BioLINCC and the NHLBI's Vision for Data Indexing

**Sean Coady, Division of Cardiovascular Sciences,
National Heart, Lung, and Blood Institute**

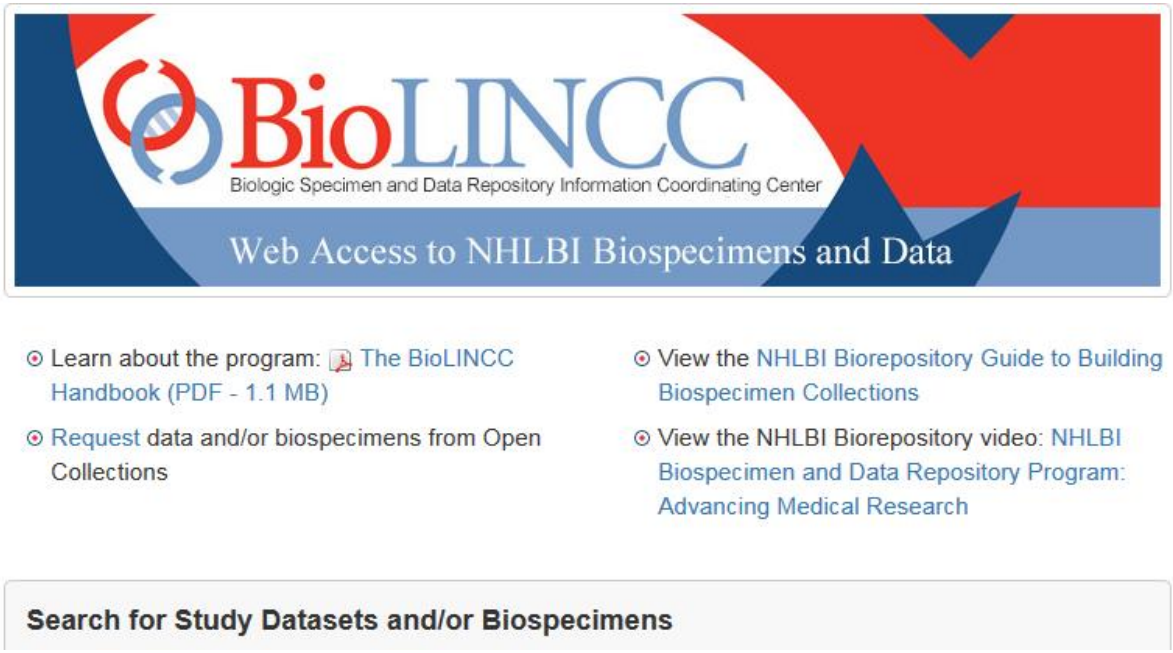
Cross-Cohort Collaboration Consortium Investigator Meeting
March 24, 2018

National Heart, Lung, and Blood Institute

Background

- 1999 NHLBI IRB approves data repository protocol. Website opens in 2000 (NHLBI IRB # 12-H-N198)

- BioLINCC established in 2009 to coordinate activities of Data Repository and Biorepository



National Heart, Lung, and Blood Institute

Eligible Studies

Observational studies and clinical trials funded by NHLBI

- ✓ **Supported by Contract (all)**
- ✓ **Supported by grant**
  - **> $500K and of high programmatic interest**
  - **Ancillary studies conducted on a majority of participants in an eligible parent study**
- **Eligible data consists of all de-identified laboratory, procedural data (exercise test, Echocardiography, Computed Tomography, etc.), questionnaires and outcome data.**
- **All appropriate documentation (protocols, annotated forms, manuals of procedures, data dictionaries, algorithms for calculated variables, etc.)**

