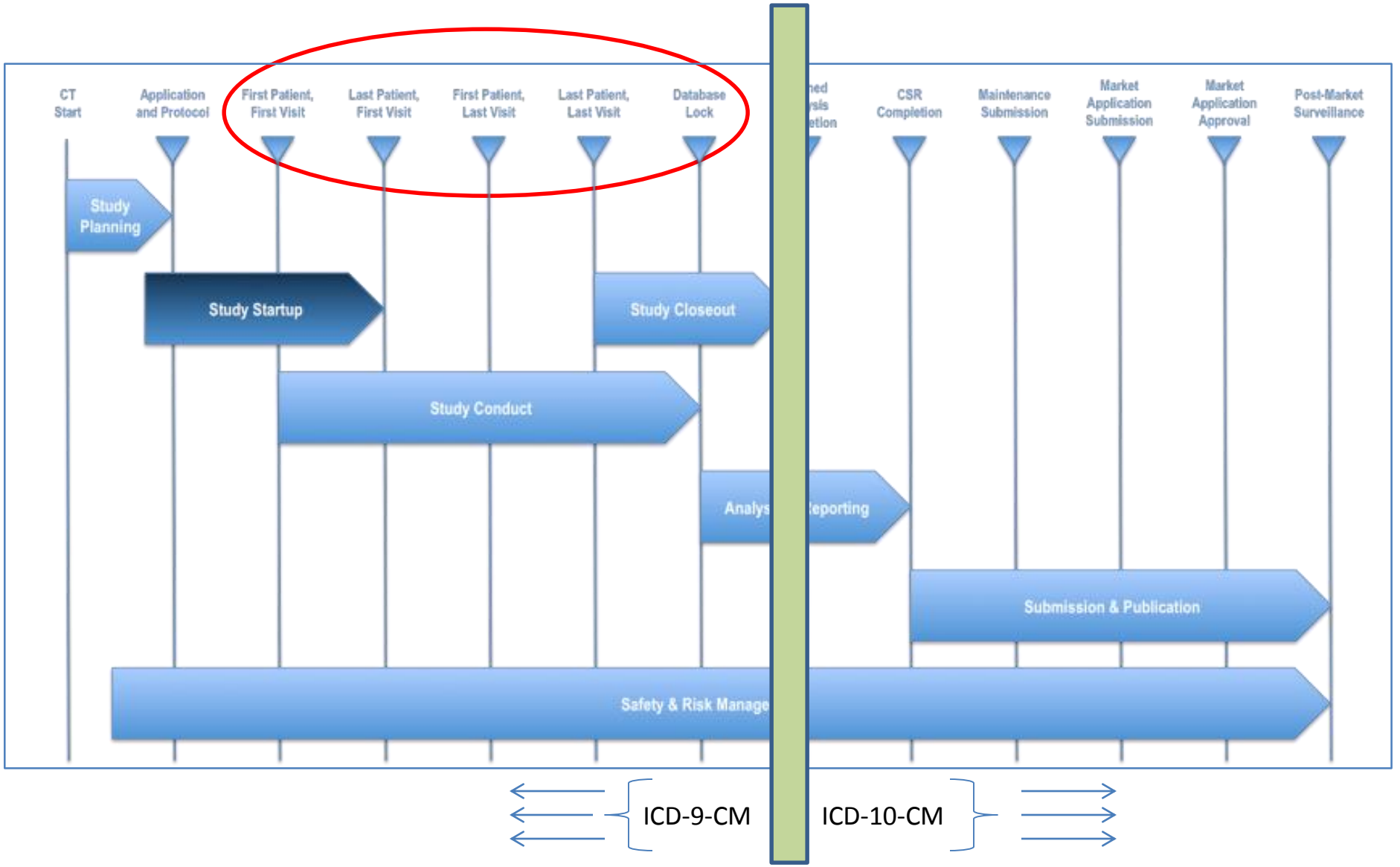
planning *data collection* *analysis & dissemination*



