

BioLINCC and the NHLBI's Vision for Data Indexing

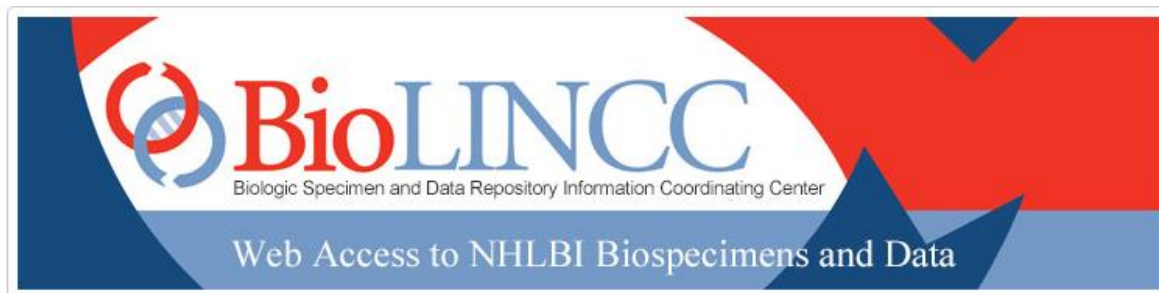
**Sean Coady, Division of Cardiovascular Sciences,
National Heart, Lung, and Blood Institute**

Cross-Cohort Collaboration Consortium Investigator Meeting
March 24, 2018



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



○ Learn about the program: [The BioLINCC Handbook \(PDF - 1.1 MB\)](#)

○ Request data and/or biospecimens from Open Collections

○ View the NHLBI Biorepository Guide to Building Biospecimen Collections

○ View the NHLBI Biorepository video: [NHLBI Biospecimen and Data Repository Program: Advancing Medical Research](#)

Search for Study Datasets and/or Biospecimens

Featured News

Funding Opportunity: NHLBI R21 RFA-HL-17-022

This funding opportunity is to support meritorious exploratory research relevant to the NHLBI mission using the biospecimen collections that are stored in the NHLBI Biorepository and that are available through BioLINCC. The next application due date is October 17, 2017.

[← Previous](#) [Next →](#)

Recent News

[New Study: CPAP \(data\)](#)
2017-05-01

Eligible Studies

Observational studies and clinical trials funded by NHLBI

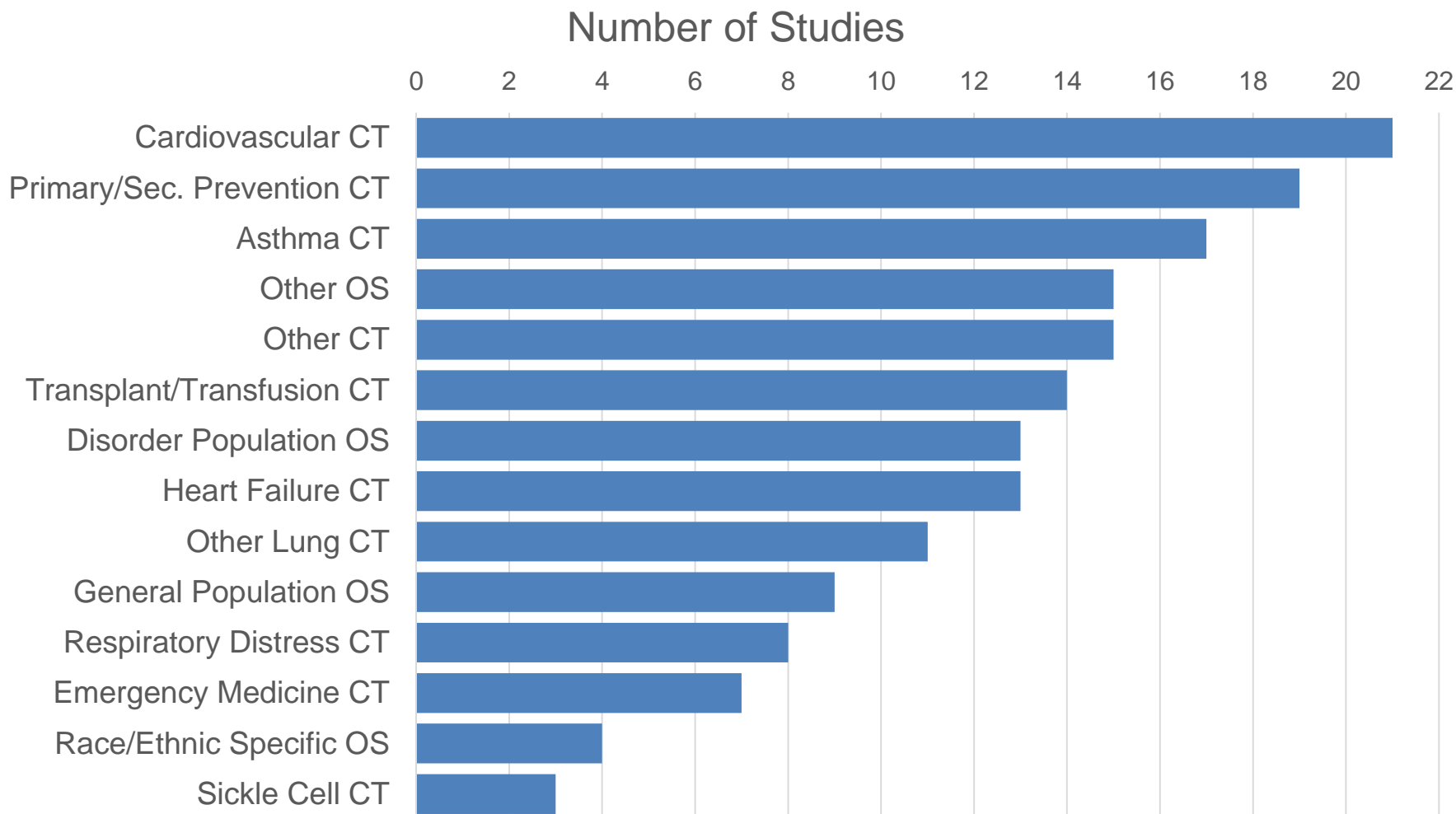
- ✓ **Supported by Contract (all)**
- ✓ **Supported by grant**
 - **> \$500K and of high programmatic interest**
 - **Ancillary studies conducted on a majority of participants in an eligible parent study**
- **Eligible data consists of all de-identified laboratory, procedural data (exercise test, Echocardiography, 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