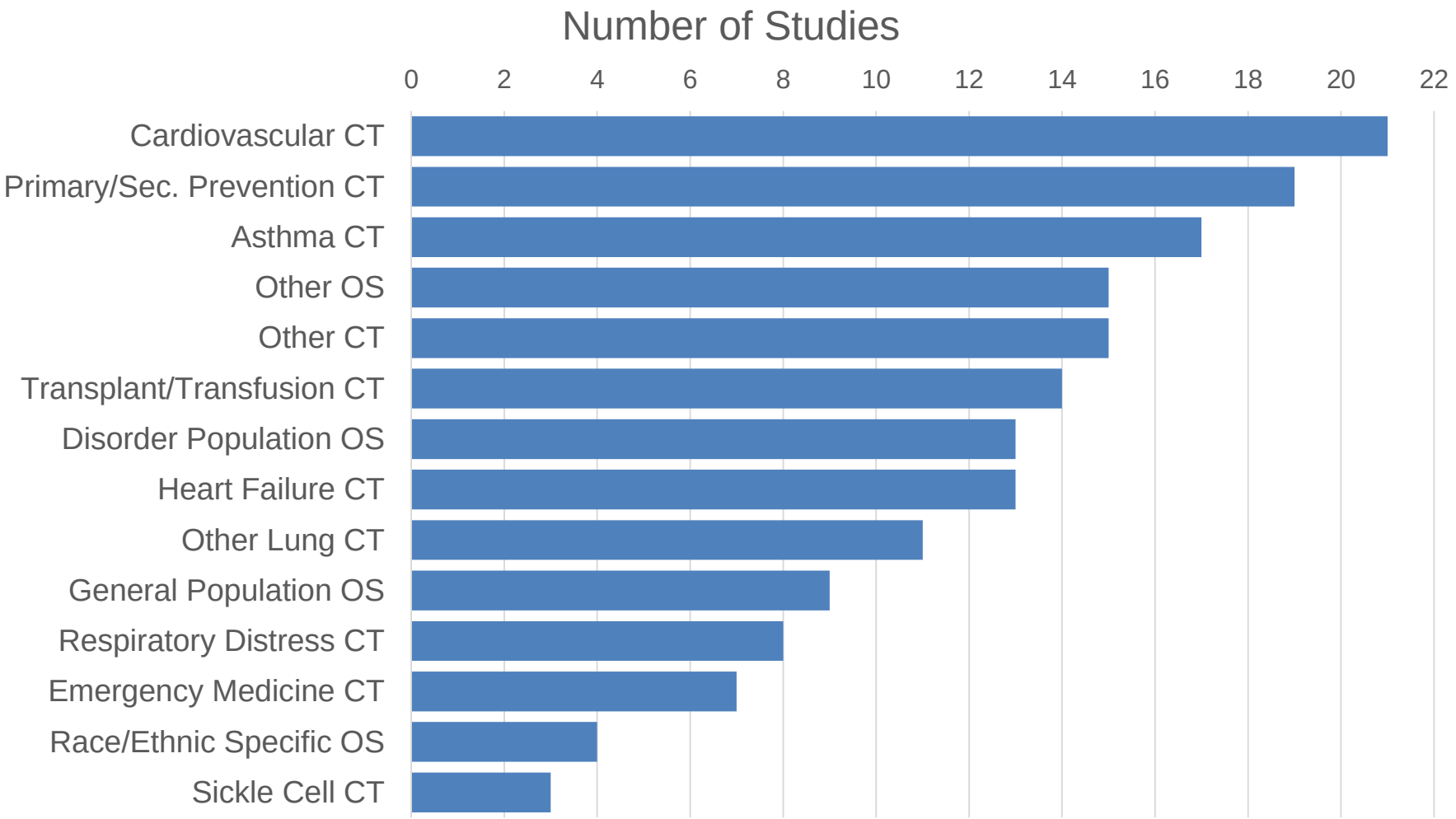
Timing for Release

- **Clinical Trials:**
  - ✓ **3 years after the final visit of the participants to their clinical trial sites, or**
  - ✓ **2 years after publication of primary outcome paper**

- **Observational Studies:**
  - ✓ **3 years after the completion of each examination or follow-up cycle, or**
  - ✓ **2 years after the baseline, follow-up, genetic, ancillary study, or other data set is finalized within the study for analysis**

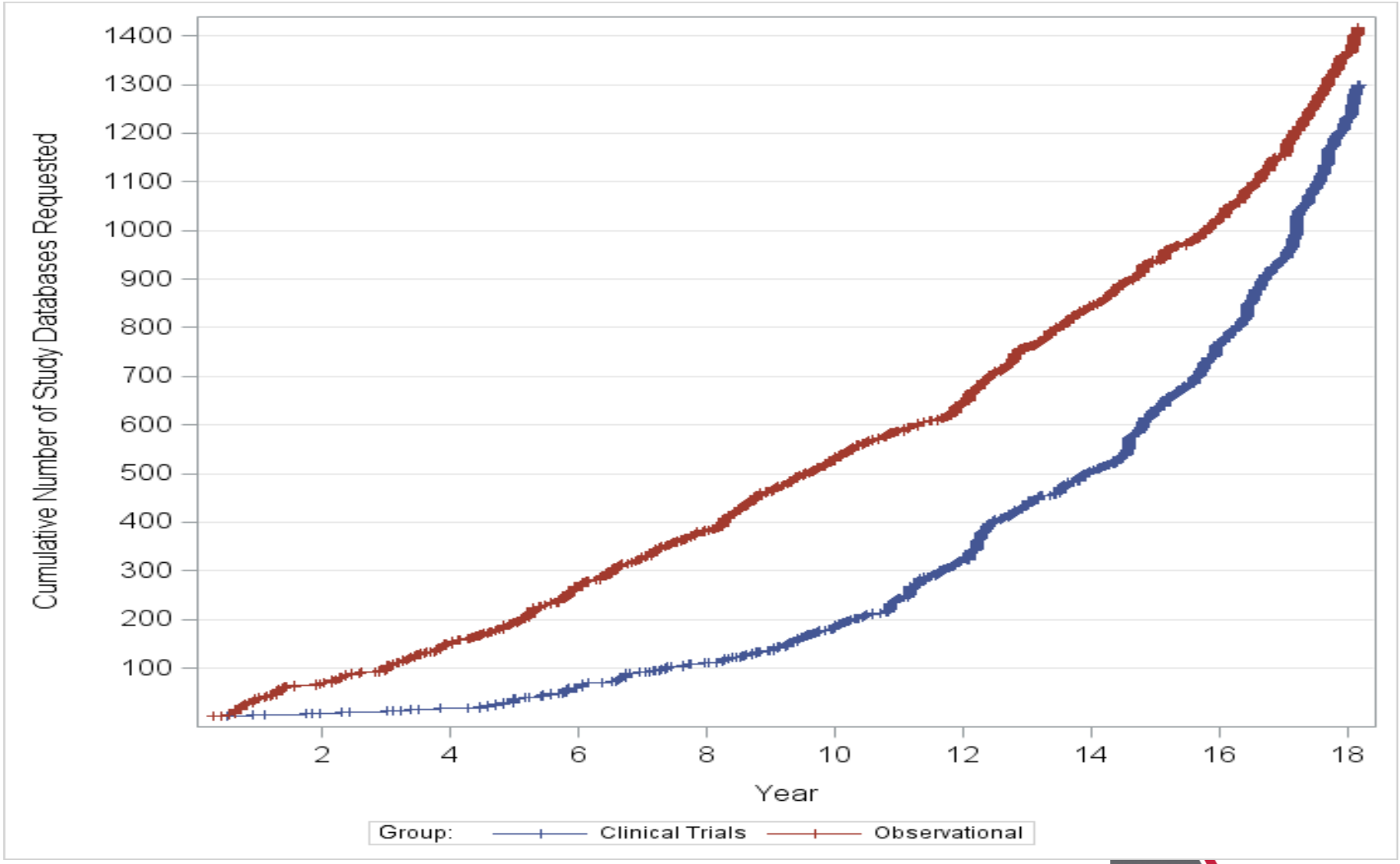
Intent: study investigators receive a 2 year protected period of time