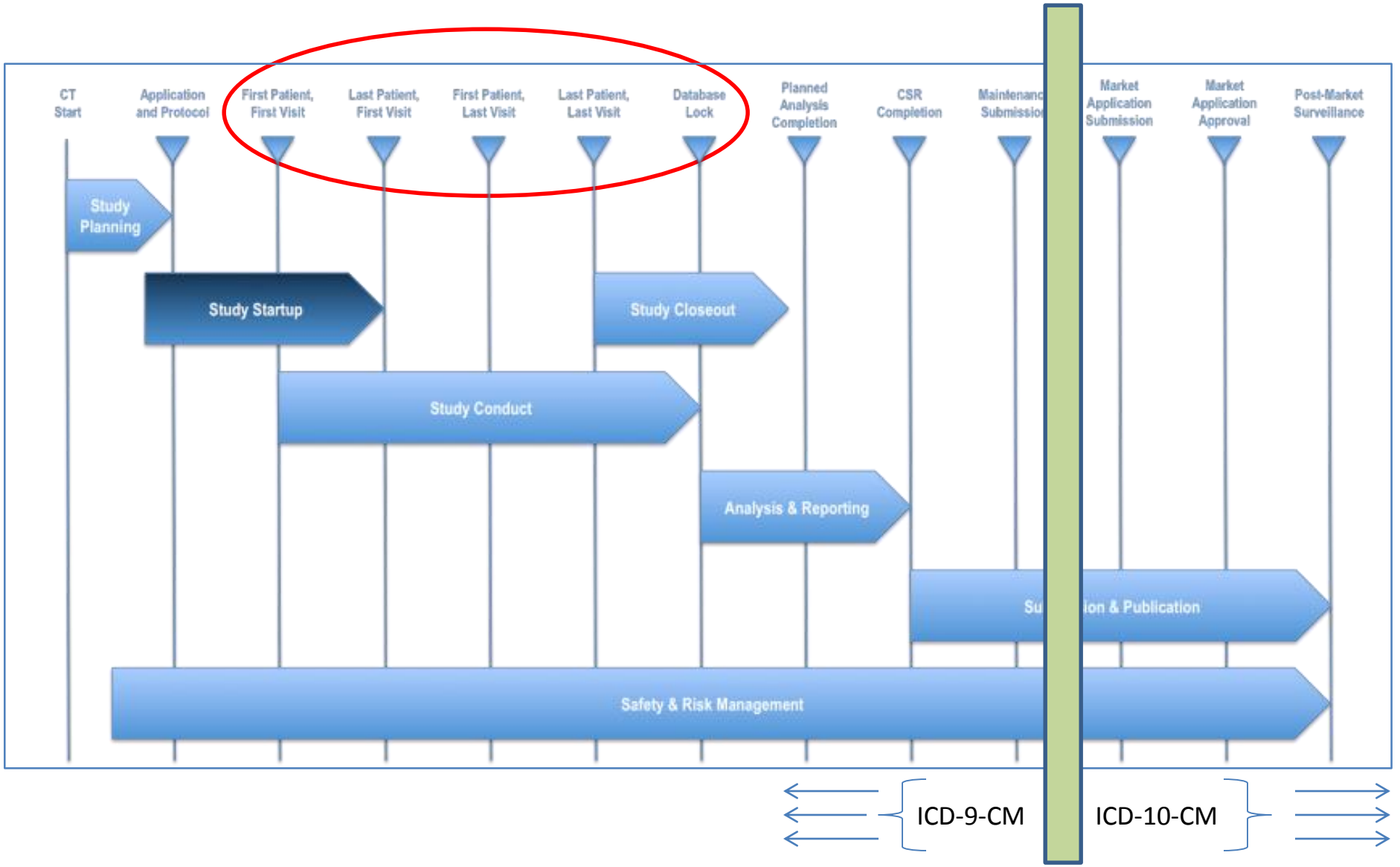
planning *data collection* *analysis & dissemination*



