Optimizing Clinical Research Participant Selection Design with Informatics

Zhe He
School of Information
Institute for Successful Longevity
Florida State University

INSTITUTE FOR SUCCESSFUL LONGEVITY®

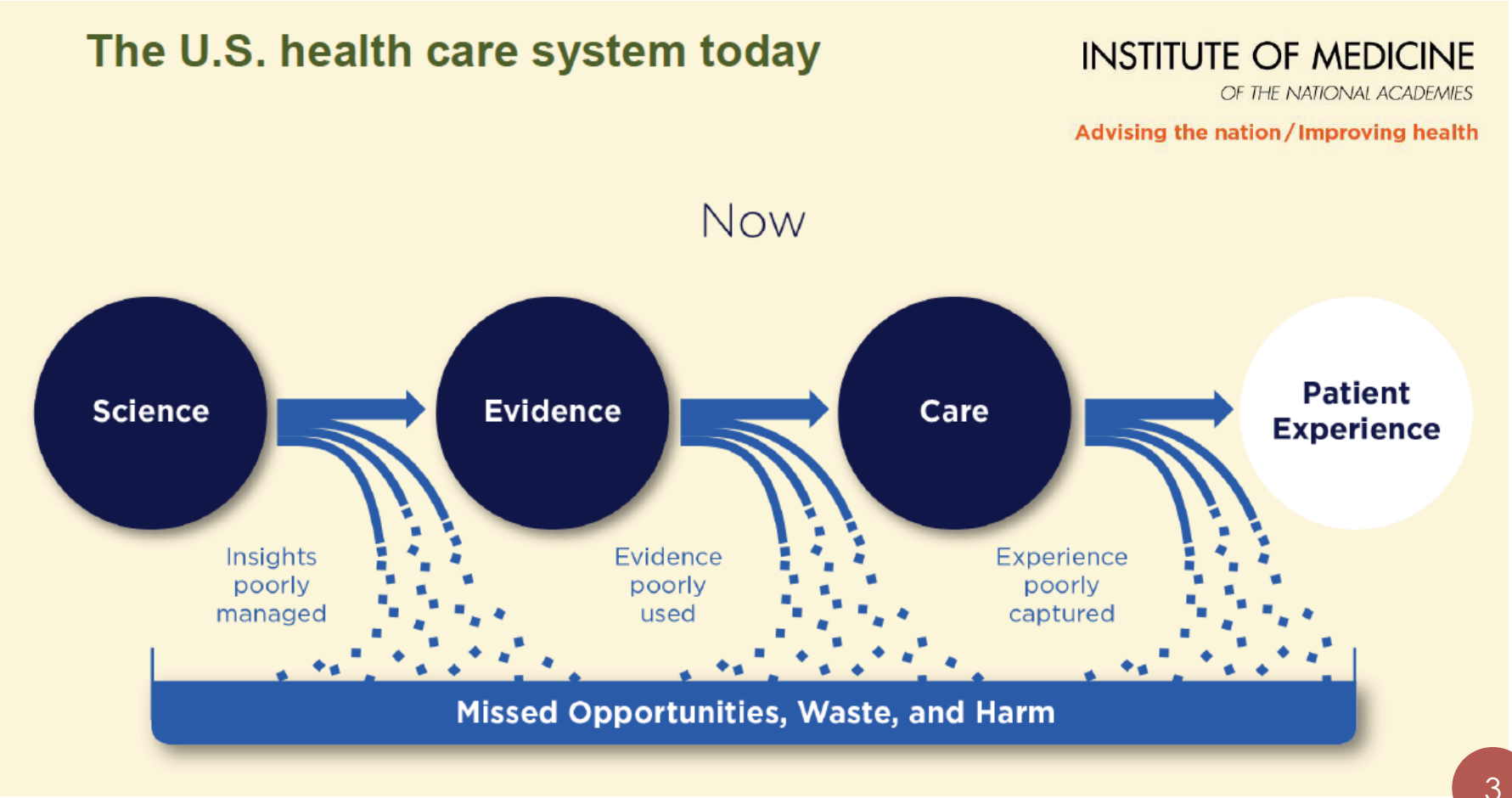
COLLEGE OF COMMUNICATION & INFORMATION
THE FLORIDA STATE UNIVERSITY

Overview

- Generalizability of Clinical Trials
- Methodologies for Quantifying the Population Representativeness of Clinical Studies
- Population Representativeness Assessment
  - Type 2 Diabetes Clinical Studies
  - Colorectal Cancer Treatment Trials
- Identifying Restrictive Qualitative Eligibility Criteria

Towards a Learning Health System

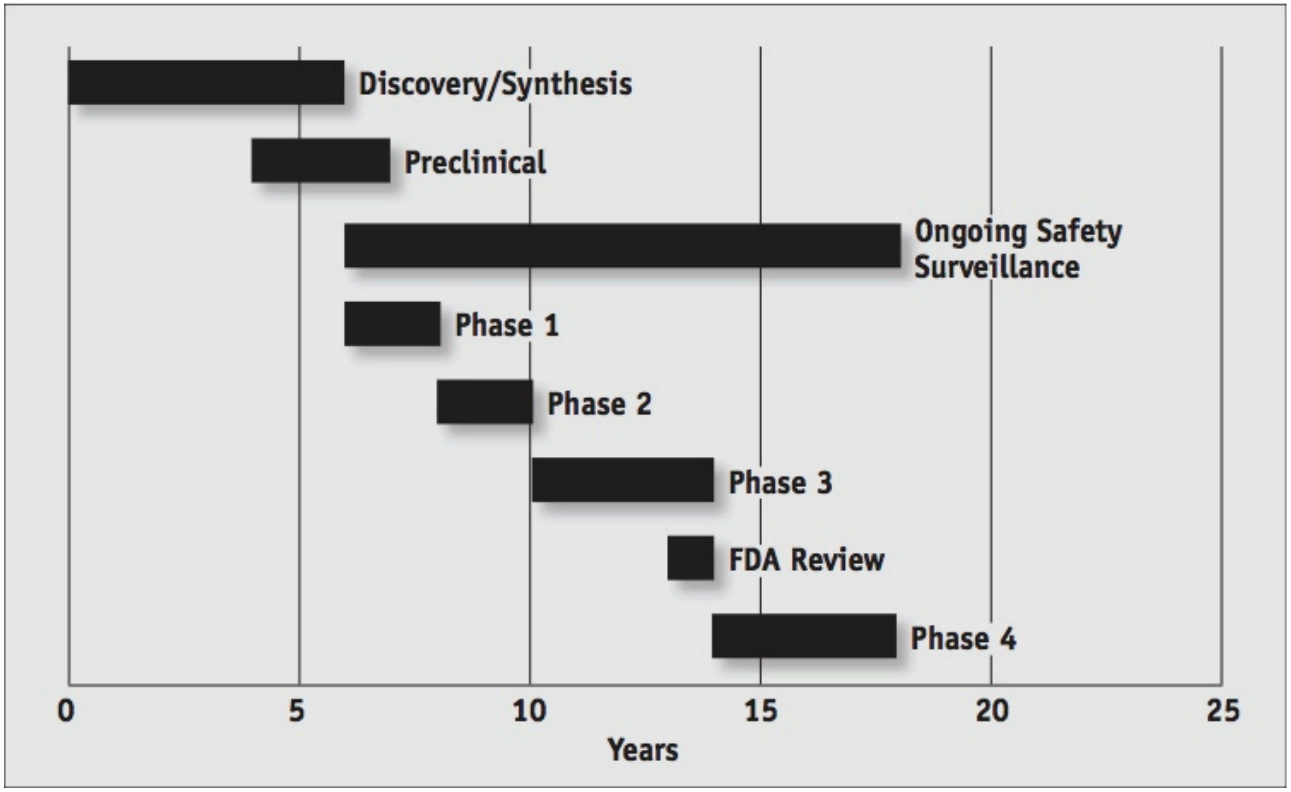
- IOM wrote that U.S. medicine wastes roughly $750 billion per year, permits tens of thousands of needless deaths and bungles its mission in ways foreign to other industries.



https://www.advisory.com/Daily-Briefing/2012/09/07/IOM-report

What are Clinical Trials?

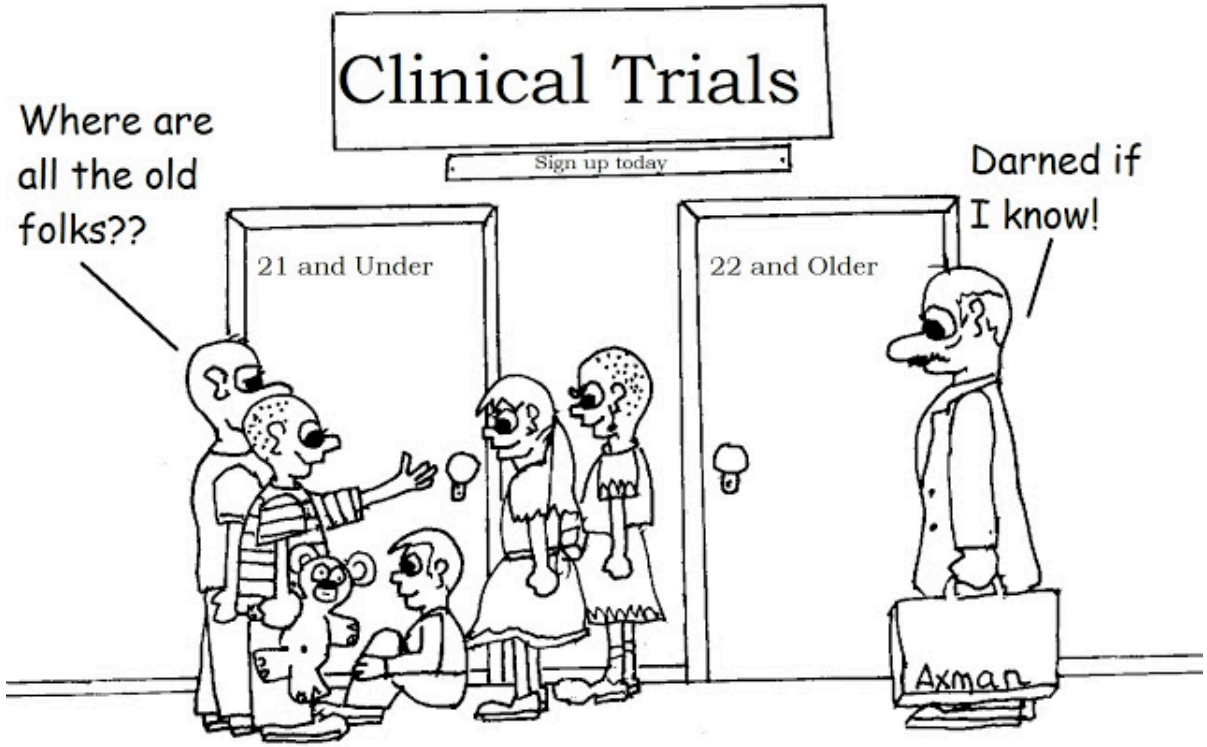
- Investigate how well a treatment for a medical condition work in people
- Generate gold standard evidence for modern medical research



http://www.cancer-clinical-trials.com/p/clinical-trials-in-cartoons.html#.VkIBYK6rTR0

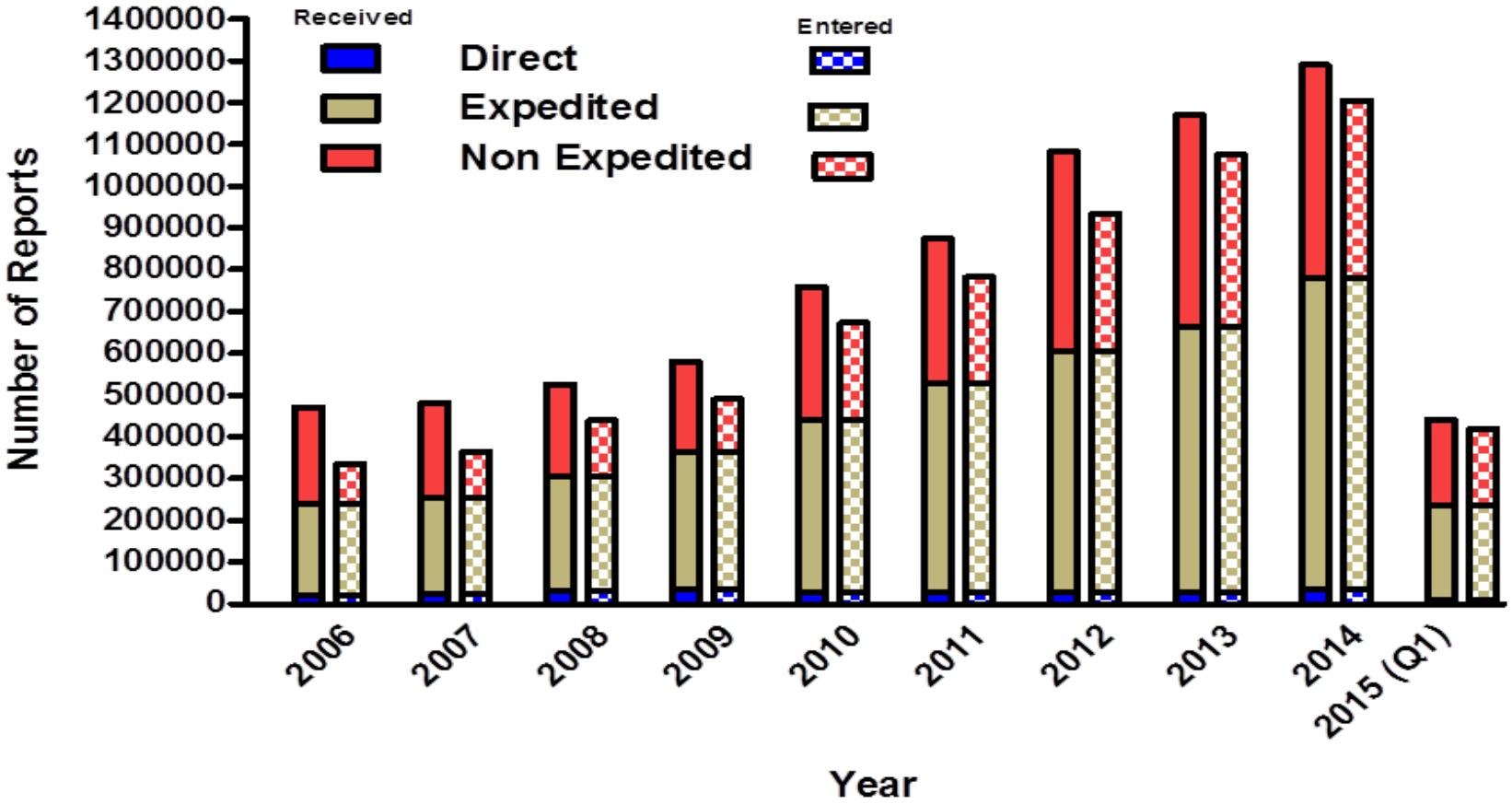
Generalizability of Clinical Trials

- Often fail to recruit enough patients
  - Stringent eligibility criteria
- Lack of population representativeness
  - Compromised generalizability
  - Adverse drug reactions



http://94.242.228.149/cancer-clinical-trials-in-cartoons-clipart-464169.html.

FDA's Adverse Event Reporting System



- More than 75 drugs/drug products have been removed from the market due to safety problems.

http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/

Celldex therapeutics

« Previous Release | Next Release »

November 20, 2015

Long-term Survival Benefit Demonstrated in Phase 2 ReACT Study of RINTEGA(R) in Recurrent Bevacizumab-naive Glioblastoma

-- At 2 years, 25% RINTEGA survival rat
-- Hazard ratio of 0.53 (p=0.0137) indicates a significant overall
-- Clinical benefit continues to be observed across multiple endpoints, including PF
favoring RINTEGA trea

HAMPTON, N.J., Nov. 20, 2015 (GLOBE NEWSWIRE) -- Celldex Therapeutics, Inc.
Company's randomized, double-blind Phase 2 study of RINTEGA® (rindopepimut) in
20th Annual Scientific Meeting of the Society for Neuro-Oncology (SNO). The data w
Clinical Director, Center for Neuro-Oncology, Dana-Farber Cancer Institute; Associat
the Society for Neuro-Oncology, as well as the lead investigator of the ReACT study.
and was granted Breakthrough Therapy Designation in February 2014. Patients with
have a worse prognosis than the overall glioblastoma population, including poor long
positive patients is 8.7 months[1]). As previously reported, the primary endpoint of the

Phase II: 40 patients

Phase III: 700 patients

cure®

Search CURE

SIGN IN | SUBSCRIBE | NEWSLETTERS

LATEST FORUM VIDEOS VOICES EVENTS ADVOCACY SUBSCRIBE SHARE YOUR STO

The Second Generation Optune® System
Smaller. Lighter. Designed for everyday life.
Explore Optune
OPTUNE

Contraindic
Do not use O
implanted me
as, missing b

Your Journey >>

Before, During & After Treatment

Cancer Types >>

Overview
Bladder
Brain
Breast
CLL
Colorectal
Head and Neck
Immuno Oncology
Kidney

Rintega (Rindopepimut) Misses Survival Endpoint in Phase 3 Glioblastoma Trial

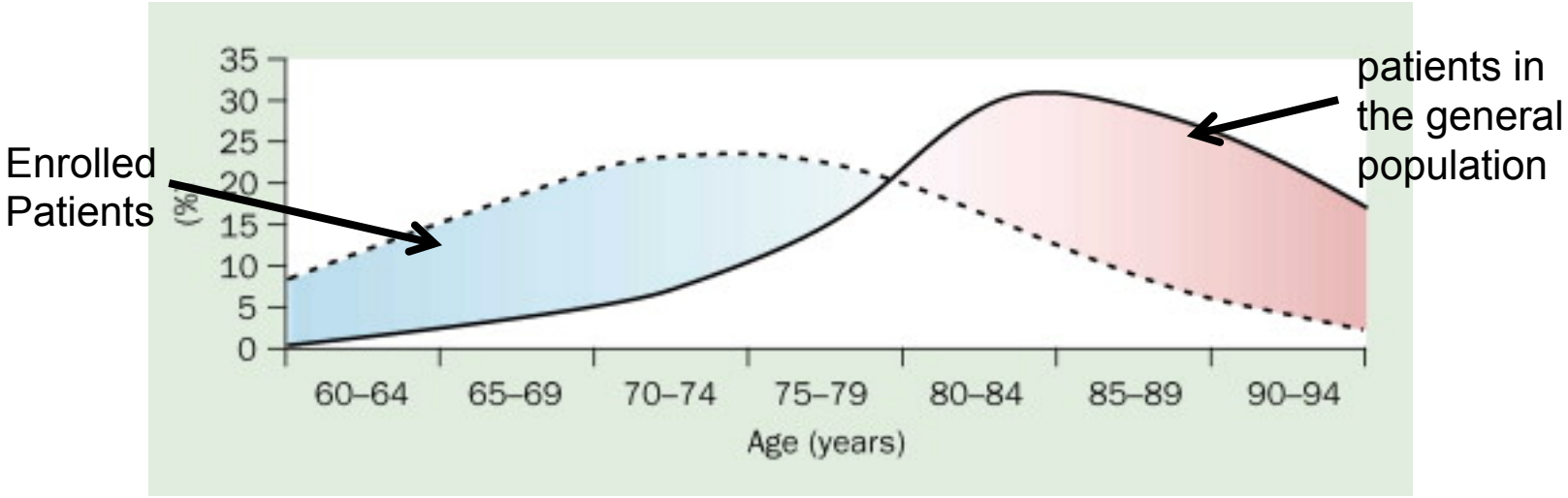
The immunotherapy will continue to be offered in the phase 3 trial and compassionate use programs.