Current portfolio: 43 Observational studies (342,000 participants), 126 trials (417,000 participants)

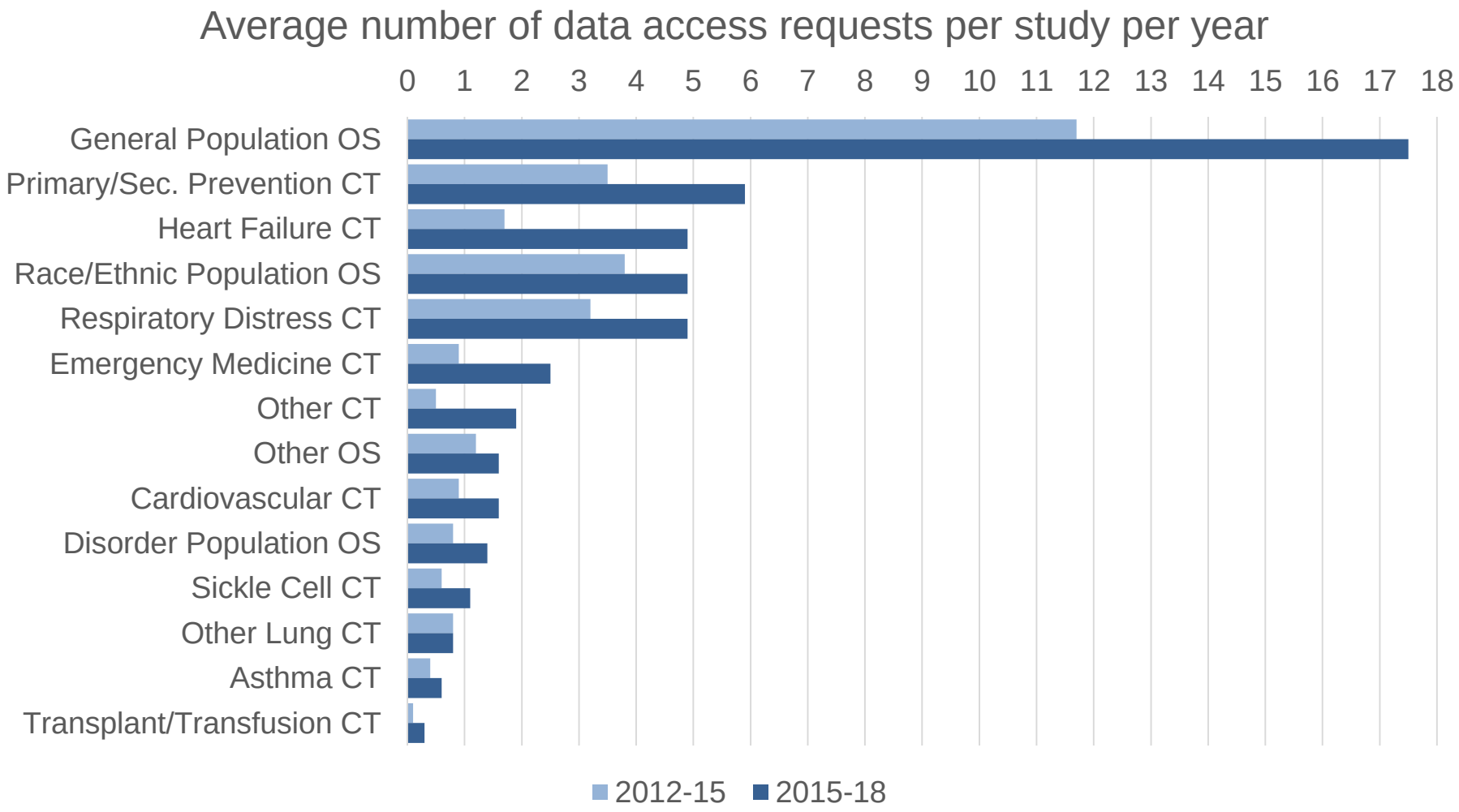


NIH National Heart, Lung, and Blood Institute

Cumulative access requests for trial and observational study data (thru Mar 15, 2018)