planning *data collection* *analysis & dissemination*



planning *data collection* *analysis & dissemination*



Timing of ICD-10 Transition Relative to Pragmatic Trials

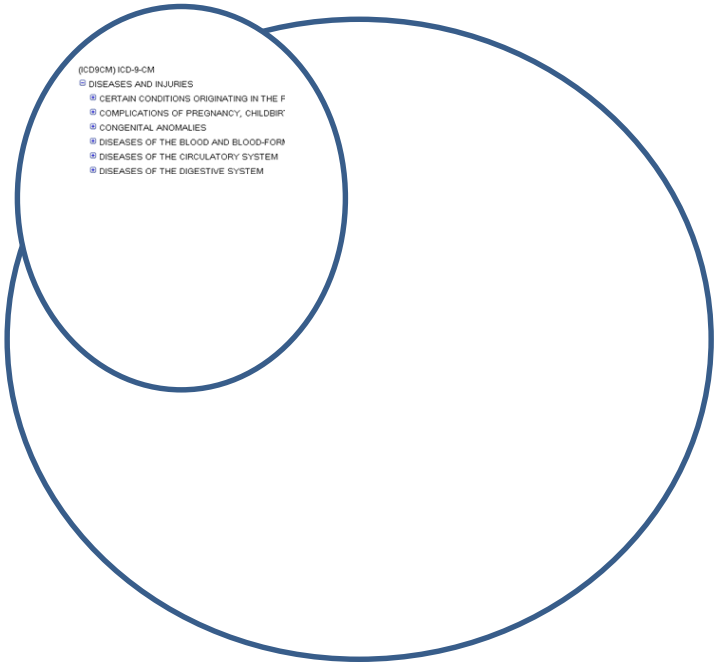
- Before trial data collection (i.e., study begins after Oct 1, 2015)
  - Can use ICD-10, but cannot re-use past tools
    i.e., Must build (and validate) new ICD-10 queries based on ICD-9
  - Historical data in ICD-9 (medical history) might be problematic

- After trial begins (i.e., study began before Oct 1, 2015)
  - ICD-10-based definitions might change the characteristics of the study population (sampling bias, ascertainment bias) or the depth/accuracy of data collection (measurement bias)

- In both cases, researchers might have data in both ICD-9 and ICD-10
- To compare, need to pick one coding system (ICD-9 or 10) or a reference standard (e.g., SNOMED CT)

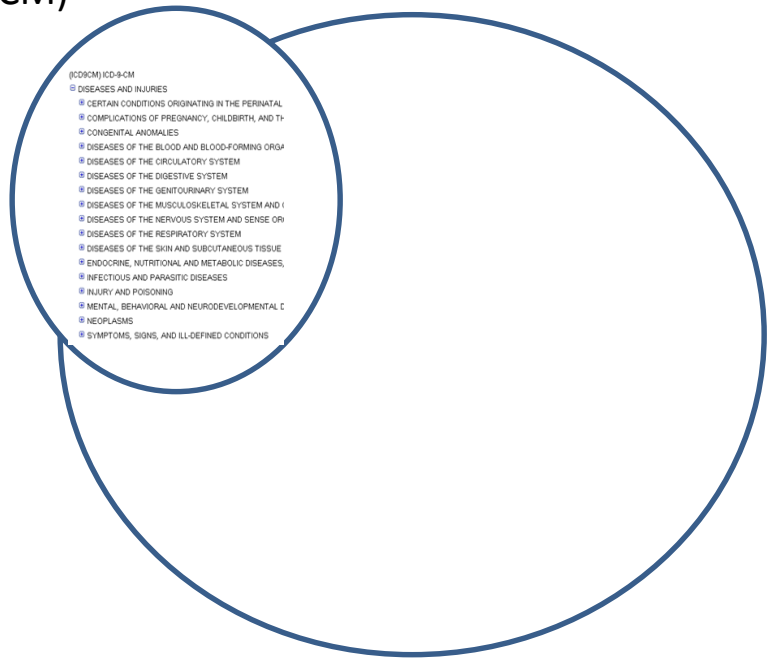
The Ultimate Challenge: Assessing the Semantic *Equivalence* of Phenotype Definitions

Phenotype definition (ICD-9-CM)



"true" population with condition

Phenotype definition (ICD-10-CM)



"true" population with condition

NIH Collaboratory *Rethinking Clinical Trials™*
Health Care Systems Research Collaboratory

Assessing Data Quality for Healthcare Systems Data Used in Clinical Research (Version 1.0)

An NIH Health Care Systems Research Collaboratory Phenotypes, Data Standards, and Data Quality Core White Paper

Meredith N. Zozus, PhD[1]; W. Ed Hammond, PhD[1]; Beverly B. Green, MD, MPH[2]; Michael G. Kahn, MD, PhD[3]; Rachel L. Richesson, PhD, MPH[4]; Shelley A. Rusincovitch[5]; Gregory E. Simon, MD, MPH[2]; Michelle M. Smerek[5]