SILAS INMAN @silasinman
PUBLISHED: MARCH 10, 2016

Be the first to discuss this article on CURE's forum. >>

Talk about this article with other patients, caregivers, and advocates in the **Brain Cancer** CURE discussion group.

Dementia Trials



(Schoenmaker *et al.* 2004)

Schoenmaker N, Van Gool WA. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. Lancet Neurol. 2004;3(10):627-30. PMID: 15380160

American Heart Journal
Volume 146, Issue 2, August 2003, Pages 250–257

Clinical investigations: congestive heart failure

Most hospitalized older persons do not meet the enrollment criteria for clinical trials in heart failure ☆

Frederick A Masoudi, MD, MSPH[a, b, c], Edward P Havranek, MD[a, b, c], Pam Wolfe, MA, MS[c], Cary P Gross, MD[d], Saif S Rathore, MPH[d], John F Steiner, MD, MPH[b], Diana L Ordin, MD, MPH[e], Harlan M Krumholz, MD[d]

Show more

http://dx.doi.org/10.1016/S0002-8703(03)00189-3

Abstract

Background

Although it is widely accepted that clinical trials in heart failure may nc

CLINICAL GASTROENTEROLOGY AND HEPATOLOGY 2012;10:1002–1007

Patients Enrolled in Randomized Controlled Trials Do Not Represent the Inflammatory Bowel Disease Patient Population

CHRISTINA HA,* THOMAS A. ULLMAN,‡ COREY A. SIEGEL,§ and ASHER KORNBLUTH‡

*Division of Gastroenterology, The Johns Hopkins School of Medicine, Baltimore, Maryland; ‡Dr. Henry D. Janowitz Division of Gastroenterology, Mount Sinai School of Medicine, New York, New York; and §Division of Gastroenterology, Dartmouth-Hitchcock Medical Center, Hanover, New Hampshire

ı activity on page e78. Learning Objectives—At the end
the potential limitations of clinical trial data when

rohn's disease (CD) and ulcerative colitis (UC) are chronic
inflammatory diseases of the gastrointestinal tract char-
erized by symptomatic relapses often requiring escalation of
dical therapy. The introduction of the biologic anti-tumor
:rosis factor alpha (anti-TNF) agents has transformed the

Cancer

Explore this journal >

Original Article

Generalizability of cancer clinical trial results†‡

Prognostic differences between participants and nonparticipants

Linda S. Elting Dr.P.H., Catherine Cooksley Dr.P.H.,
B. Nebiyou Bekele Ph.D., Michael Frumovitz M.D.,
Elenir B. C. Avritscher M.D., Charlotte Sun Dr.P.H., Diane C. Bodurka M.D.

First published: 25 April 2006 Full publication history

DOI: 10.1002/cncr.21907 View/save citation

View issue TOC
Volume 106, Issue 11
1 June 2006
Pages 2452–2458

General Public

The New York Times
SundayReview | The Opinion Pages

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

Prevnar 13®
Pneumococcal 13-valent Conjugate Vaccine
(Diphtheria CRM197 Protein)
Over 50?
Ask your doctor
about PREVNAR 13®
Learn more
AdChoices
SCROLL FOR IMPORT
IMPORT
• Prevnar 13® should
severe allergic react
diphtheria toxoid–co

OPINION

Do Clinical Trials Work?

By CLIFTON LEAF
Published: July 13, 2013 | 114 Comments

EVERY spring, some 30,000 oncologists, medical researchers and marketers gather in an American city to showcase the latest advances in cancer treatment.

Enlarge This Image

Ruth Gwily

Enlarge This Image

But at the annual meeting of the American Society of Clinical Oncology last month, much of the buzz surrounded a study that was anything but a breakthrough. To a packed and whisper-quiet room at the McCormick Place convention center in Chicago, Mark R. Gilbert, a professor of neuro-oncology at the University of Texas M. D. Anderson Cancer Center in Houston, presented the results of a clinical trial testing the drug Avastin in patients newly diagnosed with glioblastoma multiforme, an aggressive brain cancer. In two earlier, smaller studies of patients with

FACEBOOK
TWITTER
GOOGLE+
SAVE
EMAIL
SHARE
PRINT
REPRINTS

BIRDMAN
(THE UNEXPECTED VIRTUE OF IGNORANCE)
NOW PLAYING TICKETS

"Roughly **53 percent** of new cancer diagnoses are in people **65 or older**, but this age group accounts for just **33 percent** of participants in cancer drug trials."

Leaf C. "Do Clinical Trials Work?" *New York Times*. July 13, 2013.

What is Clinical Research Informatics?

Clinical Research Informatics involves the use of ***informatics*** *in the* ***discovery and management of new knowledge*** *relating to health and disease. It includes* ***management of information related to clinical trials*** *and also involves* ***informatics related to secondary research use of clinical data****.*

http://www.amia.org/applications-informatics/clinical-research-informatics

How can we optimize participant selection?

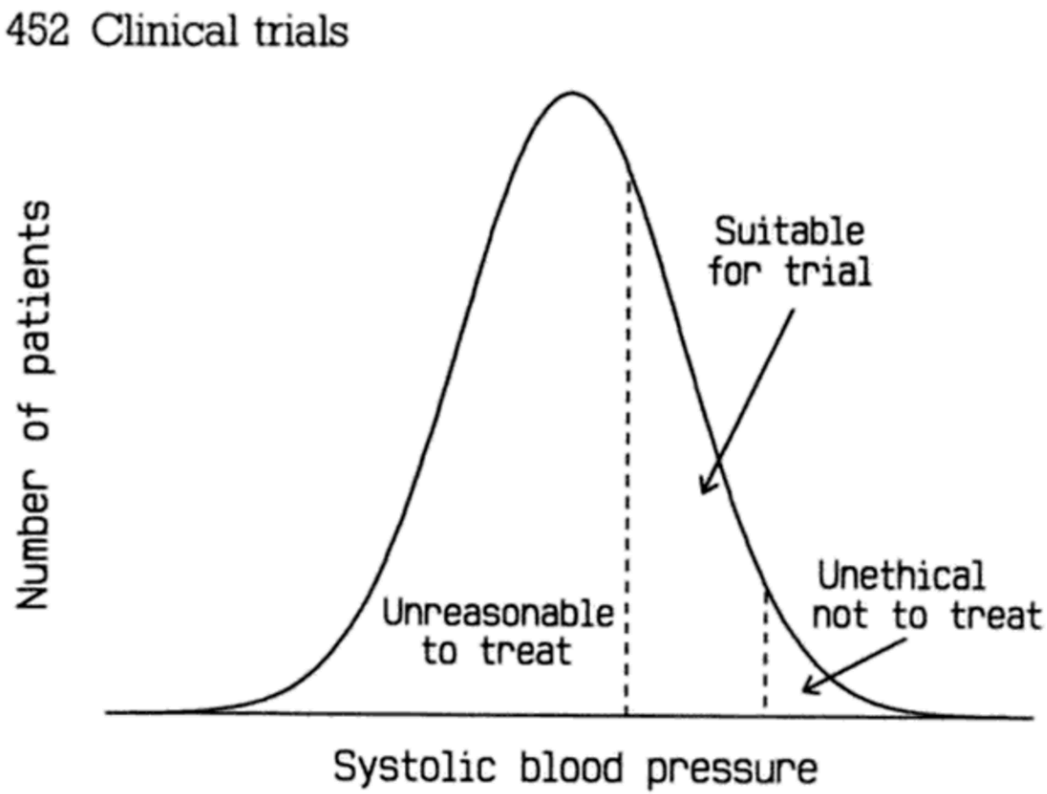
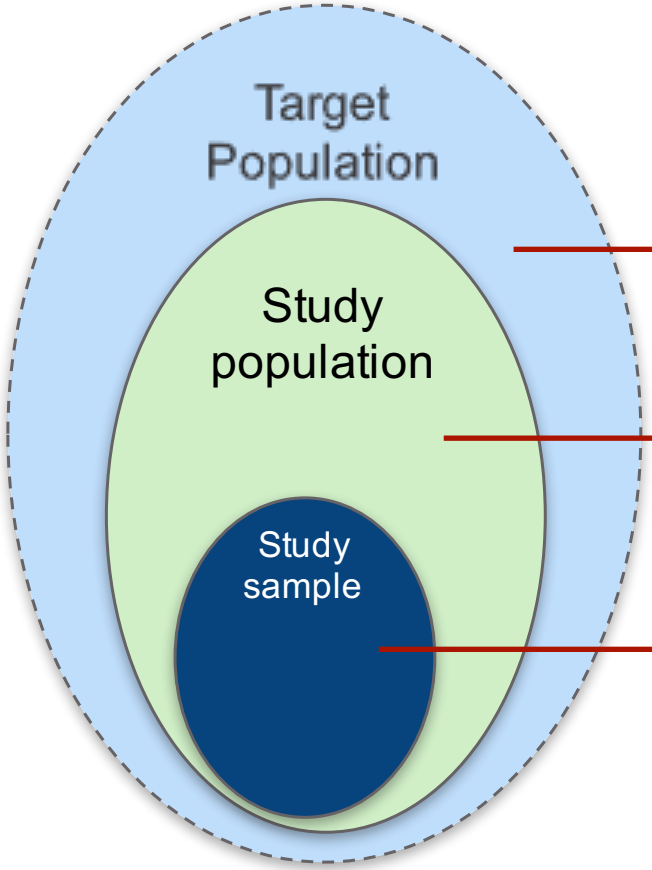


Figure 15.1 Diagram showing the eligibility of patients for a trial of a new antihypertensive agent (based on Elwood, 1982).

Practical Statistics for Medical Research, Douglas Altman, Publisher: Chapman and Hall/ CRC; 1st ed edition (November 22, 1990)Language: English, ISBN-10: 0412276305

Populations in Clinical Studies



Patients to whom the results of clinical trials are intended to be applied

Patients being sought as defined in clinical trial eligibility criteria

Patients actually enrolled in clinical trials

ClinicalTrials.gov

Eligibility

Ages Eligible for Study: 18 Years to 80 Years
Genders Eligible for Study: Both
Accepts Healthy Volunteers: No

Criteria

Inclusion Criteria:

- Patients between the ages of 18 to 80 years with Type 2 Diabetes who are on a stable dose of antidiabetic medication for at least 3 months

Exclusion Criteria:

- Patients taking Cholestin, niacin (>50 mg/day), fibrate therapy, hormonal contraceptives, intermittent Hormone Replacement Therapy, or certain corticosteroids
- Patients with any of the following conditions: active liver disease or kidney disease, poorly controlled high blood pressure, active peptic ulcer, or other heart or blood diseases
- Patients with abnormal laboratory results from a blood test that will be given before starting the study

Contacts and Locations

Choosing to participate in a study is an important personal decision. Talk with your doctor and family members or friends about deciding to join a study. To learn more about this study, you or your doctor may contact the study research staff using the Contacts provided below. For general information, see Learn About Clinical Studies.

Please refer to this study by its ClinicalTrials.gov identifier: NCT00485758

Sponsors and Collaborators

Merck Sharp & Dohme Corp.

Investigators

Study Director: Medical Monitor Merck Sharp & Dohme Corp.

Hao T, Rusanov A, Boland MR, Weng C. Clustering clinical trials with similar eligibility criteria features. *J Biomed Inform*. 2014;pii: S1532-0464(14)00011-2.

Clinical Trial Eligibility Criteria

- Impact of GBS on CVD in Type 2 Diabetes Mellitus (NCT00787670)
- Aims to determine whether surgically induced weight loss decreases the risk of CVD in morbidly obese subjects with T2DM

Inclusion criteria: Qualitative

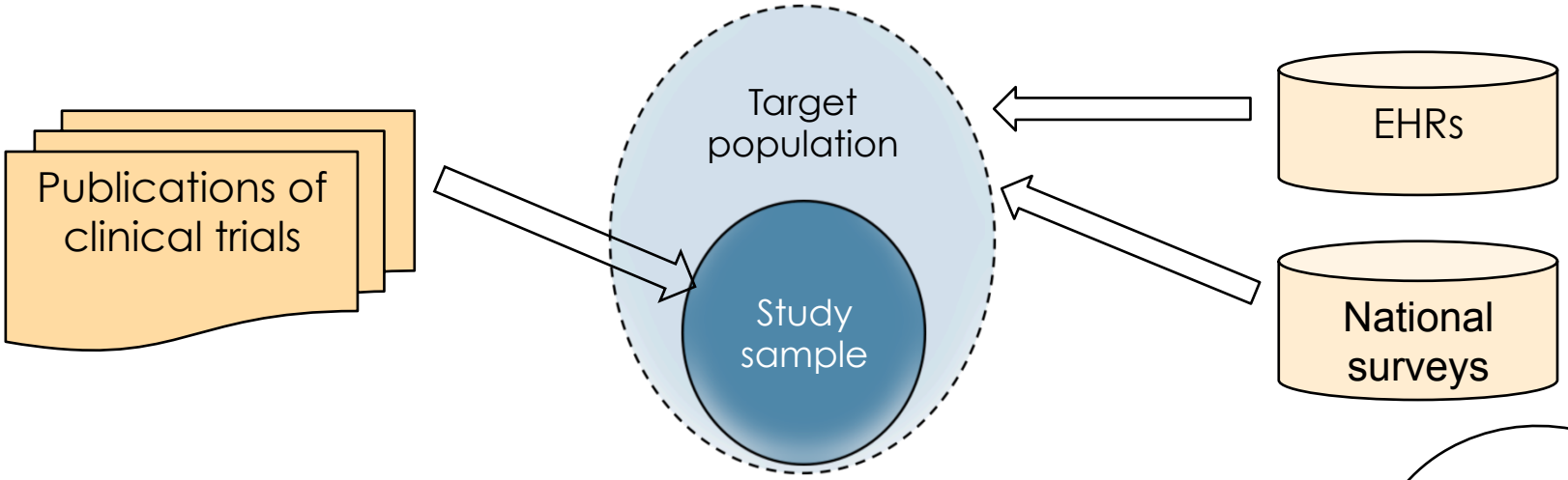
- Clinical diagnosis of T2DM diabetes mellitus with HbA1c ≤ 10.0% Quantitative
- Current regular use of insulin Qualitative
- Current regular use of oral hypoglycemic medication. Qualitative
- Body mass index ≥ 35 kg/m2 Quantitative
- Age between 18-60 years old. Quantitative
- ….

Exclusion criteria:

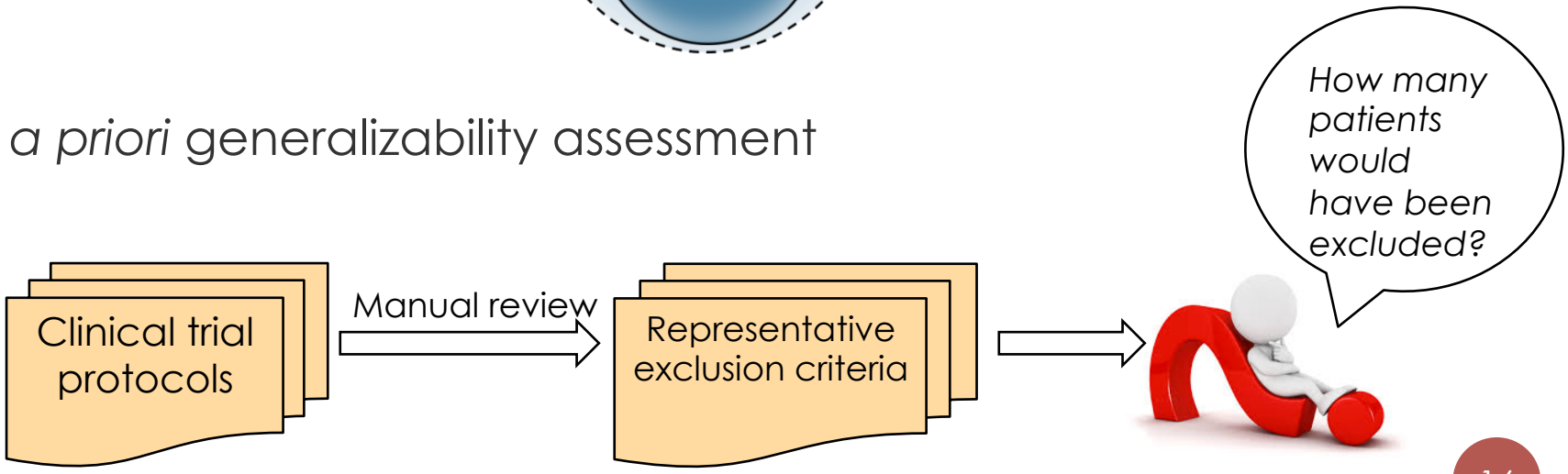
- Significant non-diabetic co-morbidity affecting life expectancy (e.g., malignancy) Qualitative
- Significant other co-morbidities (e.g. psychiatric disorder) that results in ineligibility for gastric bypass surgery Qualitative
- Pregnancy or planning pregnancy Qualitative

Traditional Generalizability Assessment

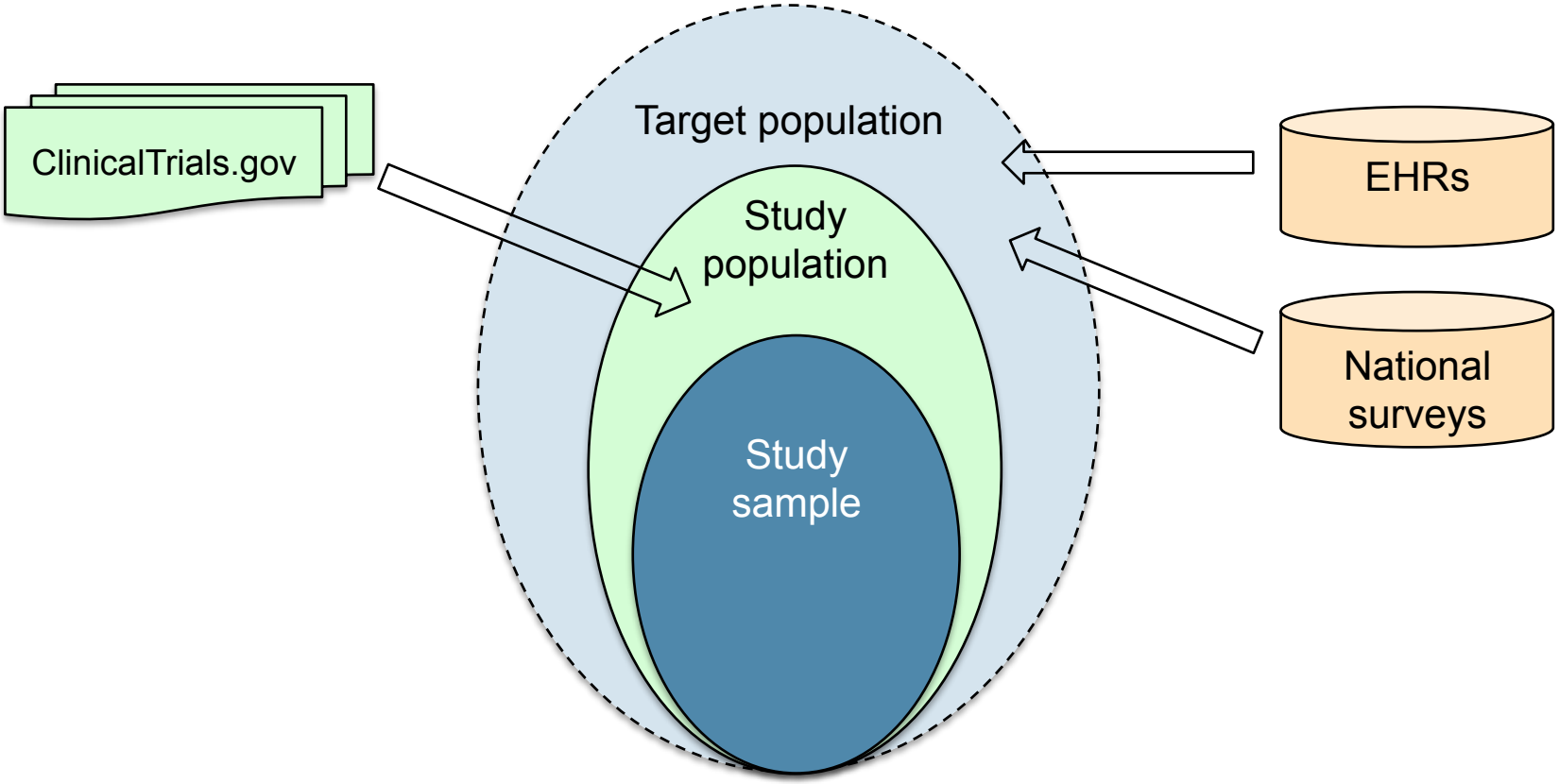
- *a posteriori* generalizability assessment