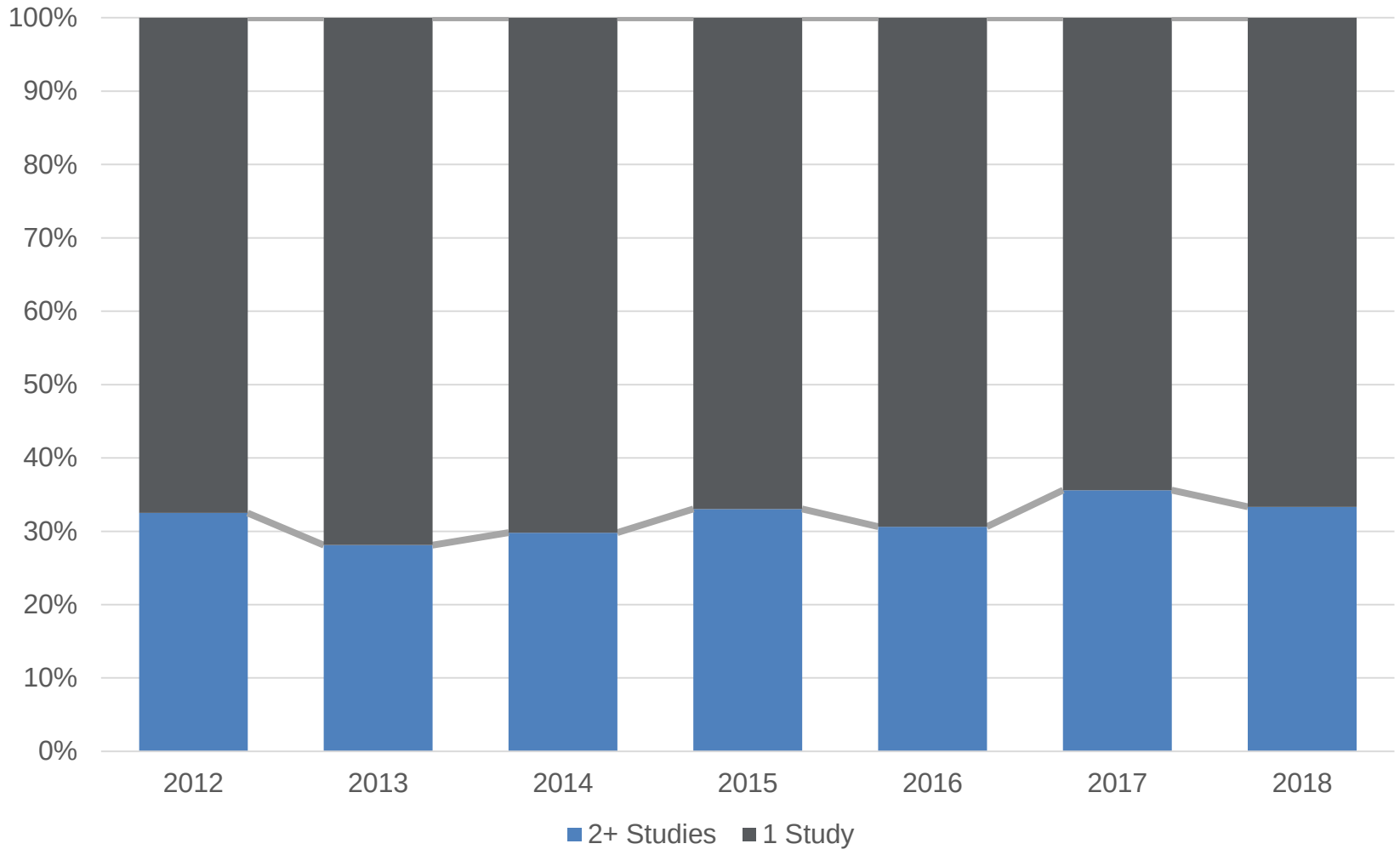


Annual average number of data access requests for study databases per study per year available (3 year rates)