[1]Duke Translational Medicine Institute, Duke University School of Medicine, Durham, North Carolina; [2]Group Health Research Institute, Seattle, Washington; [3]University of Colorado, Denver, Denver, Colorado; [4]Duke University School of Nursing, Durham, North Carolina; [5]Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina

This work was supported by a cooperative agreement (U54 AT007748) from the NIH Common Fund for the NIH Health Care Systems Research Collaboratory. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Address for correspondence:
Meredith N. Zozus, PhD
Associate Director for Clinical Research Informatics
Duke Translational Medicine Institute
DUMC Box 3850
Durham, NC 27710
Tel: (919) 668-8339
meredith.nahm@dm.duke.edu

- Completeness
- Accuracy
- Consistency

EDM Forum
EDM Forum Community

eGEMs (Generating Evidence & Methods to improve patient outcomes) | EDM Forum Products and Events

3-23-2015

Transparent Reporting of Data Quality in Distributed Data Networks

Michael G. Kahn
University of Colorado, michael.kahn@ucdenver.edu

Jeffrey S. Brown
Harvard Pilgrim Health Care Institute / Harvard Medical School, jeff_brown@harvardpilgrim.org

Alein T. Chun
Cedars-Sinai Health System, alein.chun@cshs.org

Bruce N. Davidson
Hoag Memorial Hospital Presbyterian, brucendavidson@icloud.com

See next pages for additional authors

Follow this and additional works at: http://repository.academyhealth.org/egems

Part of the Health Information Technology Commons, and the Health Services Research Commons

Recommended Citation
Kahn, Michael G.; Brown, Jeffrey S.; Chun, Alein T.; Davidson, Bruce N.; Meeker, Daniella; Ryan, Patrick B.; Schilling, Lisa M.; Weiskopf, Nicole G.; Williams, Andrew E.; and Zozus, Meredith Nahm (2015) "Transparent Reporting of Data Quality in Distributed Data Networks," *eGEMs (Generating Evidence & Methods to improve patient outcomes)*: Vol. 3: Iss. 1, Article 7.
DOI: http://dx.doi.org/10.13063/2327-9214.1052
Available at: http://repository.academyhealth.org/egems/vol3/iss1/7

This Informatics Model/Framework is brought to you for free and open access by the the EDM Forum Products and Events at EDM Forum Community. It has been peer-reviewed and accepted for publication in eGEMs (Generating Evidence & Methods to improve patient outcomes).

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Reporting recommendations related to Data processing/provenance:

– Mappings from original values to standardized values

- *"Documentation of how original data values were transformed to the target .. format."*
- *"Documentation should list source values and describe the logic or mappings used to transform original source to required target values."*

Recommendations

- Examine phenotype definitions to assess reliance on ICD-9
- Consider the phenotype definition as a “unit” or value set, and compare semantic equivalence of the set
- Consider different mapping approaches for automatic translation
- Examine research needs and nature of condition
- Be prepared to report methods for mapping
- Be prepared to validate locally
- Implement data quality assessment recommendations

Conclusion

- ICD 10 will enable researchers to make more targeted data queries and potential have more detailed data for patient risks or outcomes.
- ICD-10 transition will differentially threaten the research integrity and required resources for various types of studies.
- Studies where data collection includes the ICD-10 implementation date (October 1, 2015), researchers need to be cognizant of implications of the mapping relationships.

Acknowledgments

- **Study presented here is under review with eGEMS: "**Preparing for the ICD-10-CM Transition: Automated Methods for Translating ICD Codes in Clinical Phenotype Definitions and Quality Measure Value Sets." (Submitted July 2015)

  - **Olivier Bodenreider**, MD, PhD, National Library of Medicine
  - **Michelle Smerek**, Duke University
  - **Katherine C. Pereira**, DNP, Duke University
  - **Beverly B. Green**, MD, MPH, Group Health Research Institute
  - **Ashwin Patkar**, MD, Duke University
  - **Megan Clowse**, MD, Duke University
  - **Alan Bauck**, Center for Health Research, Kaiser Permanente Northwest

- This presentation was supported by the National Library of Medicine intramural research program, and the National Institutes of Health **(NIH) Common Fund**, through a cooperative agreement (U54 AT007748) from the Office of Strategic Coordination within the Office of the NIH Director.
- The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

- Many thanks to:
  - NIH Collaboratory – Phenotype, Data Standards, Data Quality Core
  - Duke Center for Predictive Medicine and Ben Neely, MS

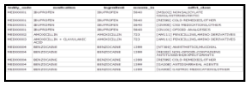
DISCUSSION

Research Concerns

- Performance
- Reproducibility
- Consistency

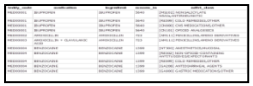
- Identify and eliminate potential bias

- Goal: at the point of randomization: a) the 2 groups have equal risk of having the outcome of interest; and b) the 2 groups are very well characterized at the point of the start of the trial.