- *a priori* generalizability assessment



Innovated Way for a priori Generalizability Assessment



Free Text => Computable Data Elements

HbA1c >= 7.0

A1c > 7.0

HbA1c <= 10.0

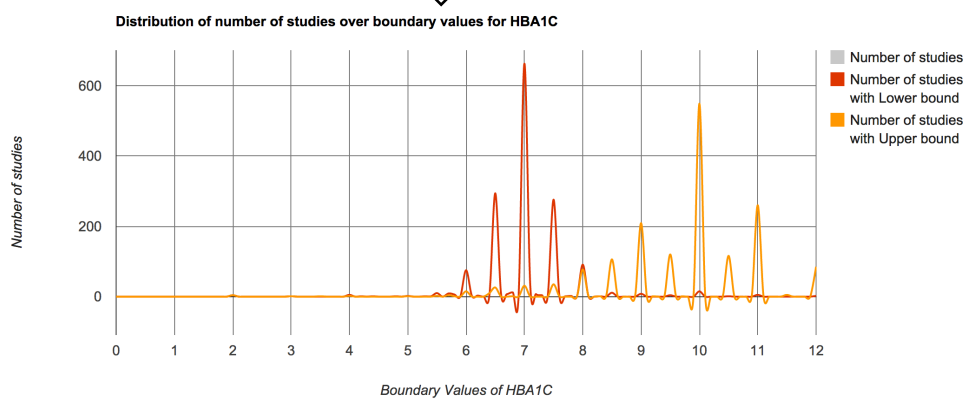
HbA1c > 11.0

A1c less than 6.5

Hemoglobin A1c < 11.0

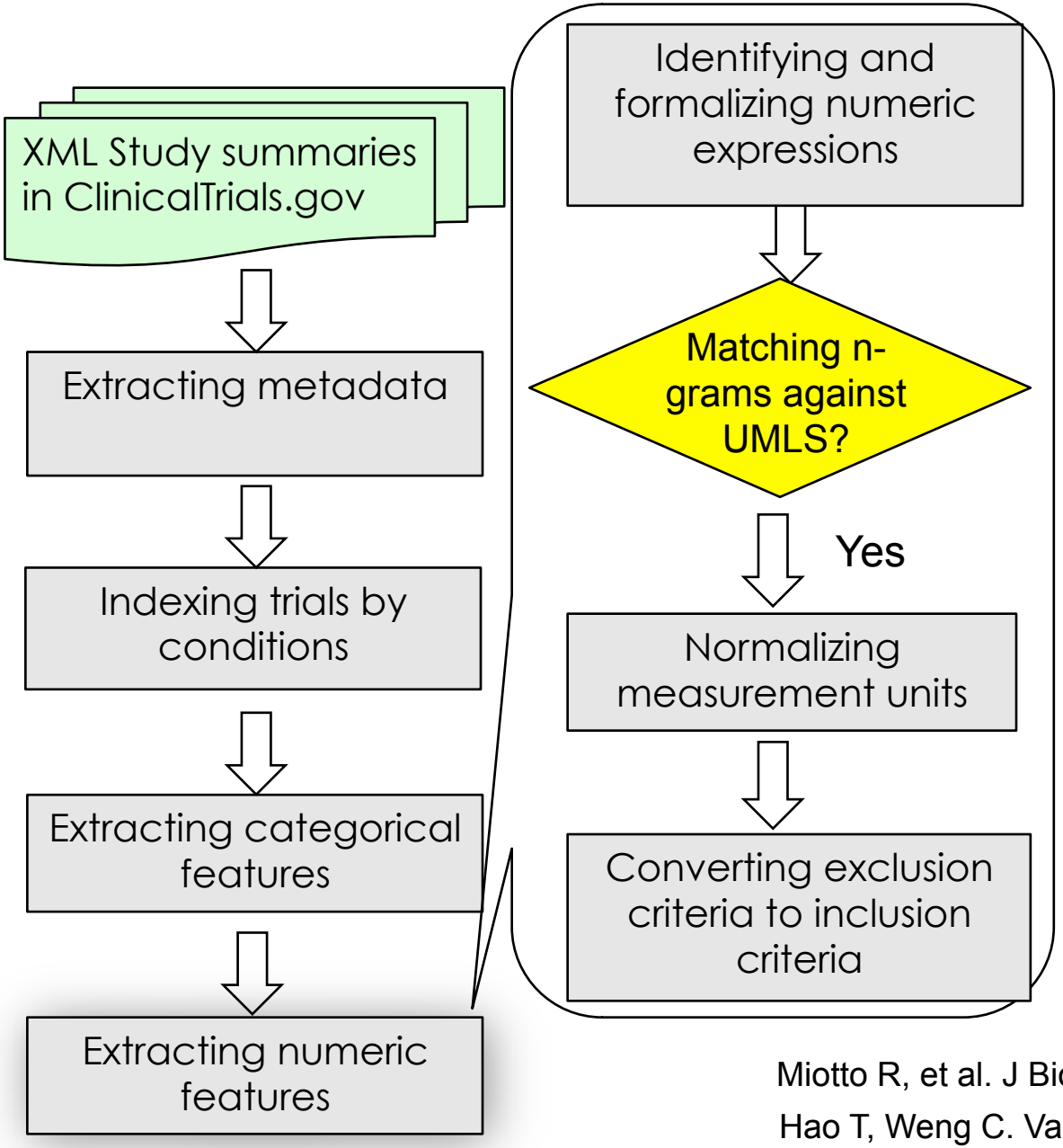
Unifying names and units

COMPACT



He et al. (2014) A Method for Analyzing Commonalities in Clinical Trial Target Populations. AMIA Annu Symp Proc: 2014: 1777-1786.

ClinicalTrials.gov → COMPACT

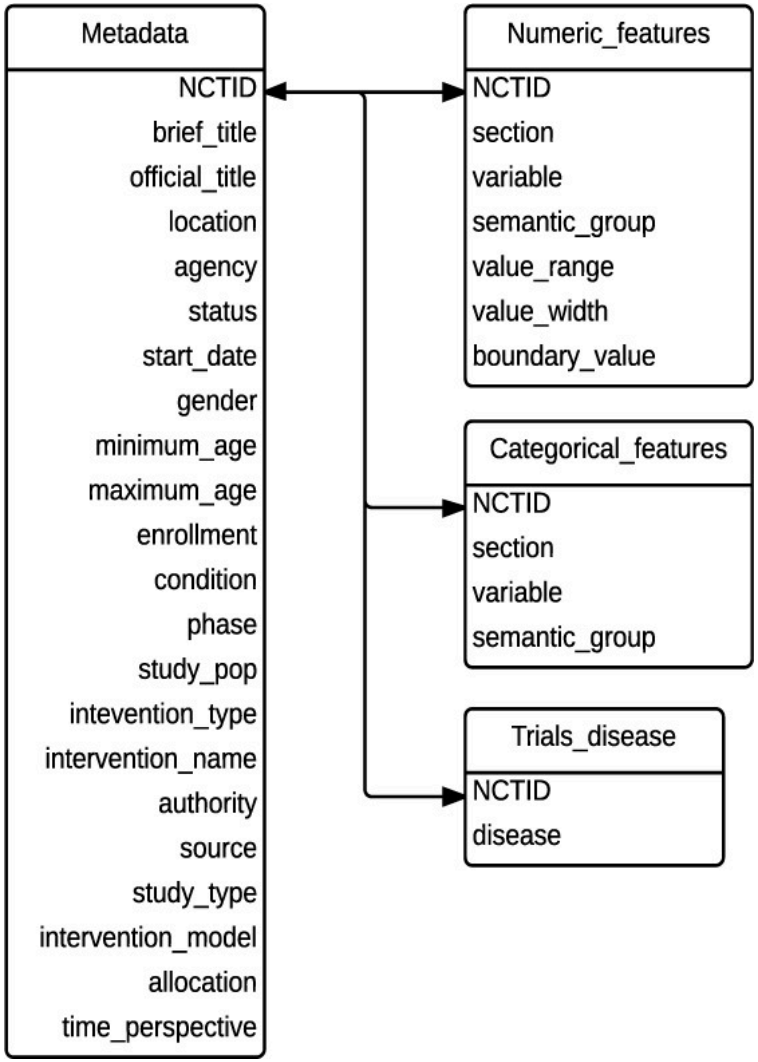


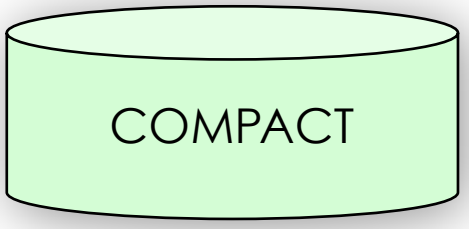
Inclusion criteria:
HbA1c value between 7.5% and 11%

Numeric Features:
["HbA1c", ">=", 7.5, "%"]
["HbA1c", "<=", 11, "%"]

Miotto R, et al. J Biomed Inform. 2013
Hao T, Weng C. Valx tool. 2014

COMPACT





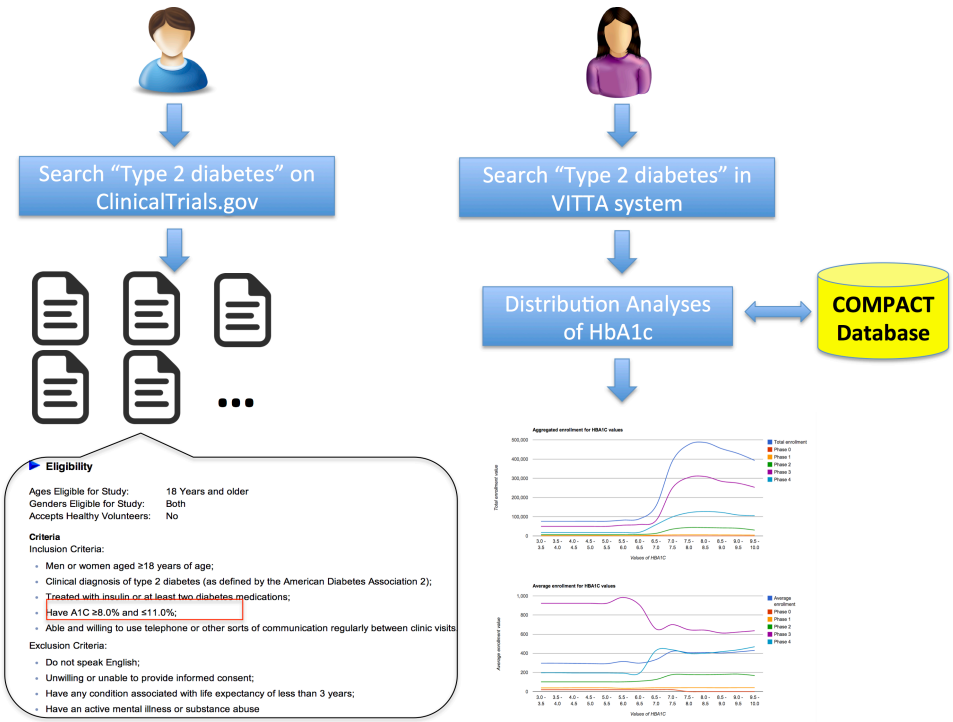
(Sept 2016)

- 220,000 + clinical studies
- 1,700+ medical conditions
- 989,042 numeric expressions
- Updated bi-annually

He et al. (2014) A Method for Analyzing Commonalities in Clinical Trial Target Populations. AMIA Annu Symp Proc: 2014: 1777-1786.

Web-Based Visual Analytics Tool for Quantitative Eligibility Features (VITTA)

- A web-based visualization tool developed to show studies vary in their study population with respect to a study trait
  - Over 1400 medical conditions
  - http://is.gd/VITTA



He Z, Carini S, Sim I, and Weng C. (2015) Visual aggregate analysis of eligibility features of clinical trials. *J Biomed Inform*. 2015, 54:241-245.

VITTA Demo

vittaec2-env-7gysespm39.elasticbeanstalk.com

Apps | Research Demo | TODO list | Impact Factors | Diary FSU | KRS Outline - Google | Teaching | Other Bookmarks

Visual Analysis Tool of Study Populations of Clinical Trials

Condition: 21-hydroxylase-deficiency

OR

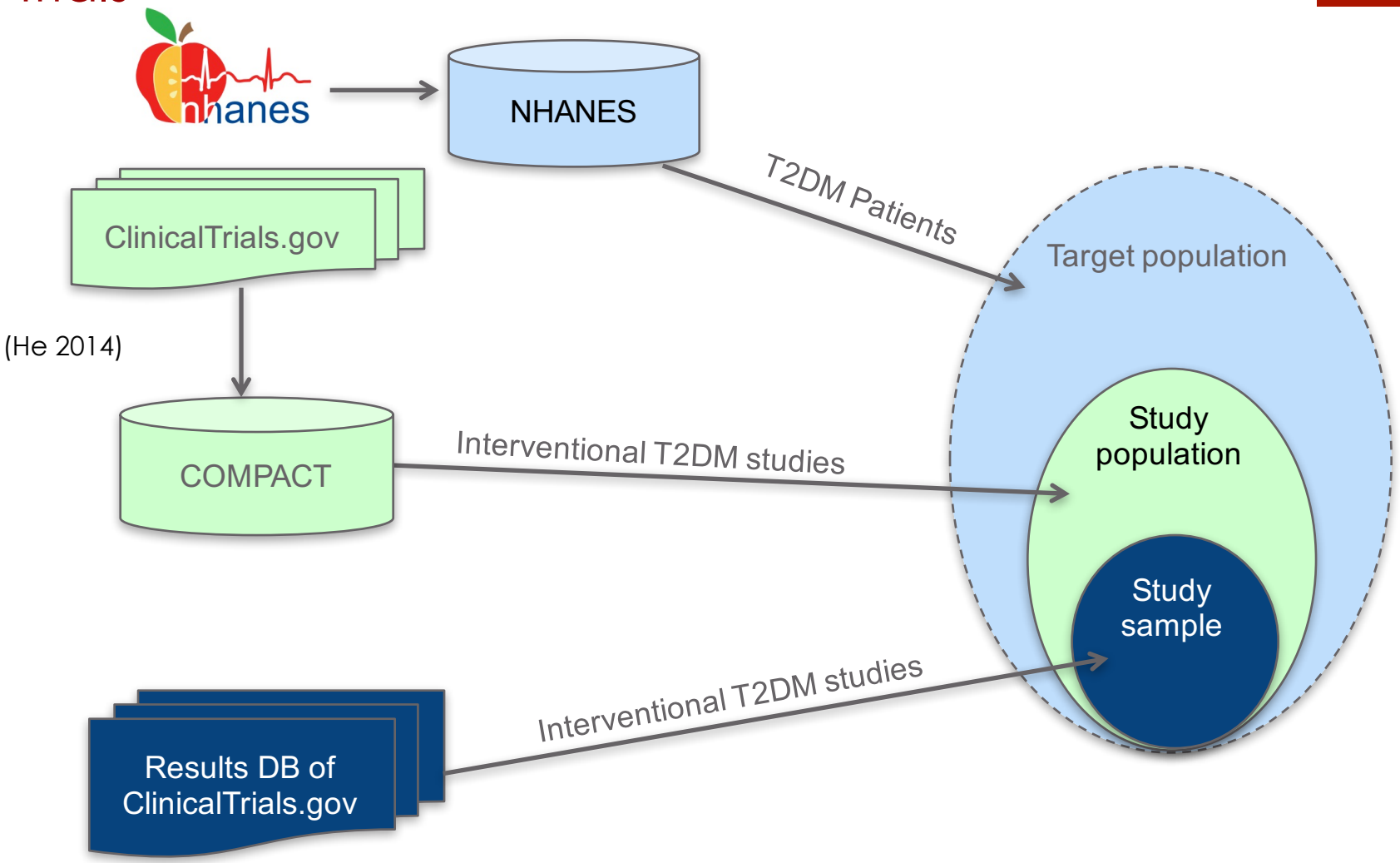
Type a condition

Top 10 Features

Process

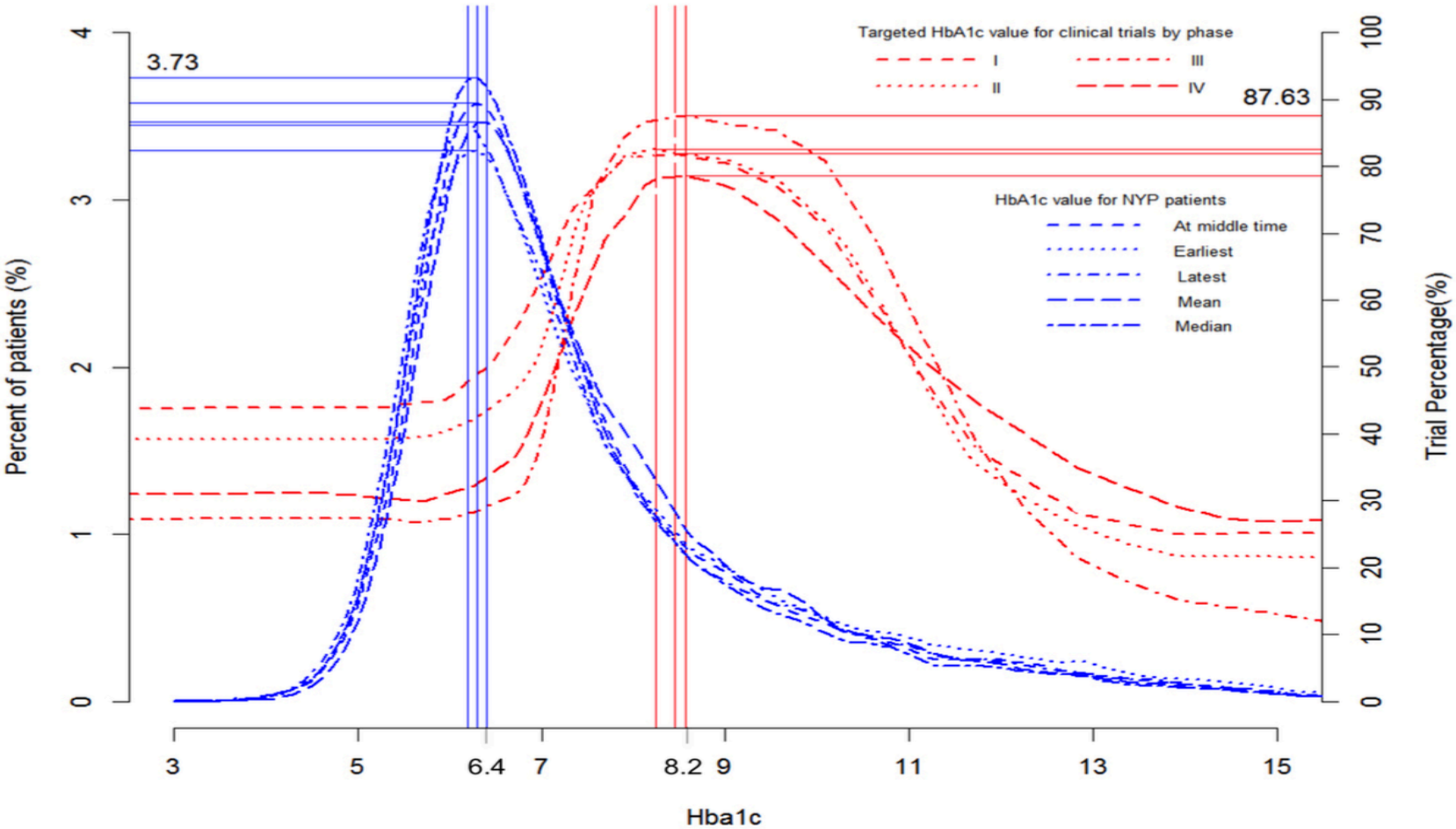
This system contains 163,285 clinical trial summaries downloaded on March 18, 2014. VITTA is optimized to work with recent versions of Google Chrome.