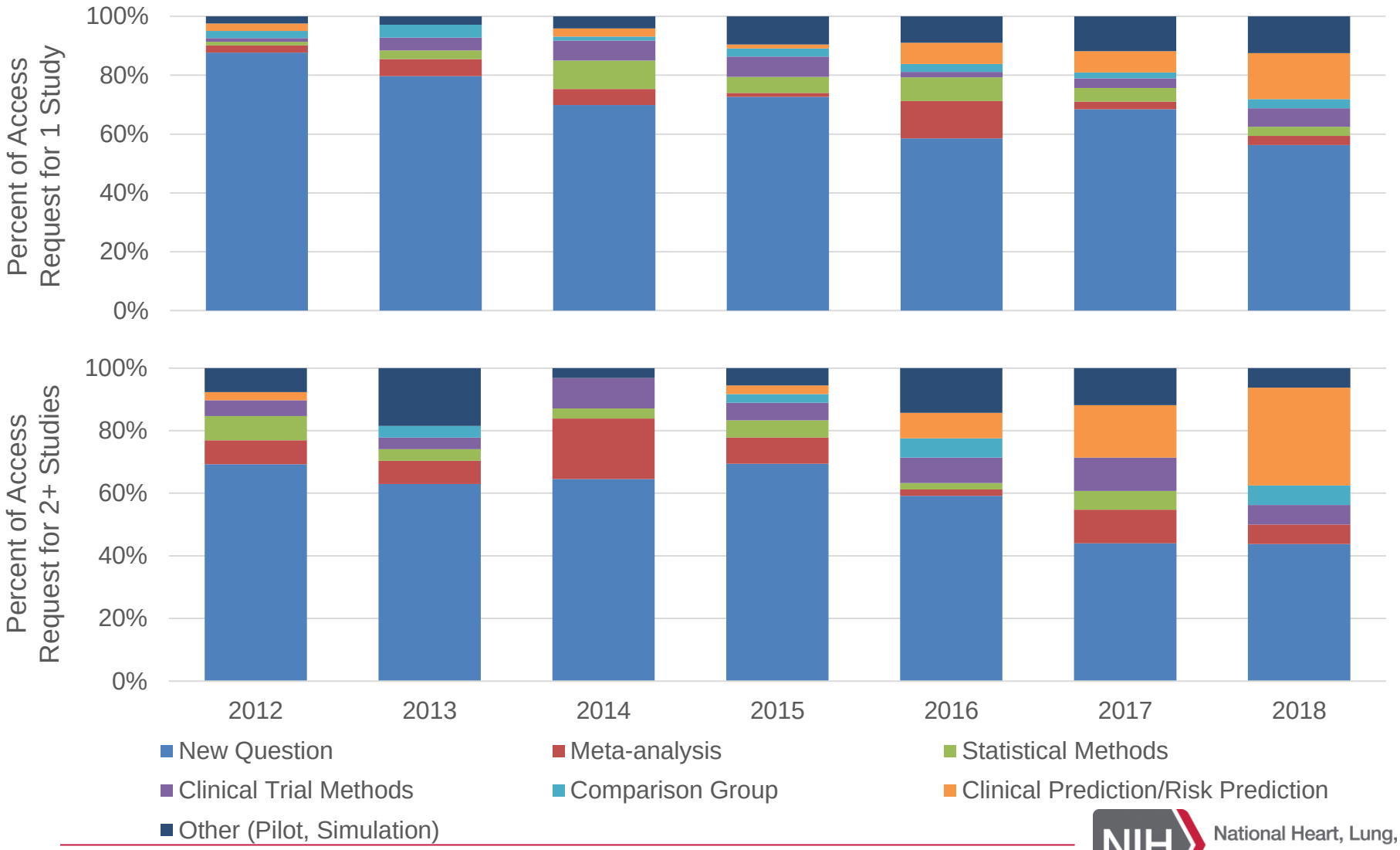
National Heart, Lung, and Blood Institute

Data Access Requests: Percent Requesting Single Study or 2+ Studies, 2012-Present



National Heart, Lung, and Blood Institute

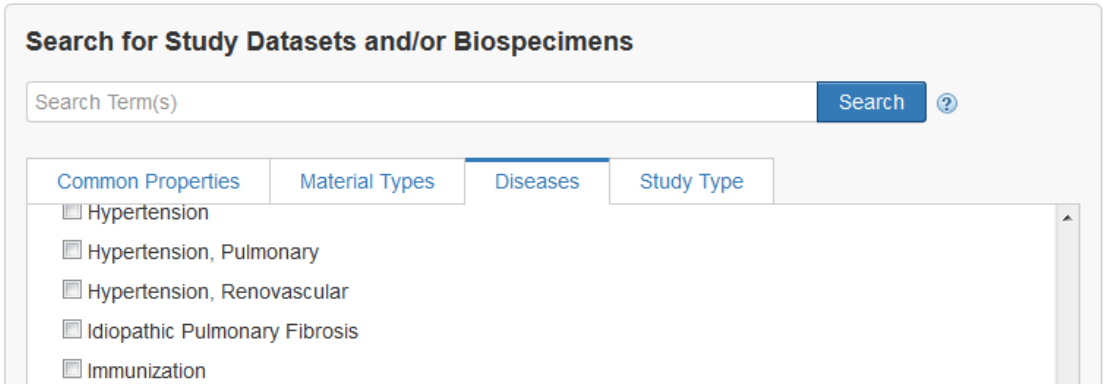
Primary Reason for Access Request for a Single Study (top panel) or Multiple Studies (bottom panel)



Need for accessible, searchable data index

- Currently repository has 169 clinical studies across the spectrum HLBS phenotypes, adding approximately 20+ new studies each year

- Approximately 1/3 of data access requests involve more than one study



New Study:
2018-03-06

A new study, S
Failure (STICH
from this study

New Study:
2018-03-05

A new study, F
Program (PRC
posted to BioL
for requests.