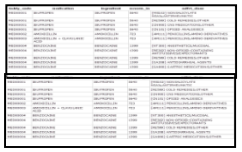
Dataset with
ICD-9 codes

+



Dataset with
ICD-10 codes

=



Merged Dataset w/
one coding system:

(ICD9CM) ICD-9-CM
- DISEASES AND INJURIES
  - CERTAIN CONDITIONS ORIGINATING IN THE PERINATAL PERIOD
  - COMPLICATIONS OF PREGNANCY, CHILDBIRTH, AND THE PUERPERIUM
  - CONGENITAL ANOMALIES
  - DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS
  - DISEASES OF THE CIRCULATORY SYSTEM
  - DISEASES OF THE DIGESTIVE SYSTEM
  - DISEASES OF THE GENITOURINARY SYSTEM
  - DISEASES OF THE MUSCULOSKELETAL SYSTEM AND CONNECTIVE TISSUE
  - DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS
  - DISEASES OF THE RESPIRATORY SYSTEM
  - DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE
  - ENDOCRINE, NUTRITIONAL AND METABOLIC DISEASES, AND IMMUNITY DISC
  - INFECTIOUS AND PARASITIC DISEASES
  - INJURY AND POISONING
  - MENTAL, BEHAVIORAL AND NEURODEVELOPMENTAL DISORDERS
  - NEOPLASMS
  - SYMPTOMS, SIGNS, AND ILL-DEFINED CONDITIONS

ICD-9-CM

SNOMED-CT

(ICD10CM) ICD-10-CM TABULAR LIST of DISEASES and INJURIES
- Certain conditions originating in the perinatal period (P00-P96)
  - Abnormal findings on neonatal screening (P09)
  - Birth trauma (P10-P15)
  - Conditions involving the integument and temperature regulation of newborn (P80-P83)
    - Hypothermia of newborn
    - Other conditions of integument specific to newborn
    - Other disturbances of temperature regulation of newborn
  - Digestive system disorders of newborn (P76-P78)
    - Necrotizing enterocolitis of newborn
    - Other intestinal obstruction of newborn
    - Other perinatal digestive system disorders
  - Disorders of newborn related to length of gestation and fetal growth (P05-P08)
    - Disorders of newborn related to long gestation and high birth weight
    - Disorders of newborn related to short gestation and low birth weight, not elsewhere classified
    - Disorders of newborn related to slow fetal growth and fetal malnutrition
  - Hemorrhagic and hematological disorders of newborn (P50-P61)
    - Disseminated intravascular coagulation of newborn
    - Hemolytic disease of newborn
    - Hemorrhagic disease of newborn
    - Hydrops fetalis due to hemolytic disease
    - Intracranial nontraumatic hemorrhage of newborn
    - Kernicterus
    - Neonatal jaundice due to other excessive hemolysis
    - Neonatal jaundice from other and unspecified causes
    - Newborn affected by intrauterine (fetal) blood loss
    - Other neonatal hemorrhages
    - Other perinatal hematological disorders
    - Umbilical hemorrhage of newborn
  - Infections specific to the perinatal period (P35-P39)
  - Newborn affected by maternal factors and by complications of pregnancy, labor, and delivery (P00-P04)
  - Other disorders originating in the perinatal period (P90-P96)
  - Other problems with newborn (P84)
  - Respiratory and cardiovascular disorders specific to the perinatal period (P15-P29)
  - Transitory endocrine and metabolic disorders specific to newborn (P70-P74)
- Certain infectious and parasitic diseases (A00-B99)
- Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99)
  - Chromosomal abnormalities, not elsewhere classified (Q90-Q99)
  - Cleft lip and cleft palate (Q35-Q37)
  - Congenital malformations and deformations of the musculoskeletal system (Q65-Q79)
  - Congenital malformations of eye, ear, face and neck (Q10-Q18)
  - Congenital malformations of genital organs (Q50-Q56)
  - Congenital malformations of the circulatory system (Q20-Q28)
  - Congenital malformations of the nervous system (Q00-Q07)
  - Congenital malformations of the respiratory system (Q30-Q34)
  - Congenital malformations of the urinary system (Q60-Q64)
  - Other congenital malformations (Q80-Q89)
  - Other congenital malformations of the digestive system (Q38-Q45)
- Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D
- Diseases of the circulatory system (I00-I99)
  - Acute rheumatic fever (I00-I02)
  - Cerebrovascular diseases (I60-I69)
  - Chronic rheumatic heart diseases (I05-I09)
  - Diseases of arteries, arterioles and capillaries (I70-I79)
  - Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified (I80-I89)
  - Hypertensive diseases (I10-I15)
  - Ischemic heart diseases (I20-I25)
  - Other and unspecified disorders of the circulatory system (I95-I99)
  - Other forms of heart disease (I30-I52)