Assessing Population Representativeness of T2DM Trials



He Z, et al. Assessing the Collective Population Representativeness of Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES. Studies in Health Technology and Informatics;216:569-73.

Assessing Population Representativeness Using EHR data



Weng C, Li Y, Ryan P, et al.. A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Applied Clinical Informatics*. 2014;5(2):463-79.

Target Population from NHANES

- Cross-section continuous national survey of NCHS
- Combined five two-year survey cycle: 2003-2012
- Diabetes? (Dodd *et al.*)
  - told by health professional to have diabetes
  - had an HbA1c measurement
- Excluding Type 1 diabetes patients (Dodd *et al.*)
  - (1) first diagnosed with diabetes before age 30
  - (2) taking insulin
- 2,695 T2DM samples ➔ 15,575,484 T2DM patients in the U.S. national population

He Z, Wang S, Borhanian E, and Weng C. (2015) Assessing the Collective Population Representativeness of Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES. Studies in Health Technology and Informatics. 216:569-73.

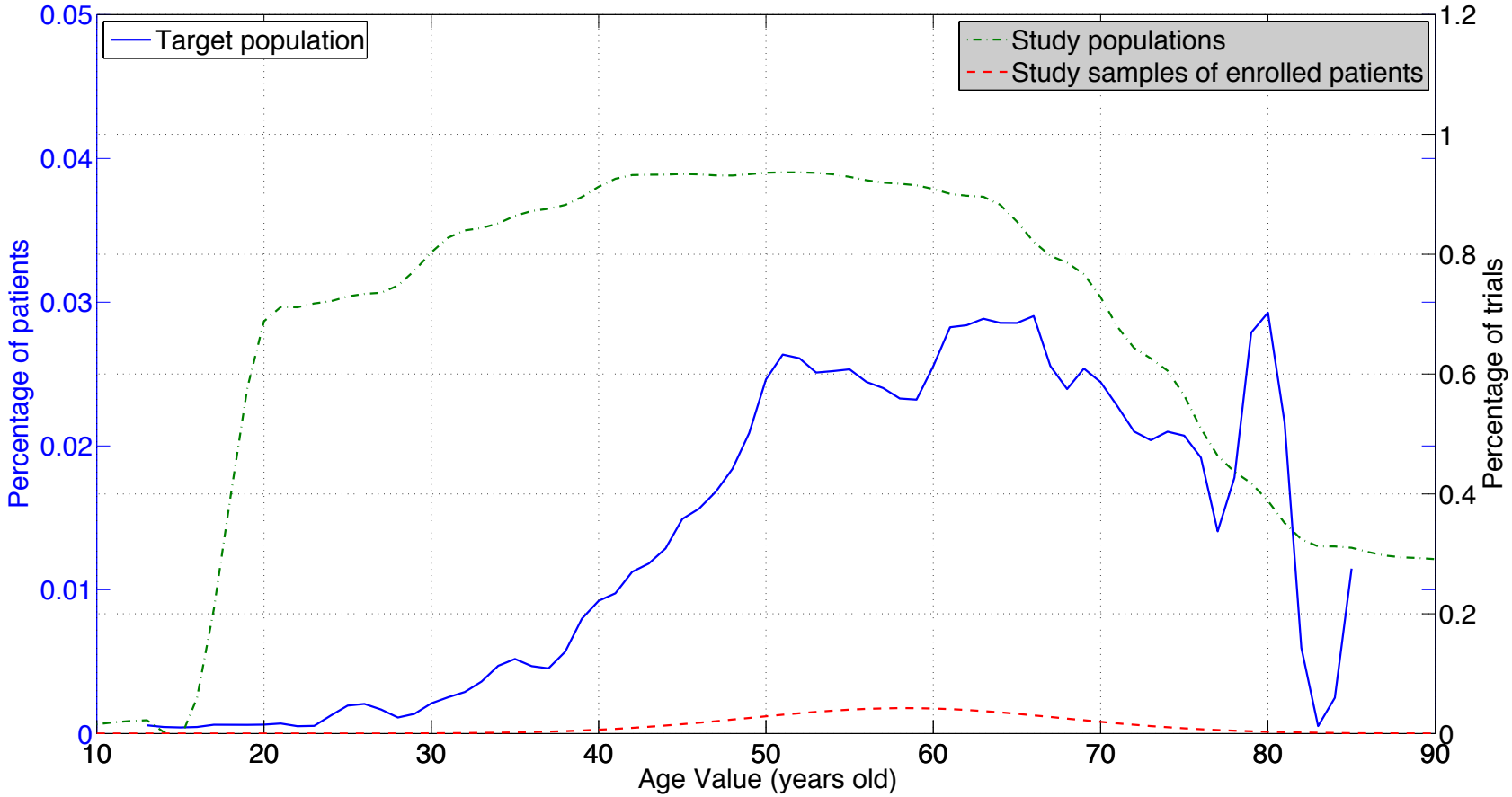
Study Sample of T2DM Trials

- Out of 2,731 T2DM trials between 2003 and 2012, 531 reported their baseline characteristics of enrolled participants in ClinicalTrials.gov.
- # of trials that reported mean and SD of age, HbA1c, and BMI of their enrolled participants are 389, 137, and 108, respectively.

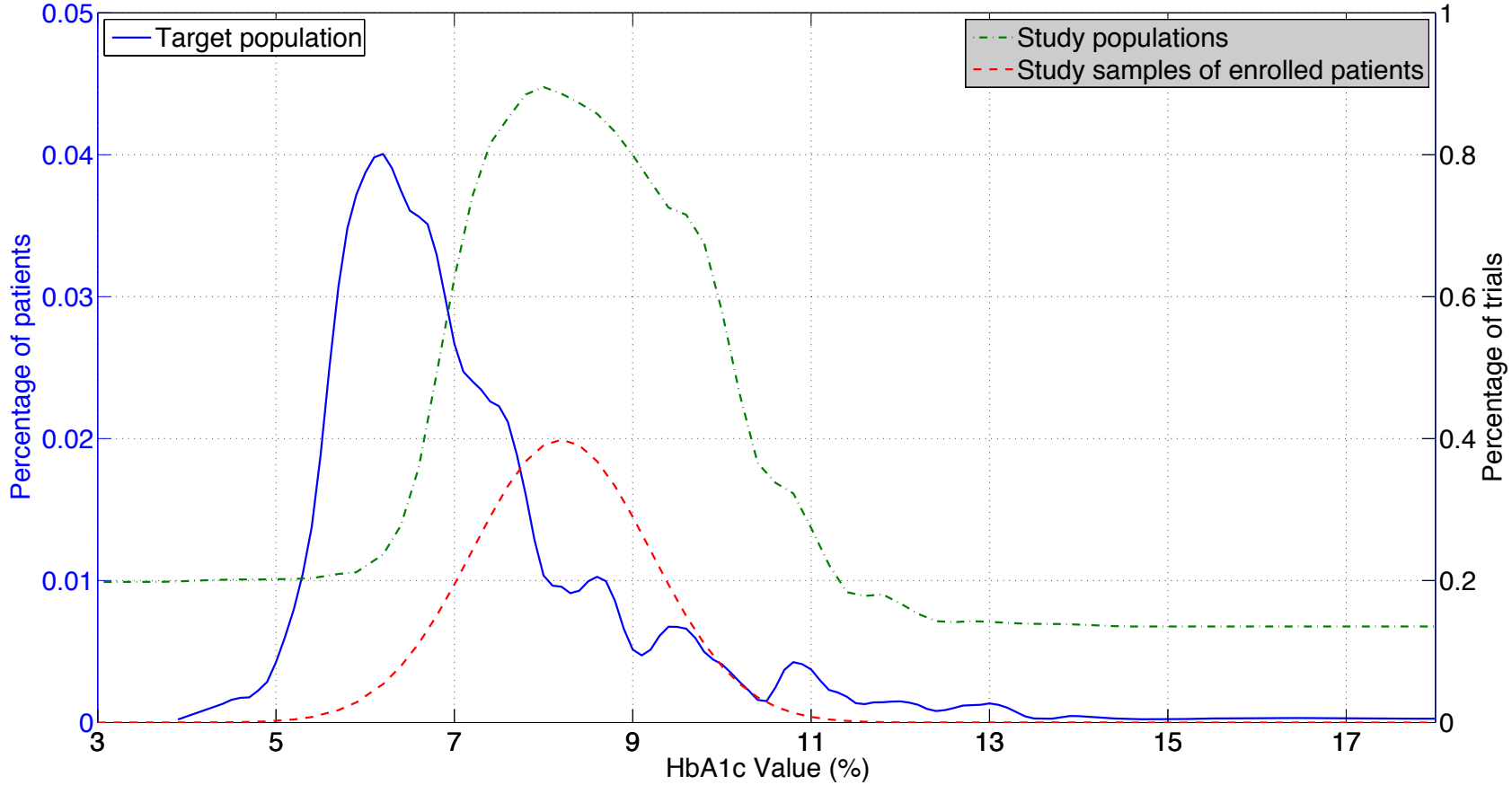
$$\boldsymbol{Weighted_mean} = \frac{\Sigma_{i=1}^{T}(mean_i * number_participants_i)}{\Sigma_{i=1}^{T} number_participants_i}$$

$$\boldsymbol{Weighted_SD} = \sqrt{\frac{\sum_{i=1}^{T}(SD_i^2 * (number_participants_i - 1))}{\sum_{i=1}^{T}(number_participants_i - 1)}}$$

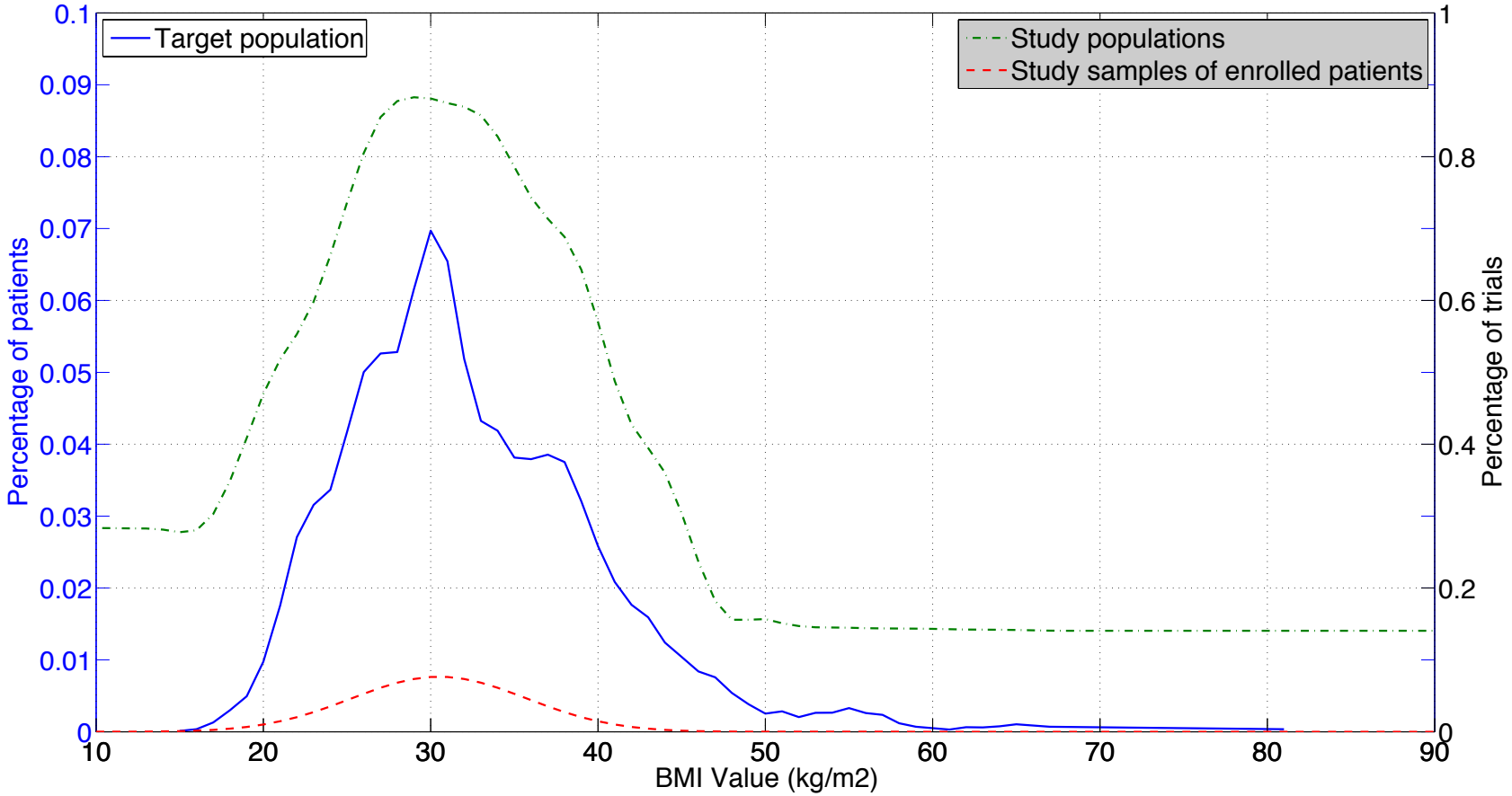
Visualization of Populations (Age)



Visualization of Populations (HbA1c)



Visualization of Populations (BMI)



Generalizability Index for Study Traits

$$GIST = \sum_{i=1}^{N} \frac{\sum_{j=1}^{T} I([i_{low}, i_{high}] \subset w_j)}{T} * \frac{\sum_{k=1}^{P} I(i_{low} \le y_k < i_{high})}{P}$$

(Weng *et al.* 2014)

A set of trials

Trial 1: Age between 20-50
Trial 2: Age >= 20

Real World Patients

10 patients < 20
60 patients 20-50
30 patients > 50

Age interval	Percentage of trials		Percentage of patients	
0 - 20	0%	x	10%	= 0 +
20 - 50	100%	x	60%	= 0.6 +
50 - inf	50%	x	30%	= 0.15
				0.75

Weng C, et al. A Distribution-based Method for Assessing the Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records *Applied Clinical Informatics*. 2014;5(2):463-79.

GIST Scores for Each Phase

GIST	Variable		
Phase	***Age***	***HbA1c***	***BMI***
I	0.60 (N=368)	0.55 (N=141)	0.64 (N=204)
II	0.77 (N=517)	0.45 (N=244)	0.71 (N=194)
III	0.87 (N=766)	0.38 (N=438)	0.77 (N=356)
VI	0.80 (N=484)	0.42 (N=306)	0.69 (N=194)
All	0.77 (N=2702)	0.44 (N=1463)	0.69 (N=1274)

He Z, et al. (2015) Assessing the Collective Population Representativeness of Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES. Studies in Health Technology and Informatics;216:569-73.

Multivariate GIST score

GIST

$$GIST = \sum_{i=1}^{N} \frac{\sum_{j=1}^{T} I([i_{low}, i_{high}] \subset w_j)}{T} * \frac{\sum_{k=1}^{P} I(i_{low} \le y_k < i_{high})}{P}$$

mGIST

$$GIST = \sum_{n=1}^{\boxed{N = \prod_{c=1}^{C} N_c}} \frac{\sum_{j=1}^{T} \boxed{\prod_{c=1}^{C} \prod_{i_c=1}^{N_c} I(i_{c,low} \le w_{j,c} < i_{c,high})}}{T} * \frac{\sum_{k=1}^{P} \boxed{\prod_{c=1}^{C} \prod_{i_c=1}^{N_c} I(i_{c,low} \le y_{k,c} < i_{c,high})}}{P}$$

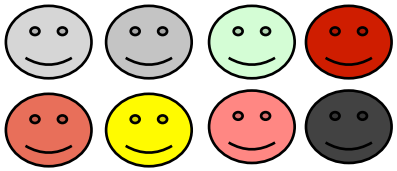
He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

mGIST Scores of T2DM Trials

Trial Characteristic	Number of Trials	GIST Score
Study Type		
Interventional	2731	0.43
Observational	423	0.69
Start Date		
01/2003 – 12/2004	333	0.52
01/2005 – 12/2006	477	0.51
01/2007 – 12/2008	743	0.47
01/2009 – 12/2010	853	0.45
01/2011 – 12/2012	752	0.44

He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

Multivariate Underrepresented Subgroup Characterization (MAGIC)



Until less than 10% well represented patients can be categorized by underrepresented subgroups

Subgroup categorization

well-represented

60%

Subgroup 1.1
age (64,85]
HbA1c (6.4,18]
BMI (25,30]
Gender: female

40%

Subgroup 2.1
age (64,85]
HbA1c (6.4,7.2]
BMI (25,30]
Gender: female

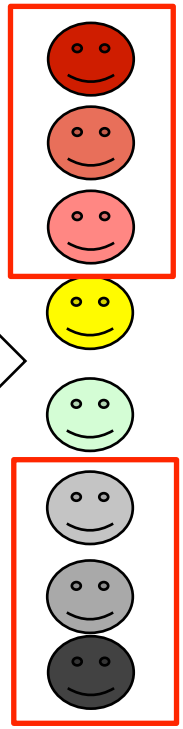
Subgroup 2.2
age (64,85]
HbA1c (7.2,8.4]
BMI (25,30]
Gender: female

Subgroup 2.3
age (64,85]
HbA1c (5.7,6.4]
BMI (25,30]
Gender: female

Subgroup 1.2
age (64,85]
HbA1c (5.7,6.4]
BMI (30,82]
Gender: male

.....