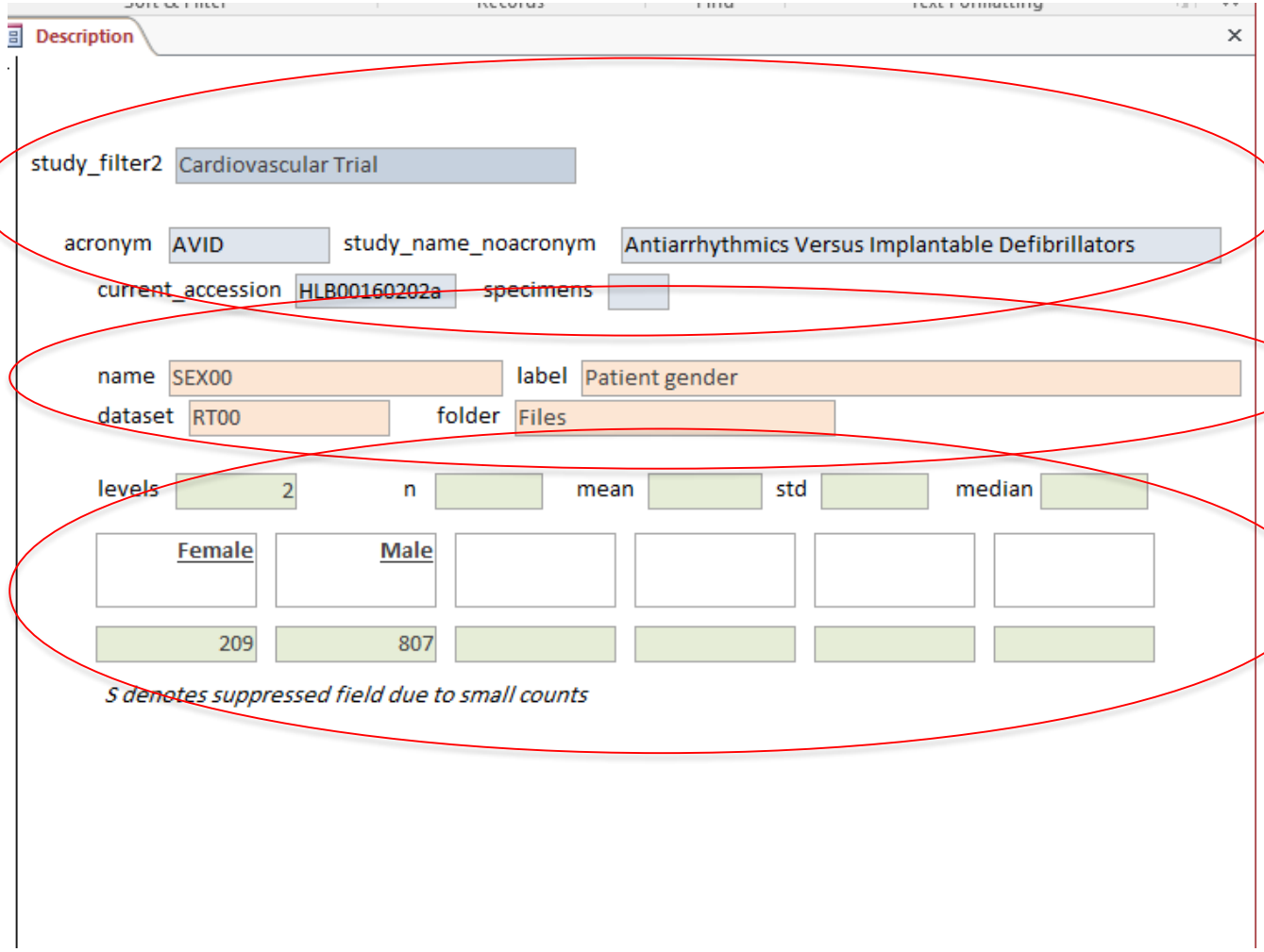
Current search is "study" based

NIH National Heart, Lung, and Blood Institute

Pilot study

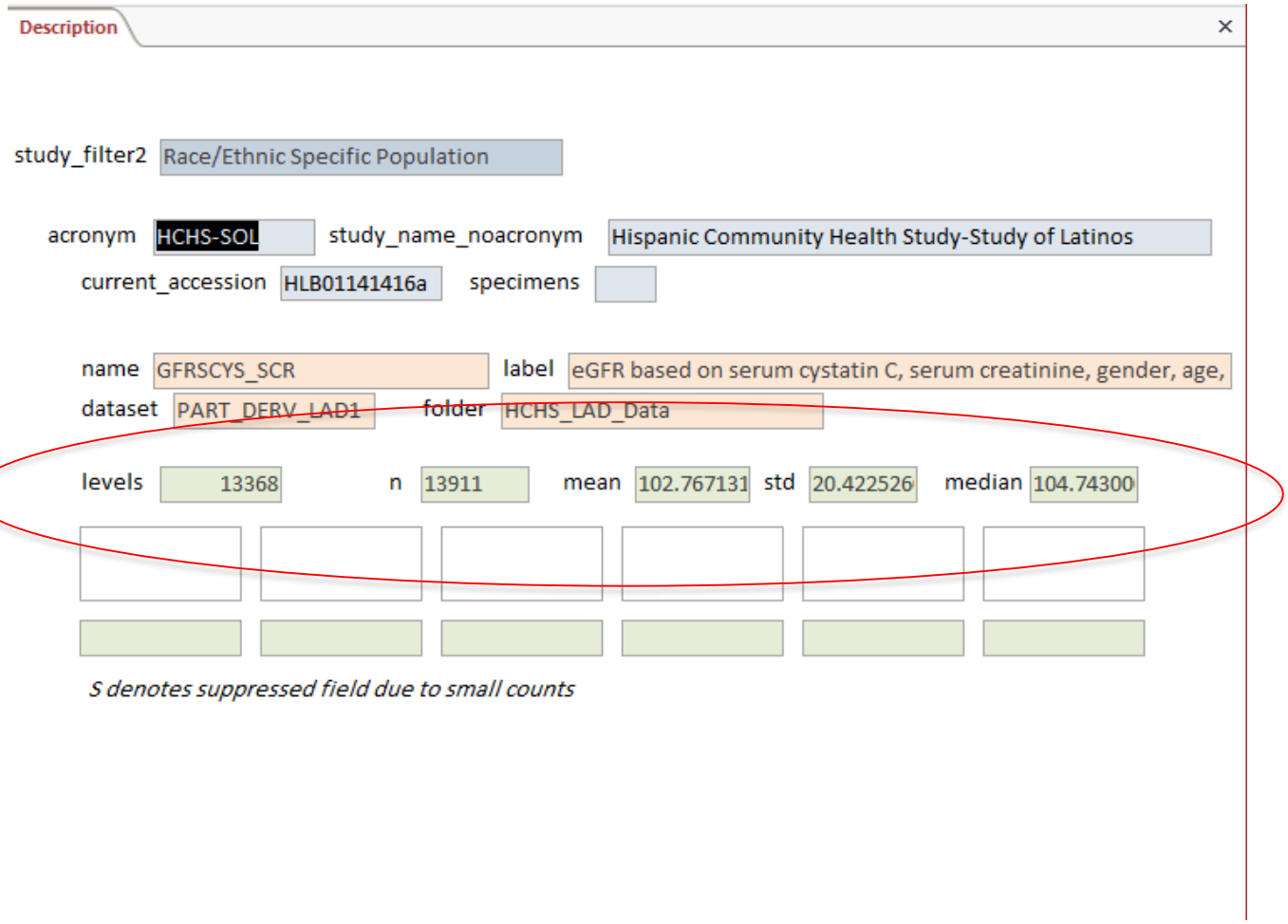
- “Come as you are” repository limits resources for indexing
- Pilot Index using 11 studies: ARIC, CARDIA, CHS, Framingham, MESA, HCHS-SOL, AVID, DASH, ROC-PRIMED, ROMICAT II, and TAAG.
- Covers 1,277 datasets and 113,437 variables

Pilot results – search “sex” / “gender”



- Basic study information
- Basic data element info
- Data distribution info

Pilot results – “sex” / ”gender”