ICD-10-CM

ICD-9-CM

ICD-10-CM

SNOMED CT

others…

ICD-10 Issues by Use Cases

- **Estimating numbers of patients potentially eligible for a proposed trial (study feasibility).**

Possible Impact: low ; research planning is a one-time activity. In the past it was done with ICD-9, but now can be done with ICD-10

Activities: ICD-9 based phenotypes will need to be converted to ICD-10.

ICD-10 Issues by Use Cases

- **Identifying patients for recruitment into prospective trials. ~cohort identification**

- ***Possible Impact:*** High. If the study recruitment occurs before and after Oct. 1, 2015, then there is a danger that those recruited after the transition are *not* the same as those before.
- Could lead to sampling bias if there are differences (including certainty of disease and severity of conditions) between patients recruited early versus late in study.

- ***Activities***: ICD-9 Based phenotypes need to be converted to ICD-10 **– and clinically validated.**

ICD-10 Issues by Use Cases

- **Describing patient cohorts for analysis of existing data for comparative effectiveness or health services research.**

Possible Impact: Moderate. If the data analyzed in the study was collected from health systems before and after Oct. 1, 2015, then there might be a systematic bias.

Activities: ICD-9 Based phenotypes need to be converted to ICD-10 **– and clinically validated.** Data quality assessment recommendations can be applied.

ICD-10 Issues by Use Cases

- **Presenting baseline characteristics or conditions to describe research populations by demographics, clinical features, and co-morbidities for clinical trials.**

Possible Impact: High. If the study recruitment occurs before and after Oct. 1, 2015, then there is a danger that those recruited after the transition are *not* the same as those before. Could lead to sampling bias if there are differences (including certainty of disease and severity of conditions) between patients recruited early versus late in study.

Activities: ICD-9 Based phenotypes need to be converted to ICD-10 **– and clinically validated.**

ICD-10 Issues by Use Cases

- **Presenting primary outcomes to test the trial hypothesis.**
- **The implementation of supportive tools for providers that are embedded in EHR systems and clinical workflows.**

Possible Impact: High. If the study outcomes are assessed for some patients before and some after Oct. 1, 2015, then there could be differences (including certainty of disease and severity of conditions) between patients assessed early versus late in study.

Activities: ICD-9 Based phenotypes need to be converted to ICD-10. Aggressive (iterative) mapping processes appropriate. New ICD-10 groups **must be clinically validated.** Data quality assessment recommendations can be applied.

Dimensions of Quality

Table 1. Data Quality Dimensions Determining Fitness for Use of Research Data

Dimension	Conceptual definition	Operational examples
Completeness	Presence of the necessary data	Presence of necessary data elements, percent of missing values for a data element, percent of records with sufficient data to calculate a required variable (e.g., an outcome)
Accuracy	Closeness of agreement between a data value and the true value*	Percent of data values found to be in error based on a gold standard, percent of physically implausible values, percent of data values that do not conform to range expectations
Consistency	Relevant uniformity in data across clinical investigation sites, facilities, departments, units within a facility, providers, or other assessors	Comparable proportions of relevant diagnoses across sites, comparable proportions of documented order fulfillment (e.g., returned procedure report for ordered diagnostic tests)