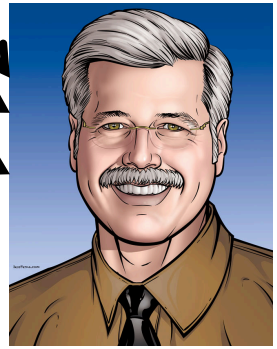
underrepresented



Top 10 Underrepresented Subgroups in T2DM Trials

Combinations of value intervals				Proportion
Age	HbA1c (%)	BMI (kg/m^2)	Gender	
(64, 85]	(5.7, 6.4]	(30, 39]	Female	8.6%
(64, 85]	(5.7, 6.4]	(25, 30]	Female	8.3%
(64, 85]	(5.7, 6.4]	(25, 30]	Male	8.1%
(64, 85]	(6.4, 7.2]	(18.5, 25]	Female	6.5%
(64, 85]	(6.4, 7.2]	(30, 39]	Female	6.3%
(64, 85]	(6.4, 7.2]	(25, 30]	Male	6.1%
(64, 85]	(5.7, 6.4]	(18.5, 25]	Male	5.9%
(64, 85]	(3.8, 5.7]	(25, 30]	Female	5.2%
(64, 85]	(6.4, 7.2]	(25, 30]	Female	4.8%

Elderly pre-diabetes or well-controlled female



Elderly pre-diabetes or well-controlled female

Pictures: Jess Perna (http://jessperna.com/budget_portraits.html)

He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

Top 10 Well Represented Subgroups in T2DM Trials

Combinations of value intervals				Proportion
Age	HbA1c (%)	BMI (kg/m^2)	Gender	
(44, 64]	(7.2, 8.6]	(30, 39]	Male	22.5%
(44, 64]	(7.2, 8.6]	(30, 39]	Female	13.8%
(44, 64]	(7.2, 8.6]	(25, 30]	Male	10.1%
(44, 64]	(8.6, 9.7]	(30, 39]	Male	8.3%
(44, 64]	(8.6, 9.7]	(30, 39]	Female	5.6%
(44, 64]	(6.4, 7.2]	(30, 39]	Female	4.2%
(44, 64]	(7.2, 8.6]	(25, 30]	Female	3.6%
(44, 64]	(8.6, 9.7]	(25, 30]	Female	3.5%
(44, 64]	(6.4, 7.2]	(30, 39]	Female	3.3%
(24, 44]	(8.6, 9.7]	(30, 39]	Male	3.0%

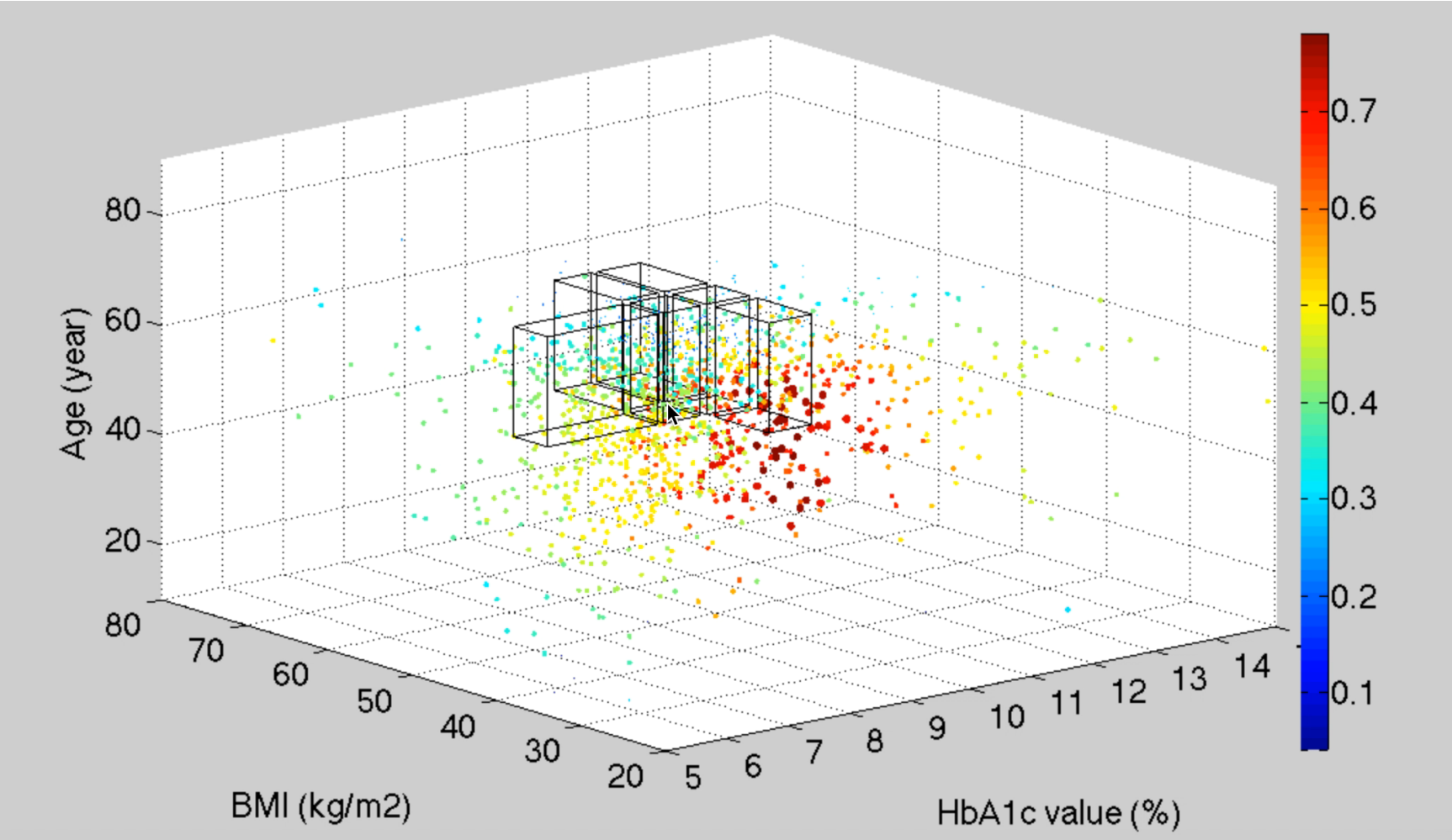
Mid-aged diabetic female

Mid-aged diabetic male

Pictures: Jess Perna (http://jessperna.com/budget_portraits.html)

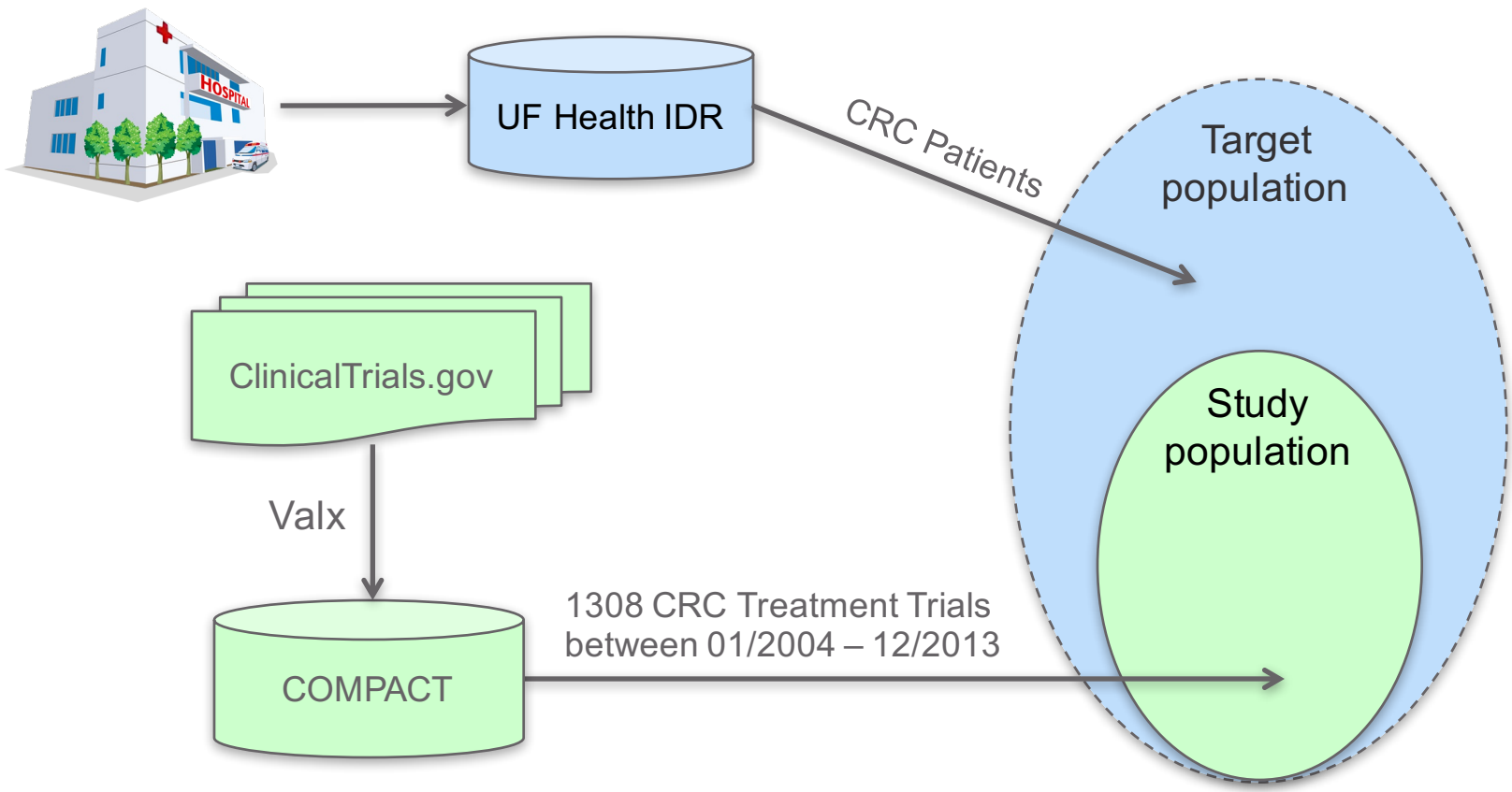
He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

Visualization of Underrepresented Subgroups in T2DM Trials



He Z, Ryan P, Hoxha J, et al. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

Assessing Population Representativeness of Colorectal Cancer Treatment Trials



He Z, Chen Z, George TJ, Lipori G, and Bian J. Assessing the Population Representativeness of Colorectal Cancer Treatment Clinical Trials. Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2016). August 16-20, 2016. Orlando, FL. In press.

GIST Scores of Age and Creatinine

Variable	GIST Score (Number of trials)	
	Trials with the variable	***All the trials***
Age	0.91 (1,308)	0.91 (1,308)
Creatinine	0.78 (422)	0.93 (1,308)
Platelet count	0.88 (332)	0.97 (1,308)
ALT	0.88 (318)	0.97 (1,308)
AST	0.91 (160)	0.99 (1,308)

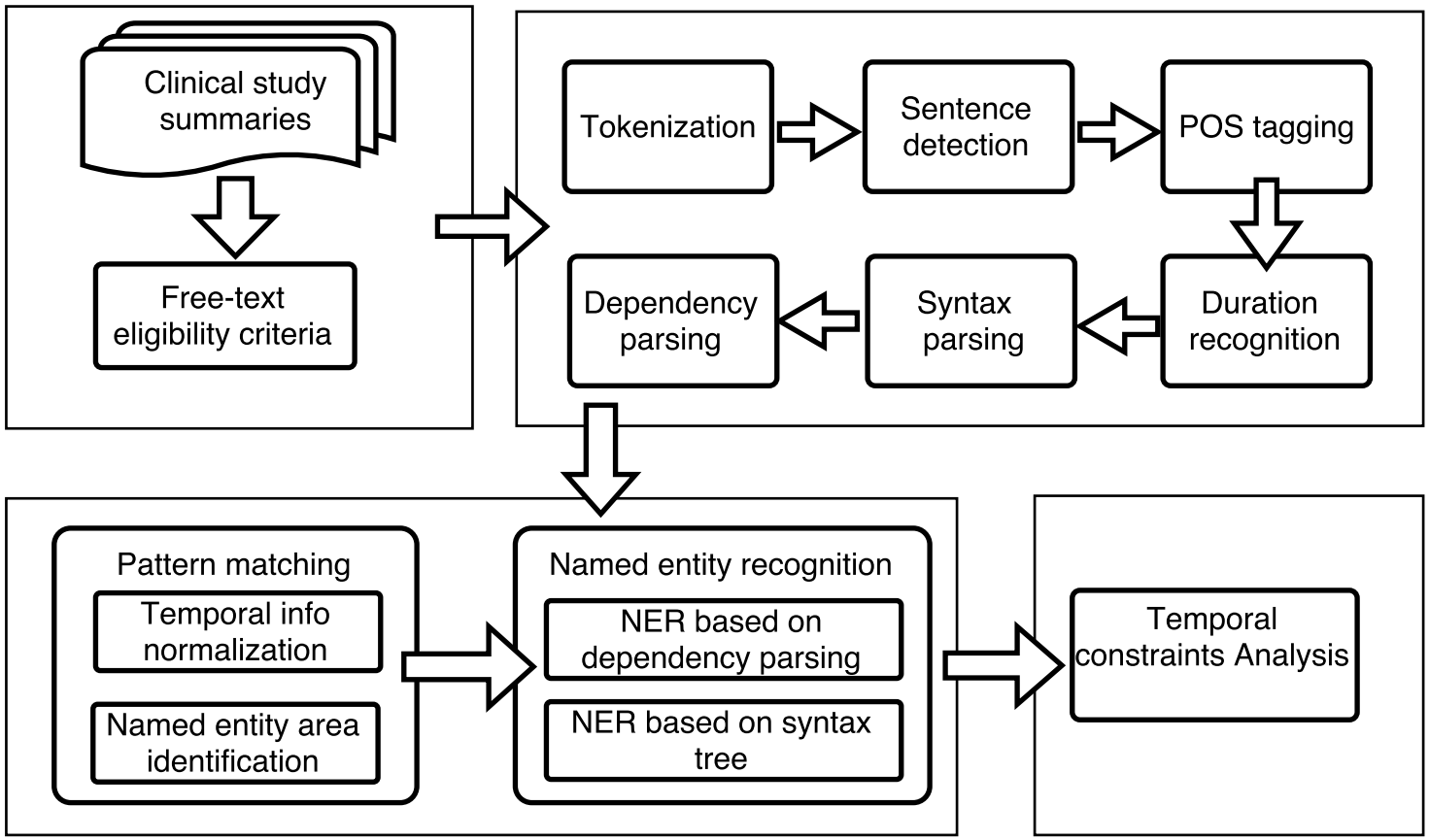
Frequency of Qualitative Variables with Restrictions

- In all 1,308 CRC treatment trials, 636 trials contain one or more keywords in the restrictive qualitative eligibility criteria in Lewis et al.
- Among 100 randomly selected CRC trials, 44 trials had 76 occurrences of the keywords of the restrictive qualitative eligibility criteria.

Qualitative Criteria	Level of Restriction	Number of Trials (%)
No clinically evident congestive heart failure	Moderate	30 (68%)
No myocardial infarction past 12 months	Moderate	17 (39%)
No myocardial infarction past 6 months	Moderate	15 (34%)
No history of congestive heart failure	Strict	5 (11.4%)
No myocardial infarction ever	Strict	2 (5%)
No history of stroke	Strict	2 (5%)

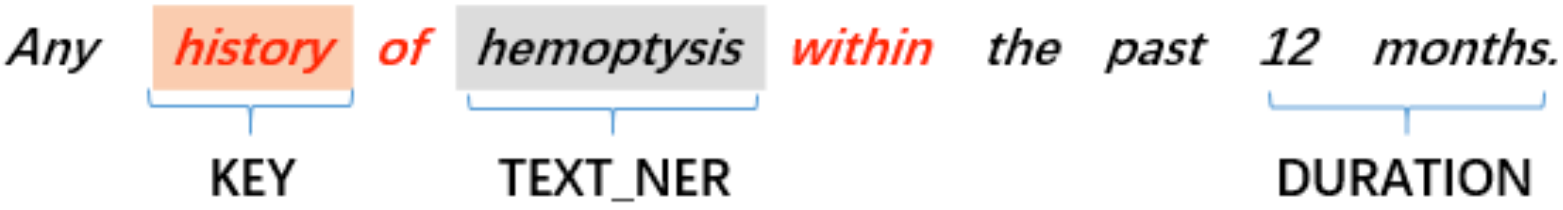
Lewis, J.H., M.L. Kilgore, D.P. Goldman, E.L. Trimble, R. Kaplan, M.J. Montello, M.G. Housman, and J.J. Escarce, Participation of patients 65 years of age or older in cancer clinical trials. J Clin Oncol, 2013. 21(7): pp. 1383-9.

Extracting Qualitative Eligibility Criteria



He, Z., Chen, Z., and Bian, J. Analysis of Temporal Constraints in Qualitative Eligibility Criteria of Cancer Clinical Studies. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016). Under review

Identifying Text Area for NER and Temporal Constraints



He, Z., Chen, Z., and Bian, J. Analysis of Temporal Constraints in Qualitative Eligibility Criteria of Cancer Clinical Studies. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016). Under review

Frequency of Lexical Patterns

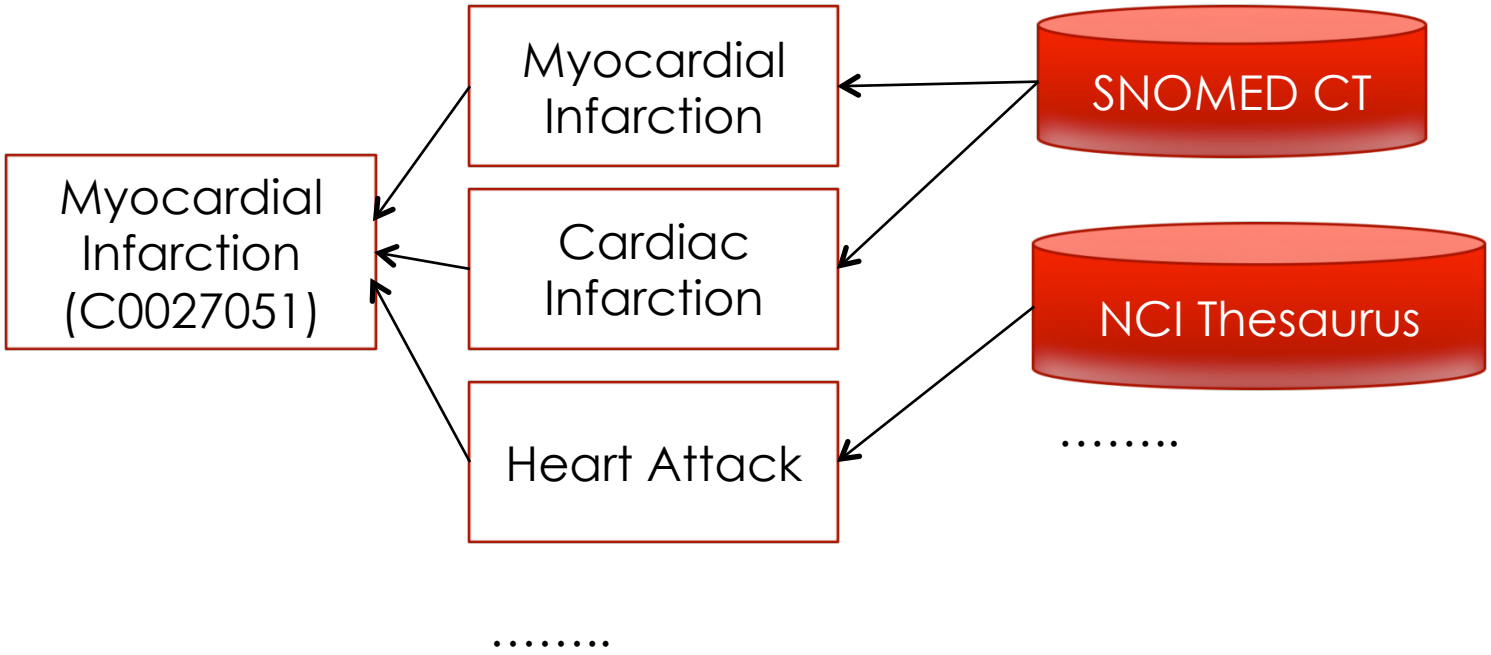
Dataset: 10848 studies on lung, breast, prostate, and colorectal cancer