Pilot results – potential filters

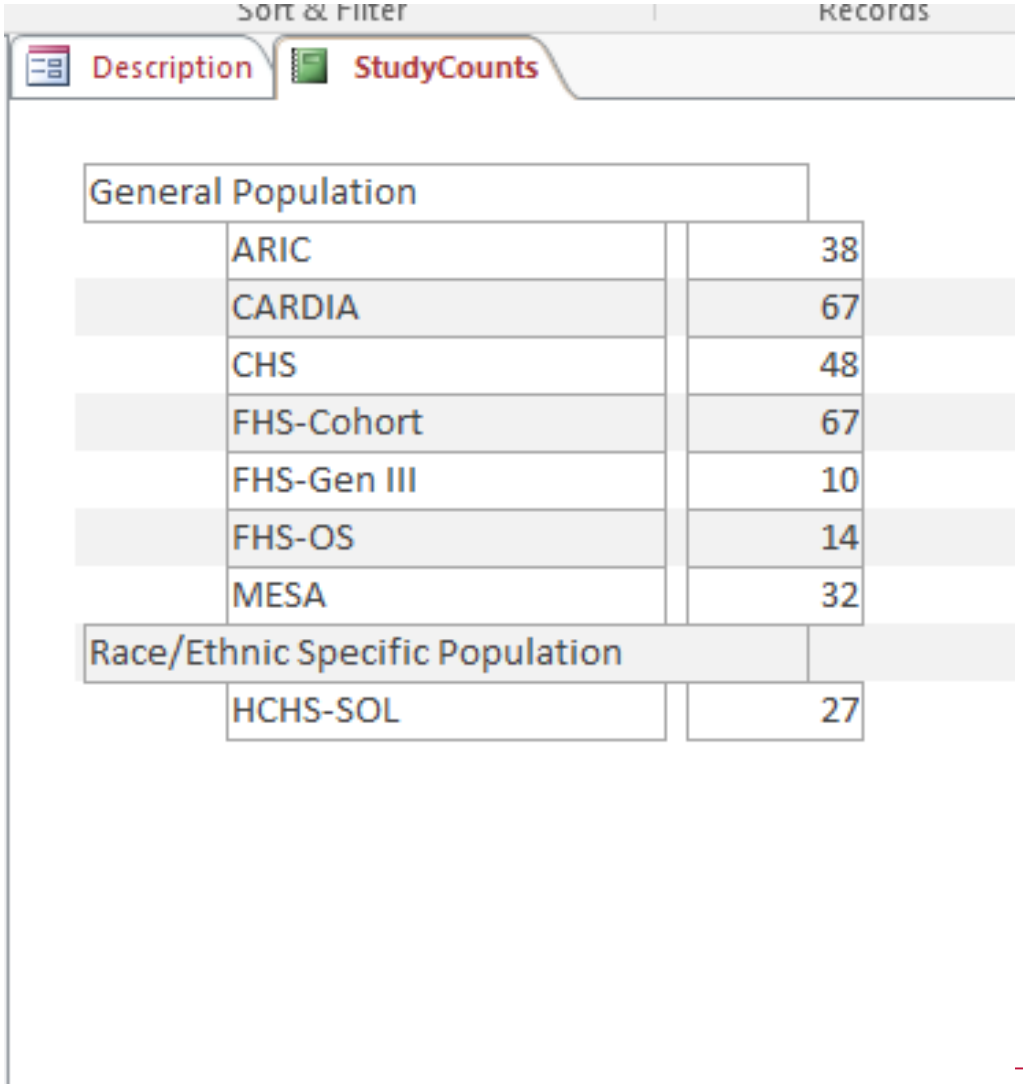
Description | StudyCounts

Cardiovascular Trial		
	AVID	39
General Population		
	ARIC	19
	CARDIA	47
	CHS	11
	FHS-Cohort	4
	FHS-Gen III	8
	FHS-OS	8
	MESA	13
Other Trial		
	DASH	1
	ROMICAT-II	2
	TAAG	24
Race/Ethnic Specific Population		
	HCHS-SOL	39

Search on “sex” or “gender” yielded 215 hits

Filters based on domain or study name relatively easy to element

Pilot results - “FEV” / ”spirometry” search

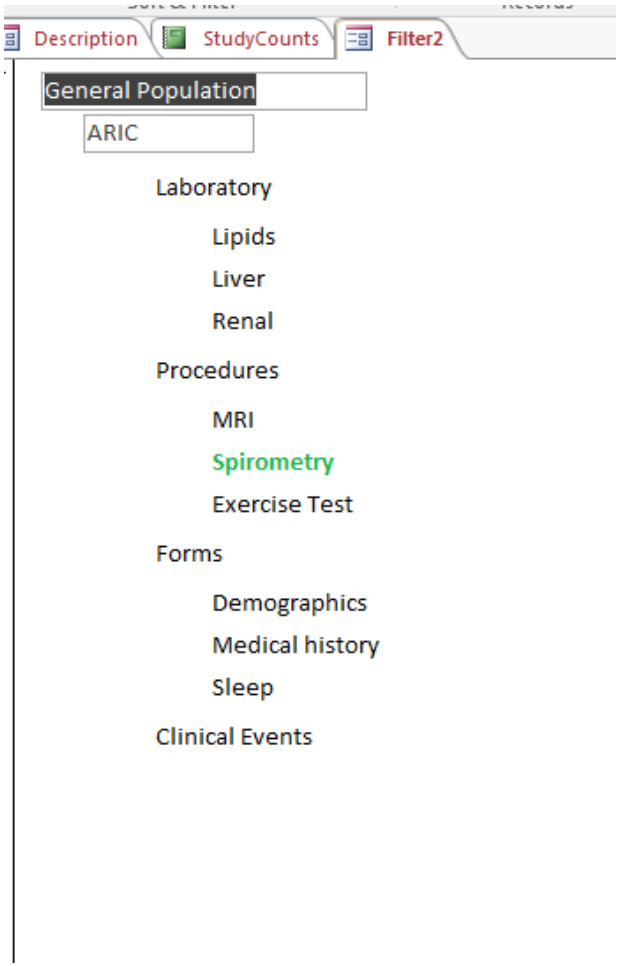


General Population	
ARIC	38
CARDIA	67
CHS	48
FHS-Cohort	67
FHS-Gen III	10
FHS-OS	14
MESA	32
Race/Ethnic Specific Population	
HCHS-SOL	27

Search on “fev” or “spirometry” yielded 303 hits

Other terms: forced expiratory, pulmonary function, pft, lung function, vital capacity

Pilot results – future directions



To be truly usable, need filters based on mapping to an ontology applicable to clinical research.

Hierarchy would need to be narrow enough to be useful, but not so narrow as to overwhelm

Summary

- Clear need to develop an accessible, searchable and usable index to enable discovery of available data elements
- Feasible to create searchable index using straight-forward programming
- To be usable, need mapping to an ontology designed for clinical research data