Pattern name	Overall freq.	Freq. in inclusion criteria	Freq. in exclusion criteria	Sample criterion
WITHIN	10,298	4,501	5,797	No myocardial infarction **within** *the past 6 months*
MORETHAN	5,184	3,998	1,186	Patients must have been off previous anti-androgen therapy for **more than** *4 weeks*.
WITHIN_PT	2,931	1,284	1,647	2 metastases seen on standard imaging **within** *30 days* **prior to** registration
ATLEAST_PT	2,279	2,132	225	**At least** *5 years* since **prior** chemotherapy
MORETHAN_PT	1,439	1,273	166	**More than** *4 weeks* since **prior** immunotherapy

He, Z., Chen, Z., and Bian, J. Analysis of Temporal Constraints in Qualitative Eligibility Criteria of Cancer Clinical Studies. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016). Under review

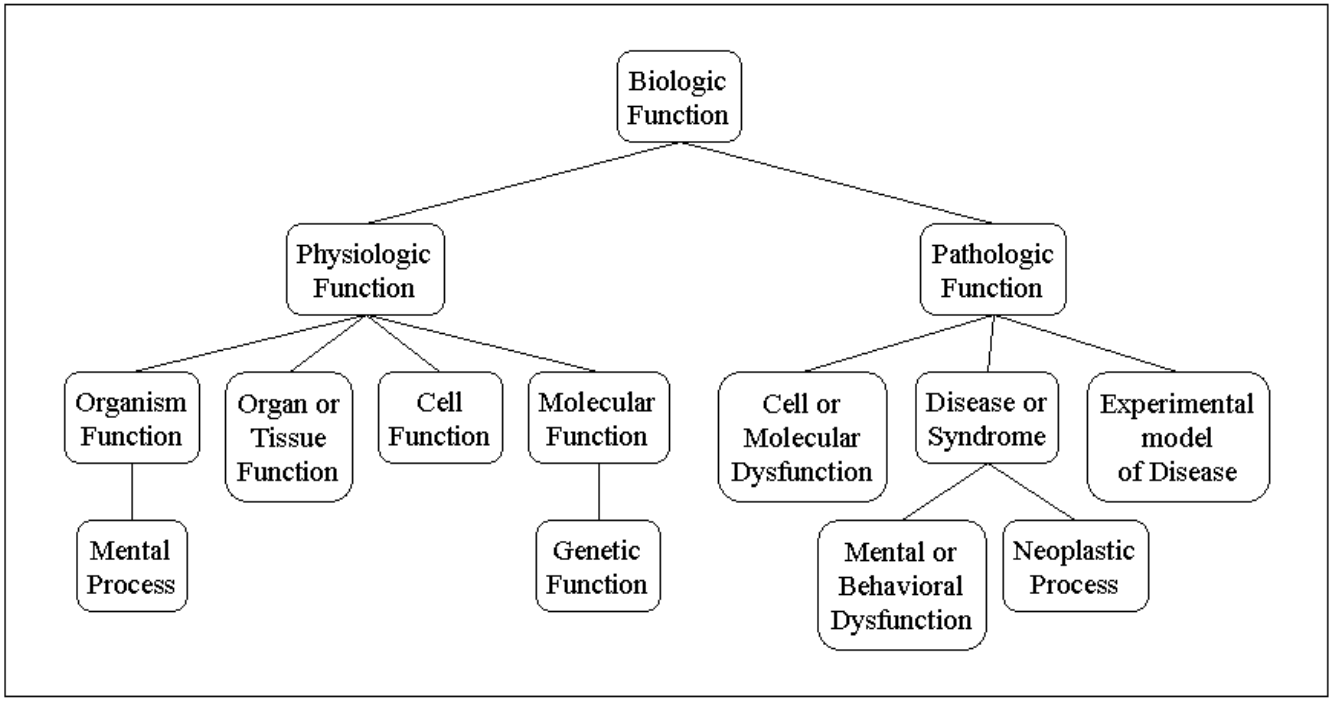
Unified Medical Language System

- The Unified Medical Language System (UMLS)
  - Over 3.1 million medical concepts; More than 9.1 million terms
  - Over 170 source terminologies: ICD, SNOMED CT, LOINC, RxNORM
  - Two-level structure: Metathesaurus & Semantic Network
  - SPECILIST Lexicon and Lexical Tools



Semantic Network & Semantic Types

127 semantic types



A Portion of the UMLS Semantic Network

"Biologic Function" Hierarchy

Semantic Types Included

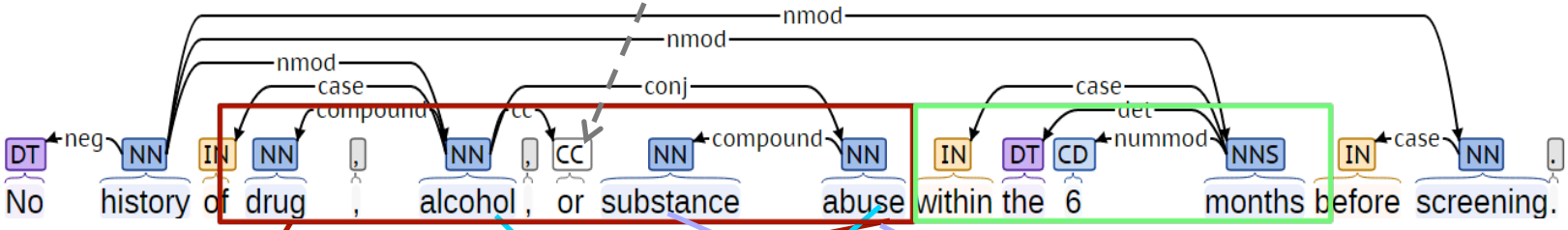
Disorders		Procedures	
TUI	**Semantic Type**	**TUI**	**Semantic Type**
T020	Acquired Abnormality	T191	Neoplastic Process
T190	Anatomical Abnormality	T046	Pathologic Function
T049	Cell or Molecular Dysfunction	T184	Sign or Symptom
T019	Congenital Abnormality	T060	Diagnostic Procedure
T047	Disease or Syndrome	T058	Health Care Activity
T050	Experimental Model of Disease	T059	Laboratory Procedure
T037	Injury or Poisoning	T063	Molecular Biology Research Technique
T048	Mental or Behavioral Dysfunction	T061	Therapeutic or Preventive Procedure
T200	Clinical Drug		

Named Entity Recognition with Dependency Parsing

*"No history of **drug, alcohol, or substance abuse** within 6 months before screening."*

Dependency Parsing by Stanford CoreNLP

Conjunction or



drug abuse

alcohol abuse

substance abuse

UMLS Matching

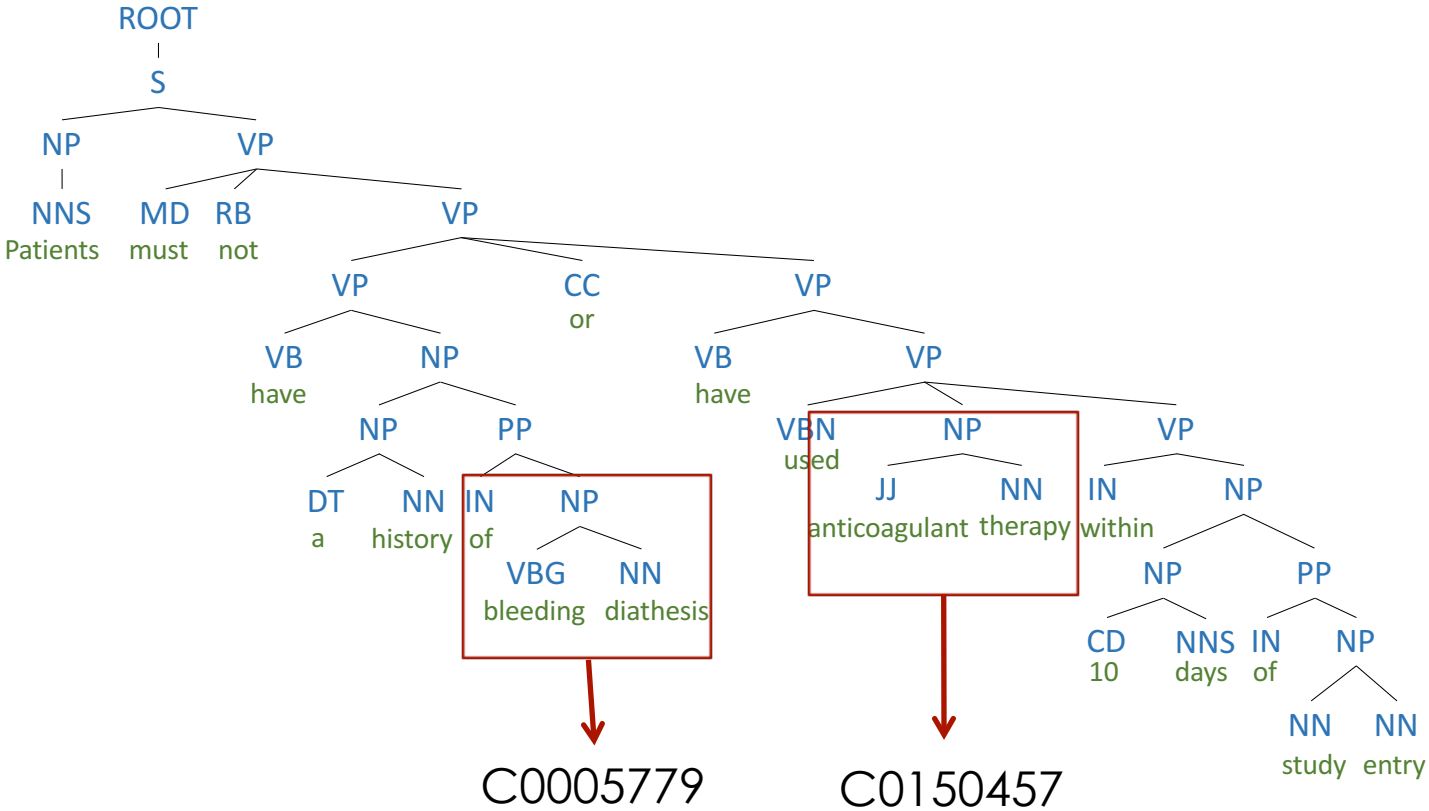
C0013146

C0085762

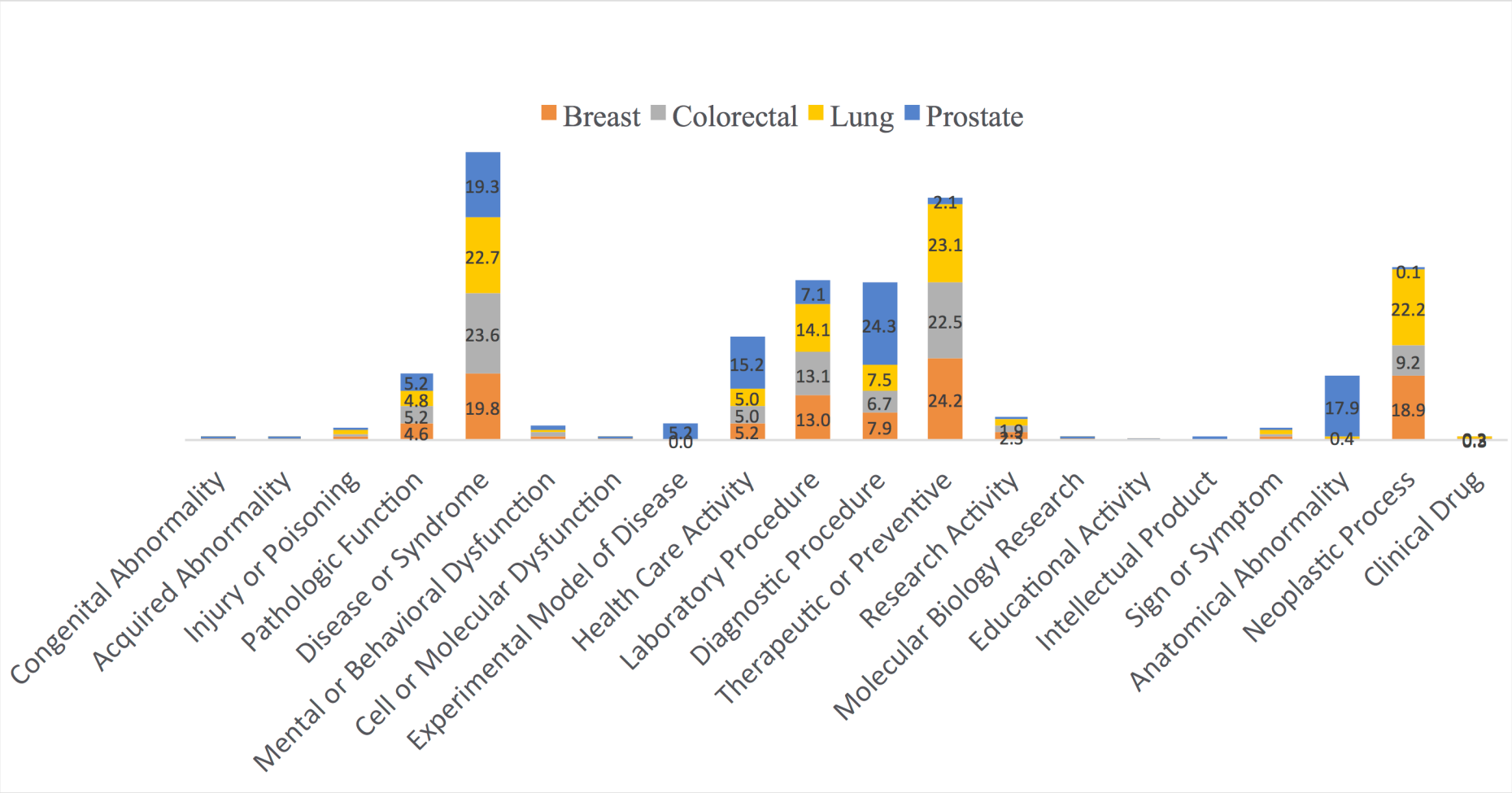
C0740858

Noun Phrases Detection Using Syntactic Tree

*"Patients must not have a history of **bleeding diathesis** or have used **anticoagulant therapy** within 10 days of study entry"*

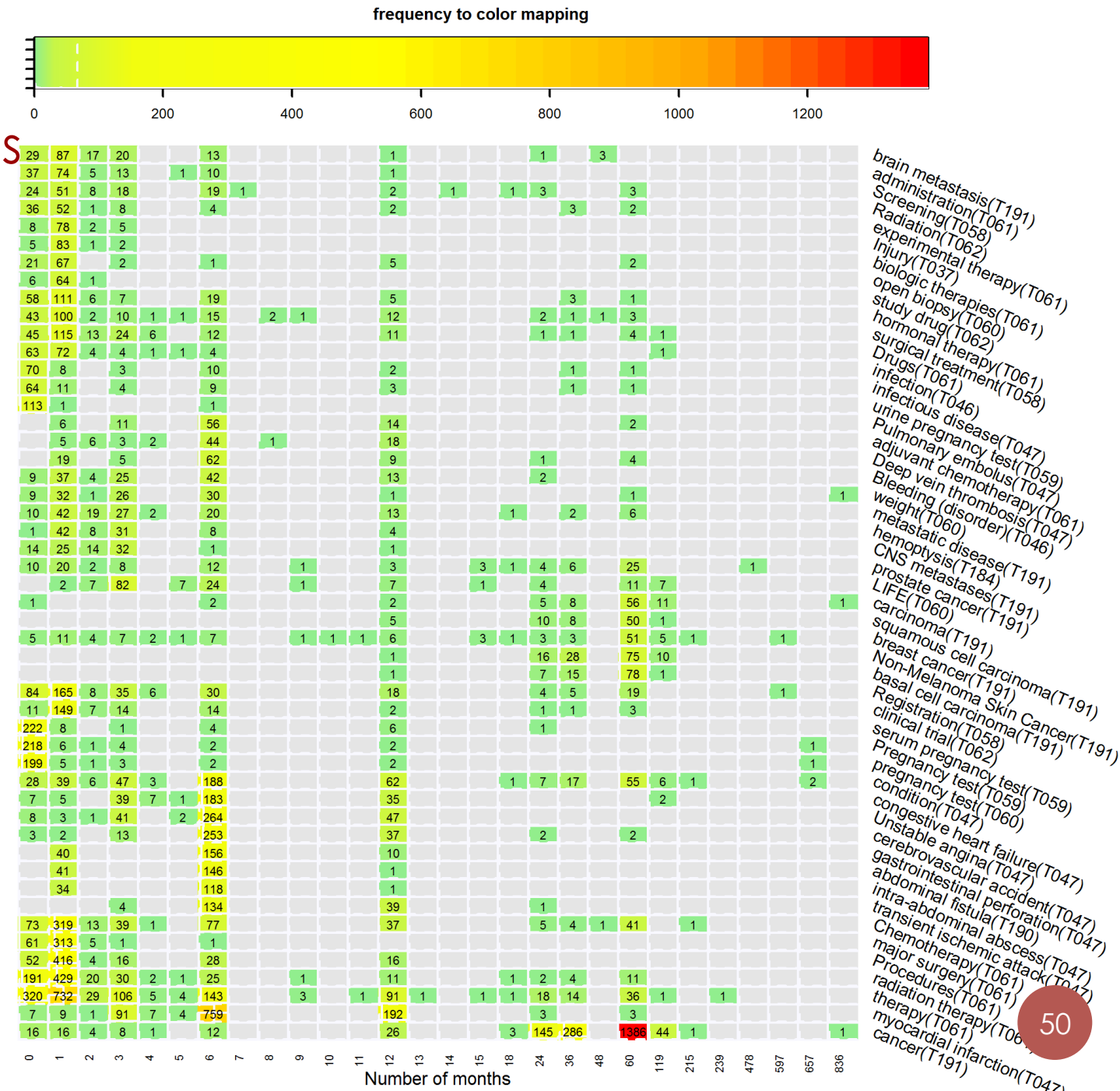


Semantic Types of Eligibility Criteria

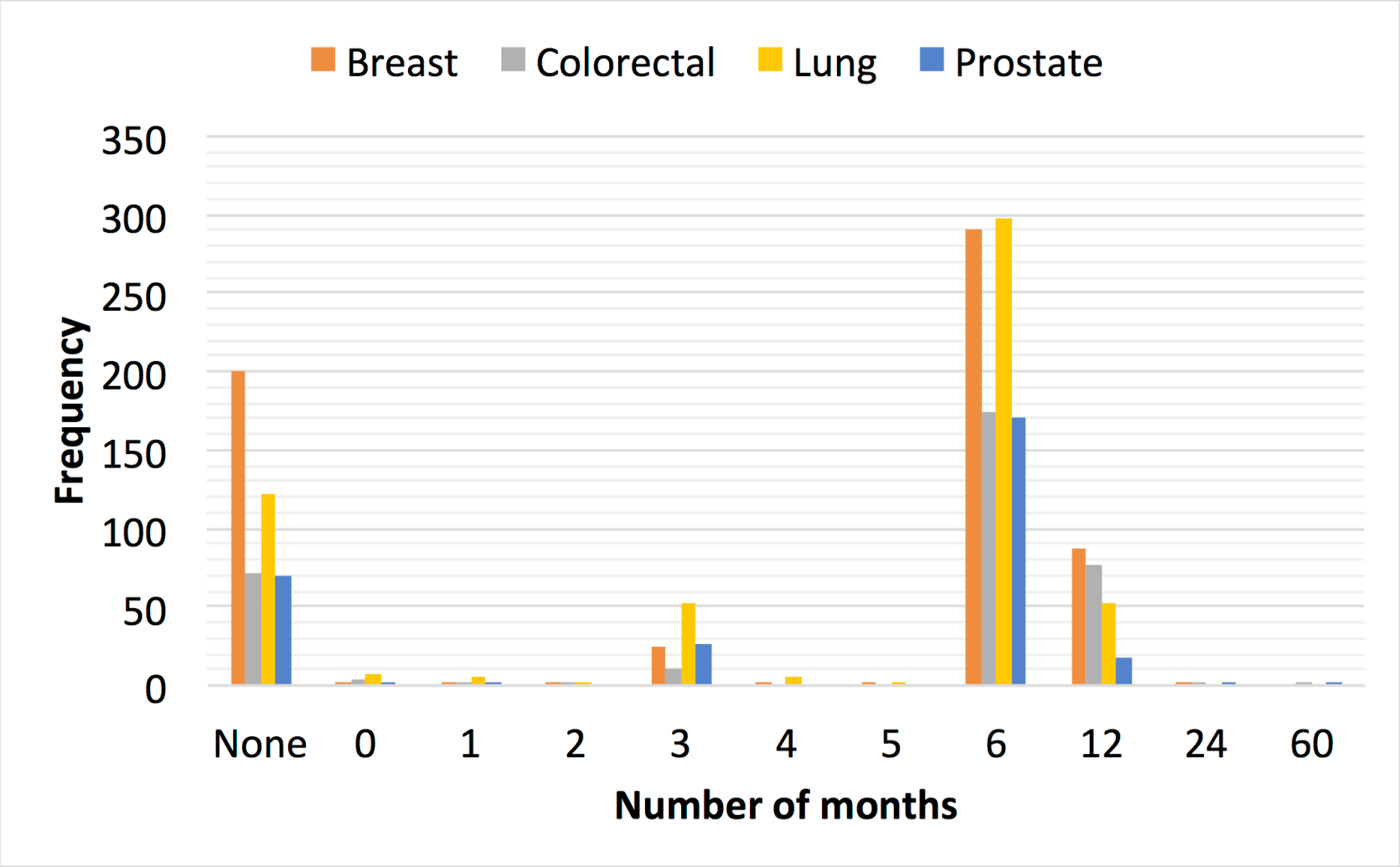


Zhe He, Zhiwei Chen, and Jiang Bian. Analysis of Temporal Constraints in Qualitative Eligibility Criteria of Cancer Clinical Studies. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016). Under review

Frequent Criteria in Cancer Studies



Temporal Constraints on Myocardial Infarction in Cancer Trials



Zhe He, Zhiwei Chen, and Jiang Bian. Analysis of Temporal Constraints in Qualitative Eligibility Criteria of Cancer Clinical Studies. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2016). Under review

VATEC Tool

Visual Analysis Tool of Qualitative Eligibility Criteria of Clinical Studies (VATEC)

Process | Input a clinical trials

Clinical study ID: (NCT00004054)
Number of UMLS terms: 37
Brief title:: Hormone Therapy Plus Radiation Therapy With or Without Combination Chemotherapy in Treating Patients With Prostate Cancer
Overall status: Completed
Study type: Interventional
Phase: Phase 3
Intervention Type: Drug
Intervention Name: bicalutamide
Link to ClinicalTrials.gov: NCT00004054

Disease	type	pattern	duration	month	negation	sty	org_str	cui_str	sentence
Prostate	INCLUSION	None	-1	-1	0	Neoplastic Process	Histologically proven prostate cancer	prostate cancer	Histologically proven prostate cancer at high risk for relapse as determined by either of the following:
Prostate	INCLUSION	None	-1	-1	0	Disease or Syndrome	high risk for relapse	relapse	Histologically proven prostate cancer at high risk for relapse as determined by either of the following:
Prostate	INCLUSION	None	-1	-1	0	Laboratory Procedure	Prostate Specific Antigen	Prostate Specific Antigen	Prostate-specific antigen (PSA) 20-100 ng/mL and Gleason score at least 7 (any T stage)

Analysis Query Builder

Current condition is Breast_cancer

Show feature list | Top 10 Features

Basic options ('All Studies' by default)

Conditions: Breast_cancer (e.g. 'Lung cancer')

Gender: (Both: 30.12%, Male: 0.21%, Female: 69.58%) | Age: Minimum age: Maximum age:

Start date after: Month: Year: (e.g., 2005)
Start date before: Month: Year: (e.g., 2005)

Advanced Options ('All Studies' by default)

Study Types:

- Interventional (80.86%)
- Observational (18.51%)
- Observational [Patient Registry] (0.49%)
- Expanded Access (0.05%)

Intervention types:

- Drug (56.24%)
- Procedure (8.06%)
- Biological (7.33%)
- Device (4.31%)
- Behavioral (7.80%)
- Dietary Supplement (1.92%)
- Genetic (1.51%)
- Radiation (4.05%)
- Other (8.78%)

Status:

- Open Studies
- Recruiting (19.21%)
- Not yet recruiting (1.08%)
- Closed Studies
- Active, not recruiting (20.15%)
- Completed (45.35%)
- Withdrawn (2.76%)
- Suspended (0.47%)
- Terminated (10.02%)
- Enrolling by invitation (0.82%)

Sponsor Types:

- NIH (4.56%)
- Industry (20.64%)
- Other U.S. Federal Agency (0.21%)
- Other (74.49%)

Phases:

Check multiple phases to stratify the distributions of studies or enrollment by phases

- Phase 0 (1.24%)
- Phase 1 (10.72%)
- Phase 2 (30.35%)
- Phase 3 (12.54%)
- Phase 4 (2.81%)
- Not applicable or not specified (35.24%)

Study design - intervention model:

Study design - allocation:

Study design - time perspective:

Pattern:

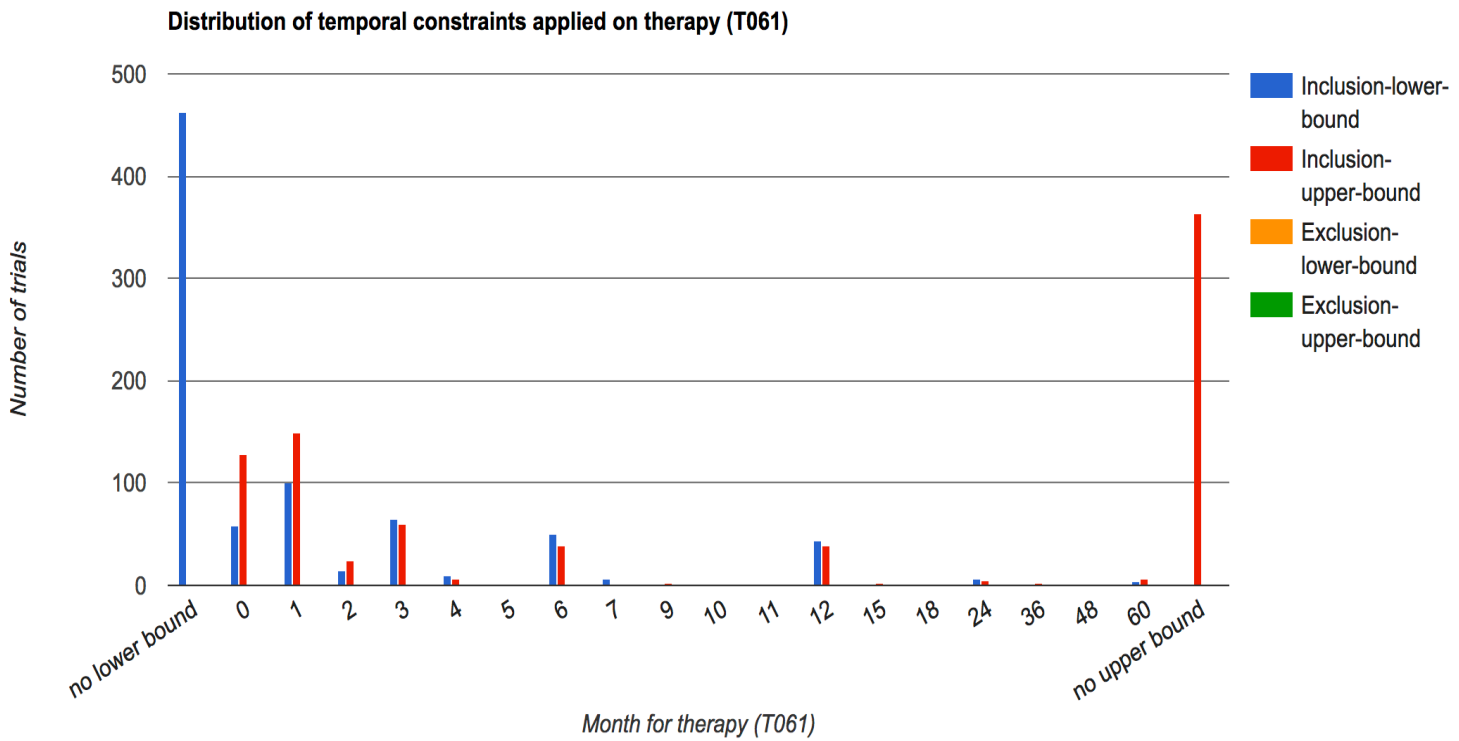
Analysis Results

Choose another condition | Export output in CSV

Condition: Breast_cancer
Feature: therapy (T061)

Number of occurrence: 7136
Number of trials using therapy (T061) with a temporal constraint: 828
Number of trials using therapy (T061) without a temporal constraint: 914
Number of Clinical Trials: 1742

Trial characteristics:
study_type: Interventional;
intervention_type: Drug;



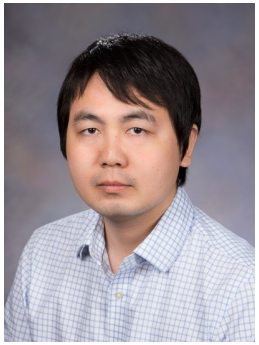
Limitations and Take Home Message

- Lack of population representativeness reduces the cost-effective ratio of clinical trials
- Effective use of the colossal health data repositories is promising to improve the transparency of the design issues of clinical studies

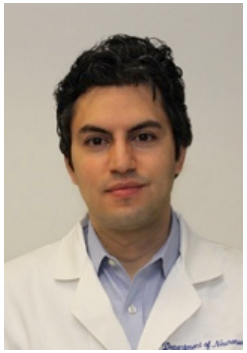
Collaborators and Students

Henry Carretta,
PhD, FSU

Neil Charness,
PhD, FSU



Jiang Bian,
PhD, UF



Elias Sayour,
MD, PhD, UF

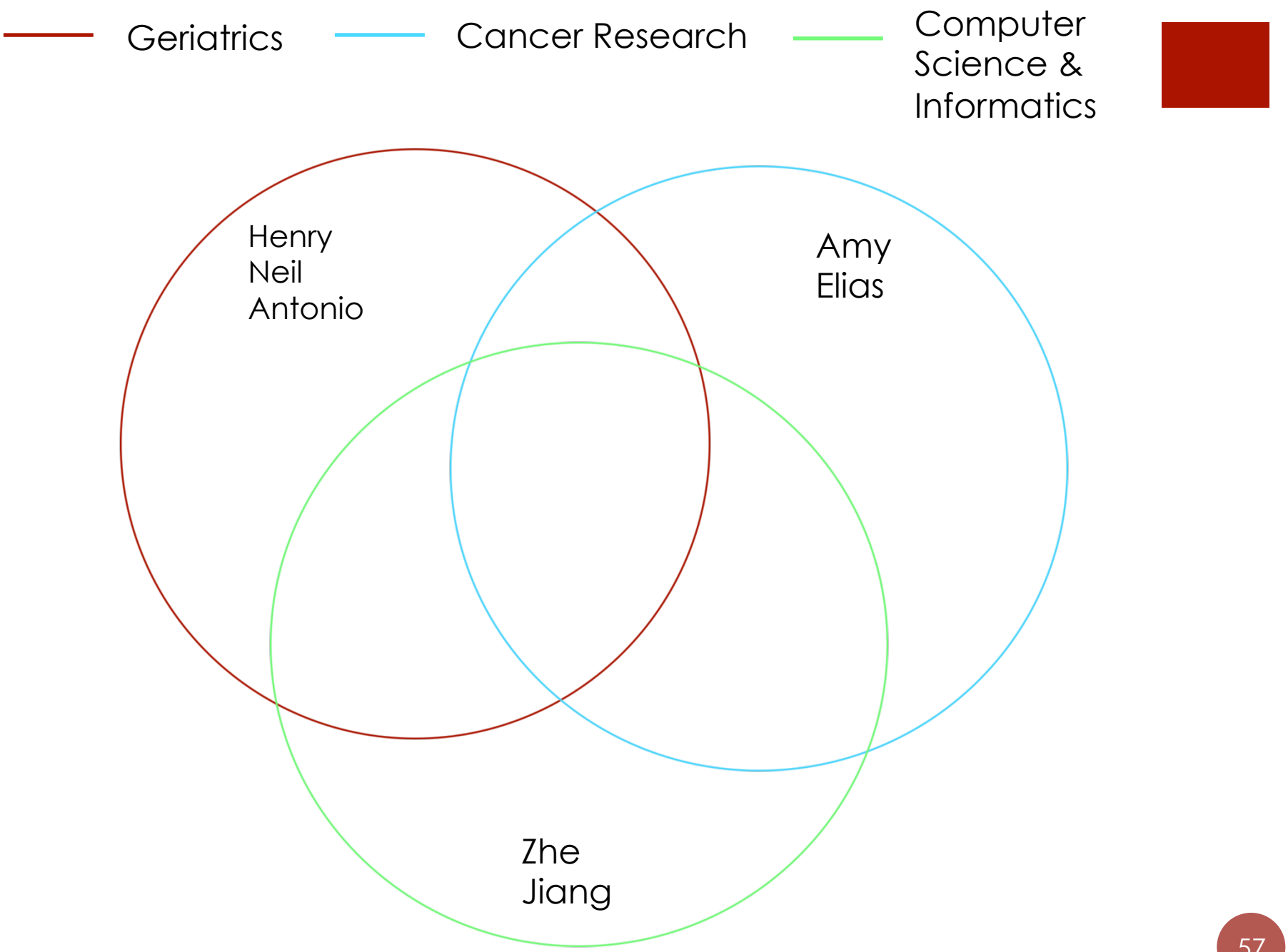
Aisha Langford,
PhD, MPH, NYU

Keith Lohse,
PhD, Auburn

Antonio Terraciano,
PhD, FSU

Qing-Xiang, Amy, Sang
PhD, FSU

Students: Zhiwei Chen (MS Student), Jiwon Lee (PhD Student), Sarah Mixon (BS Student)


Geriatrics
Cancer Research
Computer Science & Informatics
Henry
Neil
Antonio
Amy
Elias
Zhe
Jiang

Acknowledgments

The development of the COMPACT database and the VITTA tool was supported by

- U.S. National Library of Medicine Grant R01LM009886 (PI: Weng)
- U.S. National Center for Advancing Translational Science (NCATS) Award UL1TR000040 (PI: Ginsberg)

This work was supported by:

- Amazon Web Service in Education Research Grant Award (PI: He)
- Planning Grant of FSU Institute for Successful Longevity (PI: He)
- Seed Grant of Florida State University Council for Research and Creativity (PI: He)
- NIH/NCATS Clinical and Translational Science Award UL1TR001427 (PIs: Nelson & Shenkman)

References

1. Rothwell, P.M., External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*, 2005. 365(9453): pp. 82-93.
2. Schoenmake,r N., Van Gool, W.A. The age gap between patients in clinical studies and in the general population: a pitfall for dementia research. Lancet Neurol. 2004;3(10):627-30. PMID: 15380160
3. Lewis, J.H., M.L. Kilgore, D.P. Goldman, E.L. Trimble, R. Kaplan, M.J. Montello, M.G. Housman, and J.J. Escarce, Participation of patients 65 years of age or older in cancer clinical trials. J Clin Oncol, 2013. 21(7): pp. 1383-9.
4. Filion, M., G. Forget, O. Brochu, L. Provencher, C. Desbiens, C. Doyle, B. Poirier, M. DuRocher, S. Camden, and J. Lemieux, Eligibility criteria in randomized phase II and III adjuvant and neoadjuvant breast cancer trials: not a significant barrier to enrollment. *Clin Trials*, 2012. 9(5): pp. 652-659.
5. Lee, P.Y., K.P. Alexander, B.G. Hammill, S.K. Pasquali, and E.D. Peterson, Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA*, 2001. 286(6): pp. 708-13.
6. Weng, C., Y. Li, P. Ryan, Y. Zhang, J. Gao, F. Liu, J.T. Bigger, and G. Hripcsak, A Distribution-based Method for Assessing The Differences between Clinical Trial Target Populations and Patient Populations in Electronic Health Records. *Applied Clinical Informatics*, 2014. 5(2): pp. 463-479.
7. He, Z., S. Wang, E. Bornanian, and C. Weng, Assessing the Population Representativeness of Type 2 Diabetes Trials by Combining Public Data from ClinicalTrials.gov and NHANES. *Stud Health Technol Inform*, 2015. 2015(216): pp. 569-73.
8. He, Z., P. Ryan, J. Hoxha, S. Wang, S. Carini, I. Sim, and C. Weng. Multivariate analysis of the population representativeness of related clinical studies. 2016;60:66-76.

LIS 5916 Data Analytics

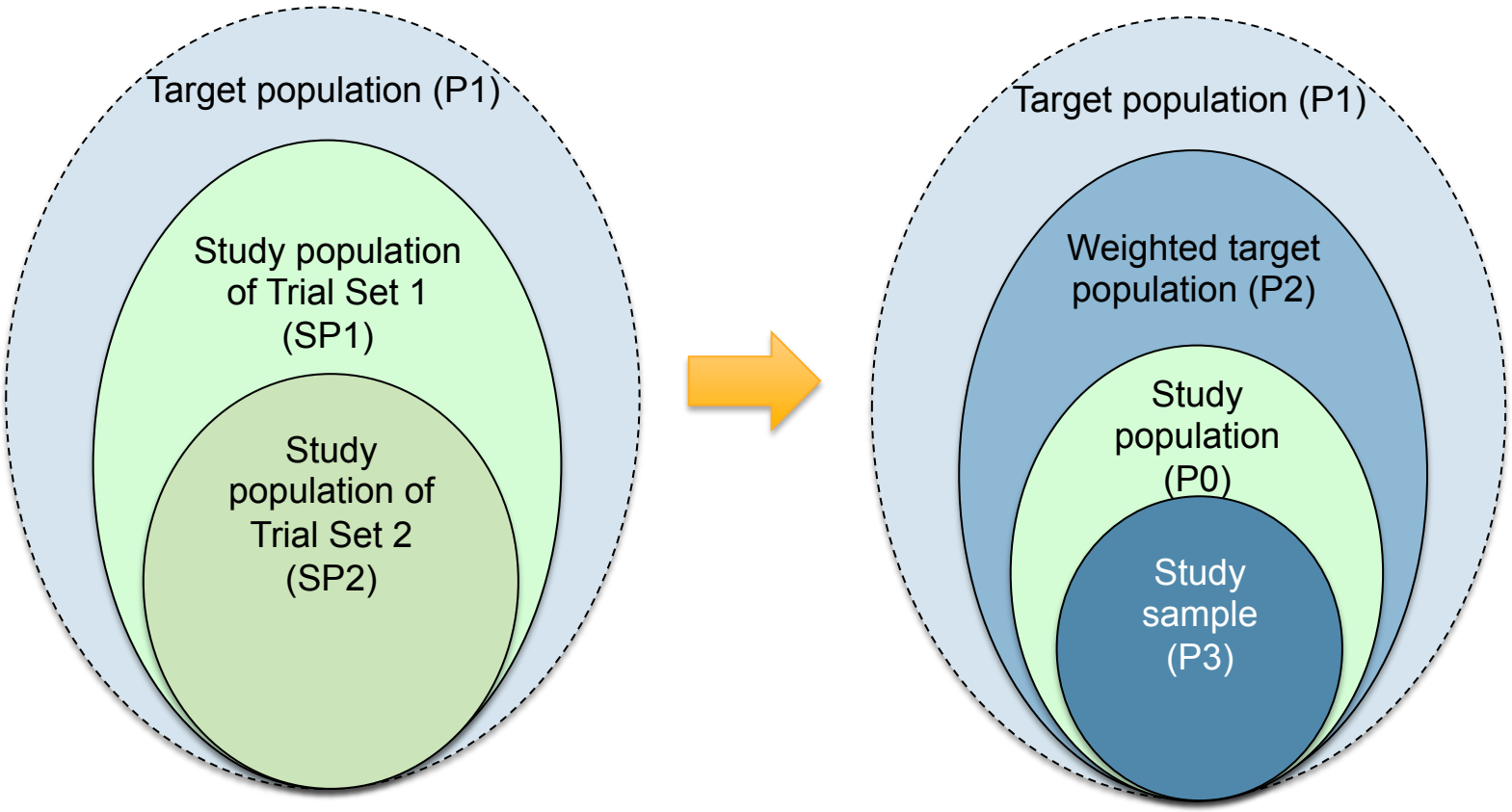
- Tuesday 8 pm – 10 pm on BlackBoard Collaborate
- This course provides an introduction to data analytics
- Students will learn basic concepts and tools for data analytics, including:
  - data sources
  - data cleaning tools and methods
  - clustering analysis
  - association rule mining
  - predictive modeling
  - text mining

- Students will also learn how data mining approaches are used in various application areas, particularly in healthcare.

Thank you!

Dr. Zhe He
School of Information
Florida State University
zhe@fsu.edu
http://www.zhehe.info

Methods for Evaluating GIST



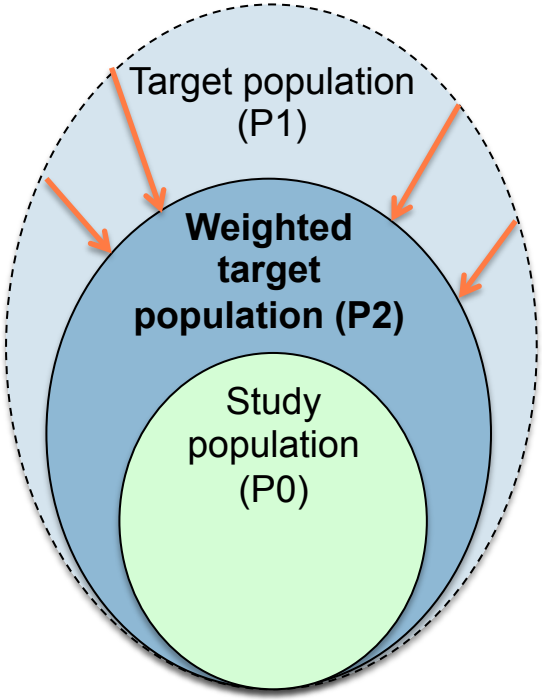
Test if GIST(SP1, P1) > GIST (SP2, P1)

Test if GIST(P0, P1) < GIST (P0, P2) < GIST (P0, P3)

He et al. (2015) Simulation-based Evaluation of the Generalizability Index for Study Traits. AMIA Annu Symp Proc. 2015. In press.

SP-Weighted Target Population (P2)

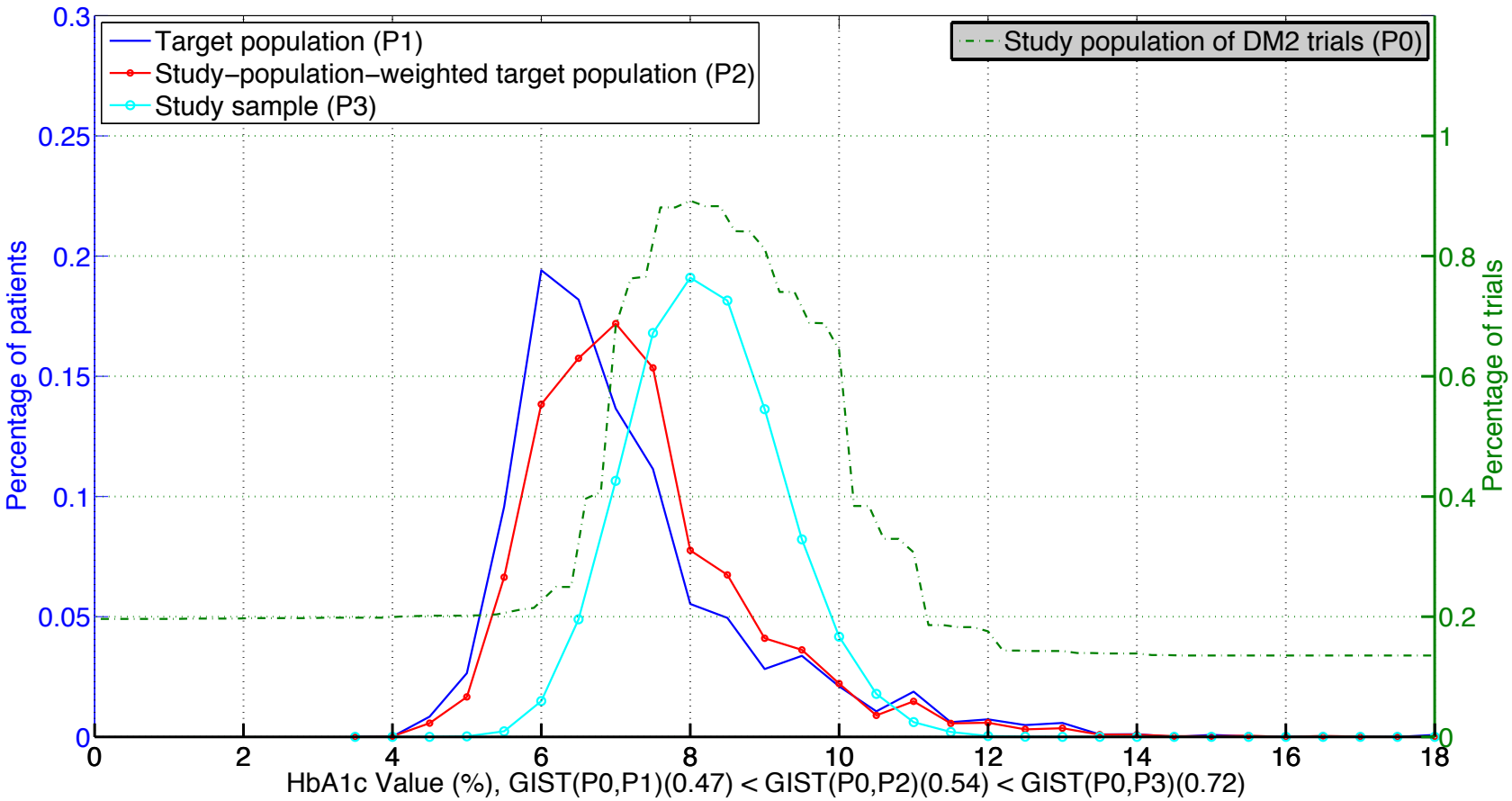
- Average of normalized NHANES sample weight and percentage of trials
- Oversampled patients who are considered by more trials and under-sampled patients who are considered by fewer trials



Evaluation of the GIST score

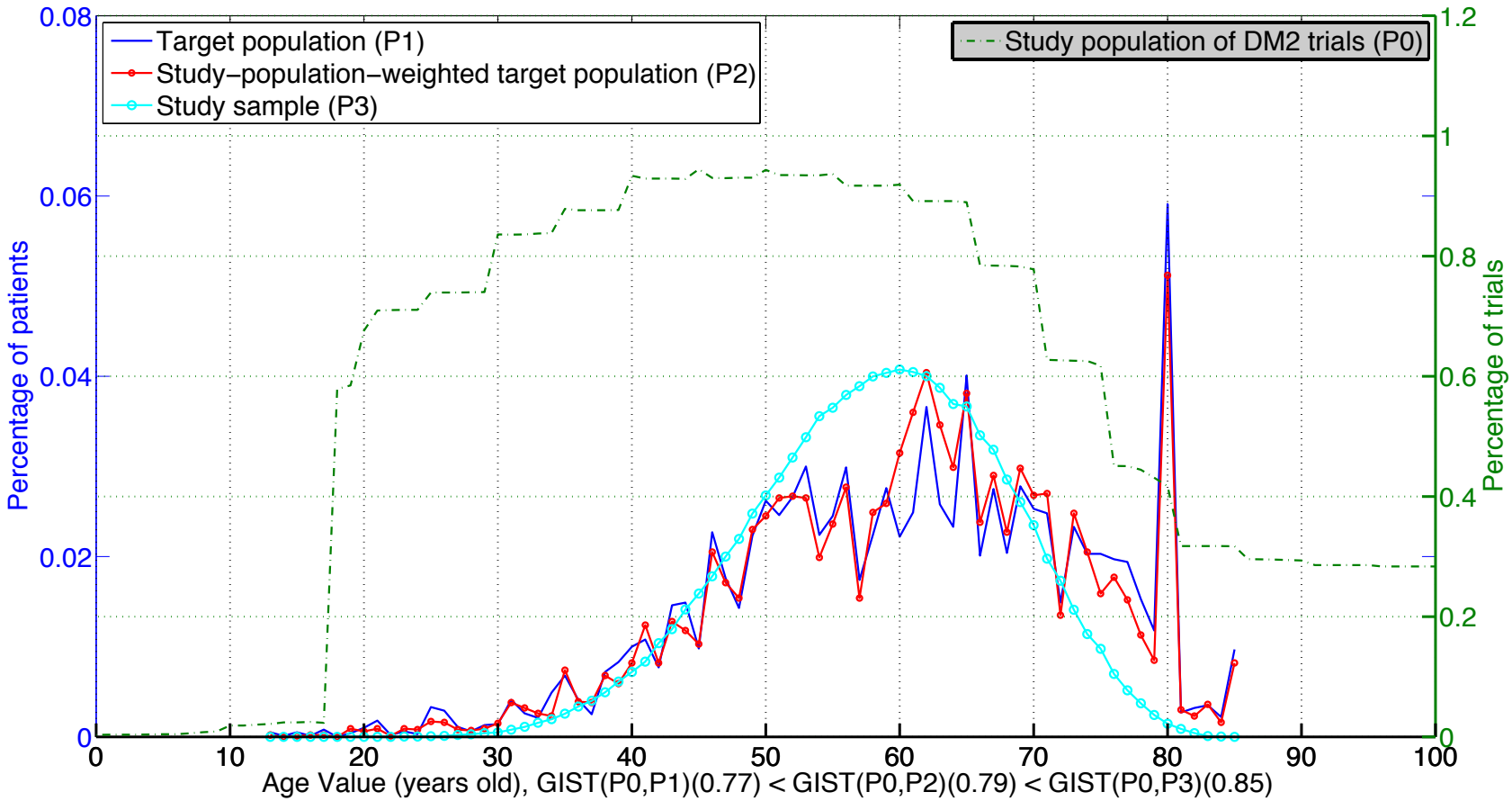
- Generate a random sample of 10000 patients of P1, P2, P3
- Calculate GIST(P0,P1), GIST(P0,P2), GIST(P0,P3)
- Test if GIST(P0,P1) < GIST(P0,P2) < GIST(P0,P3)
- Simulate a hundred times for each variable

Visualization of Simulation (HbA1c)

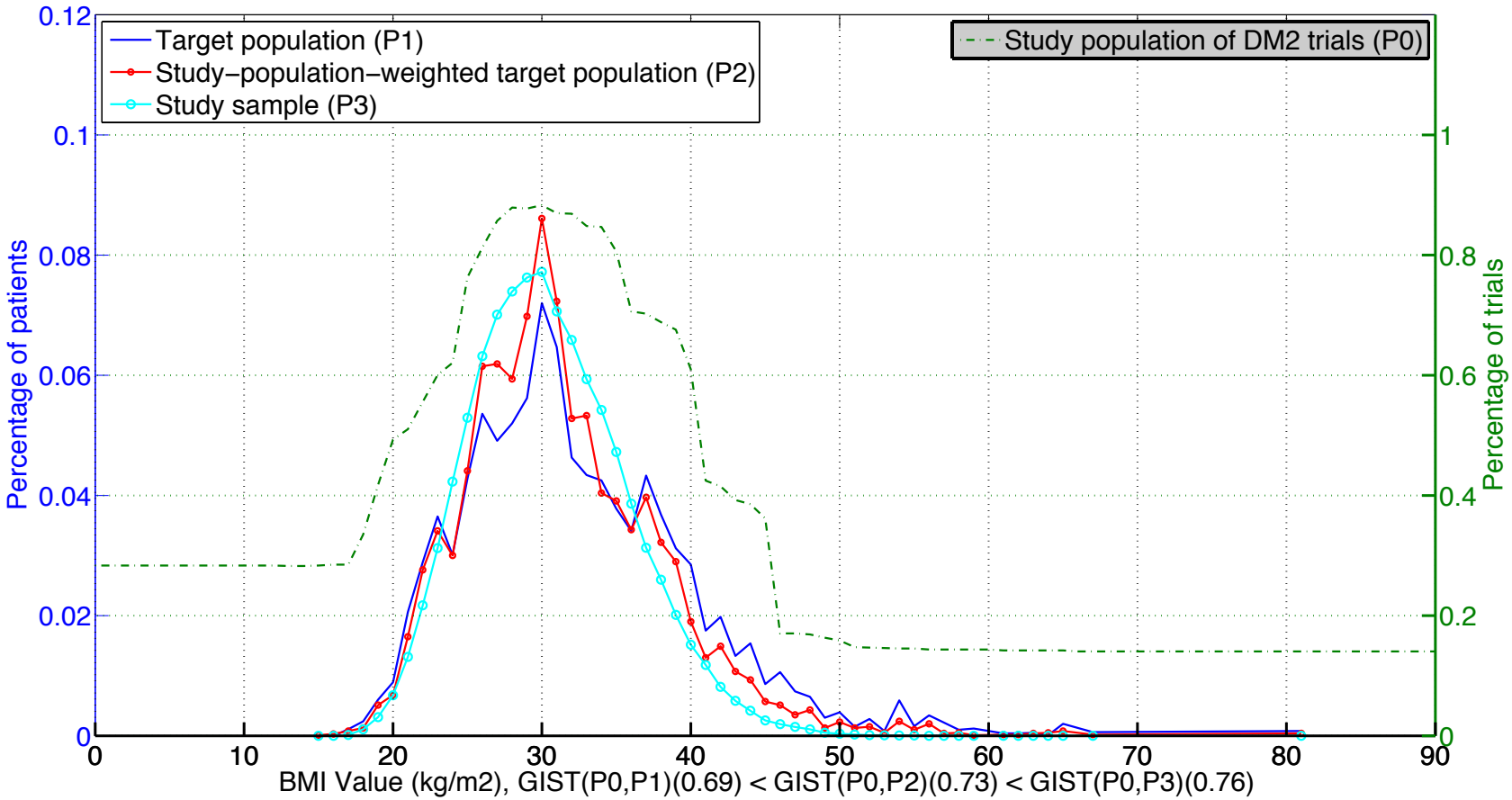


He et al. (2015) Simulation-based Evaluation of the Generalizability Index for Study Traits. AMIA Annu Symp Proc. 2015. In press.

Visualization of populations (Age)



Visualization of populations (BMI)

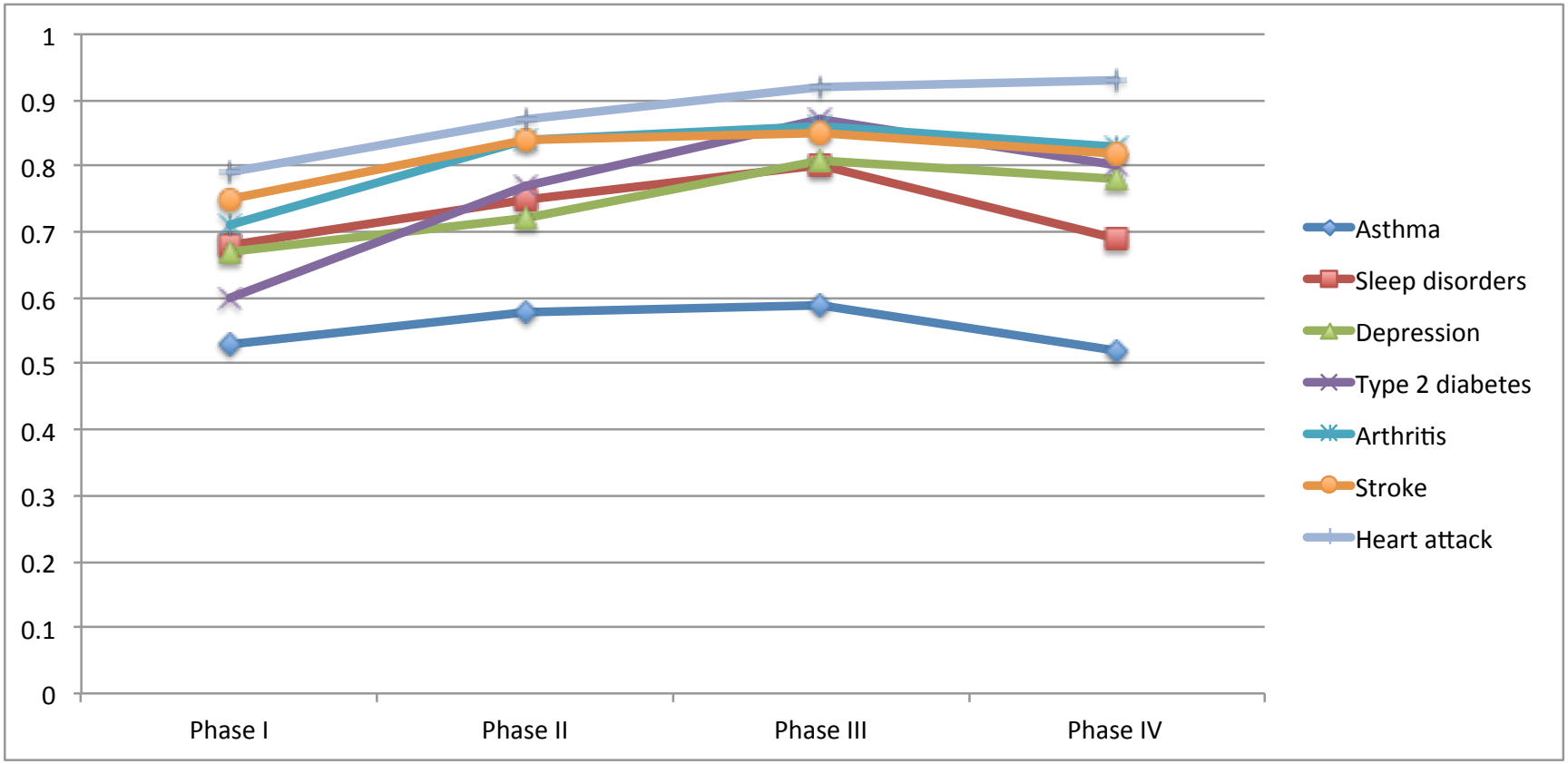


GIST Score for P1, P2, P3

Eligibility Feature	GIST(P0, P1)	GIST(P0, P2)	GIST(P0,P3)
Age	0.77 ± 0.0021	0.79 ± 0.0019	0.85 ± 0.0002
HbA1c	0.47 ± 0.0029	0.55 ± 0.0025	0.72 ± 0.0007
BMI	0.69 ± 0.0020	0.73 ± 0.0019	0.76 ± 0.0005

He et al. (2015) Simulation-based Evaluation of the Generalizability Index for Study Traits. AMIA Annu Symp Proc. 2015. In press.

GIST of Age Stratified by Study Phase



COLLEGE OF COMMUNICATION & INFORMATION
THE FLORIDA STATE UNIVERSITY