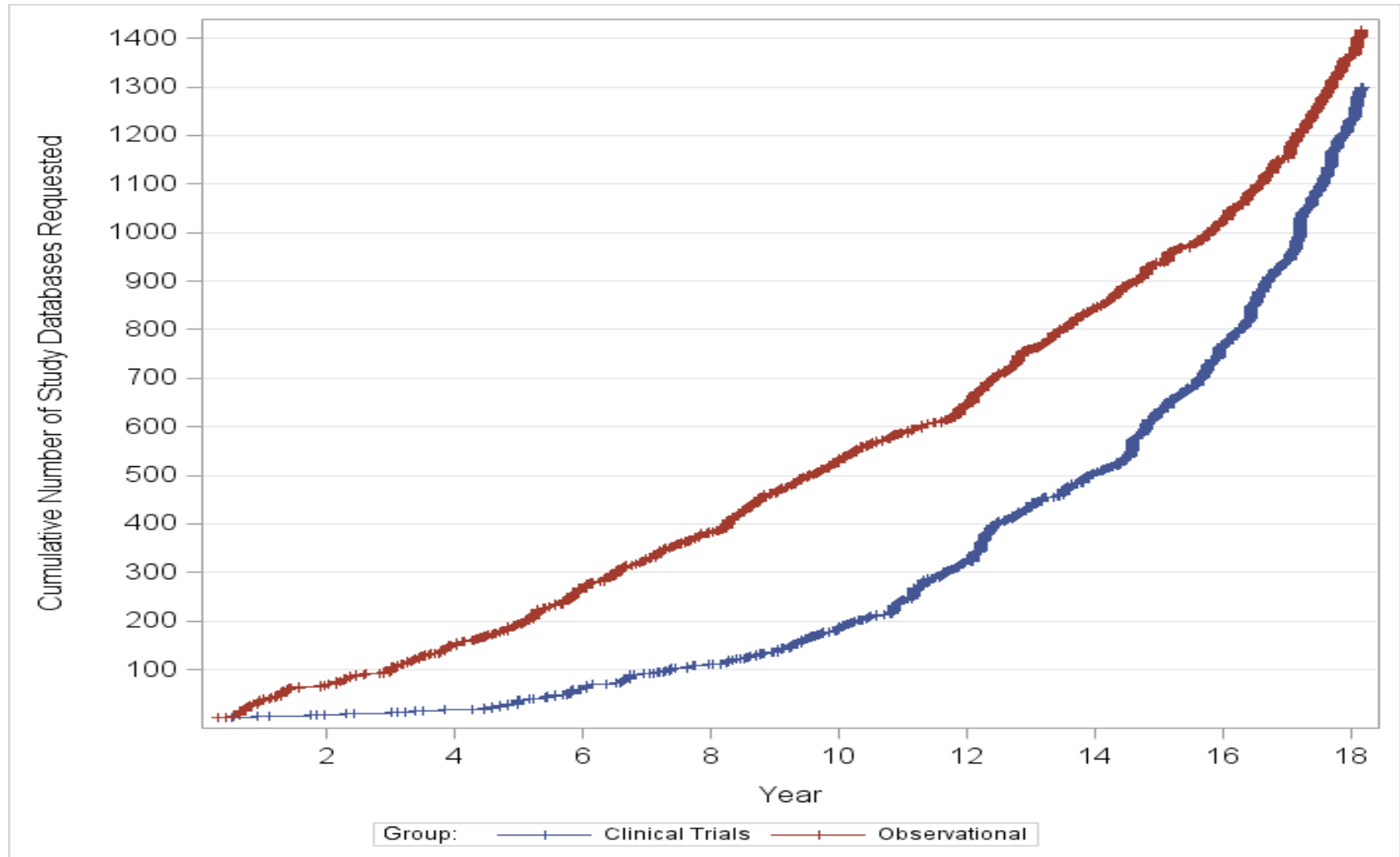
- **Clinical Trials:**
 - ✓ 3 years after the final visit of the participants to their clinical trial sites, or
 - ✓ 2 years after publication of primary outcome paper
- **Observational Studies:**
 - ✓ 3 years after the completion of each examination or follow-up cycle, or
 - ✓ 2 years after the baseline, 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)

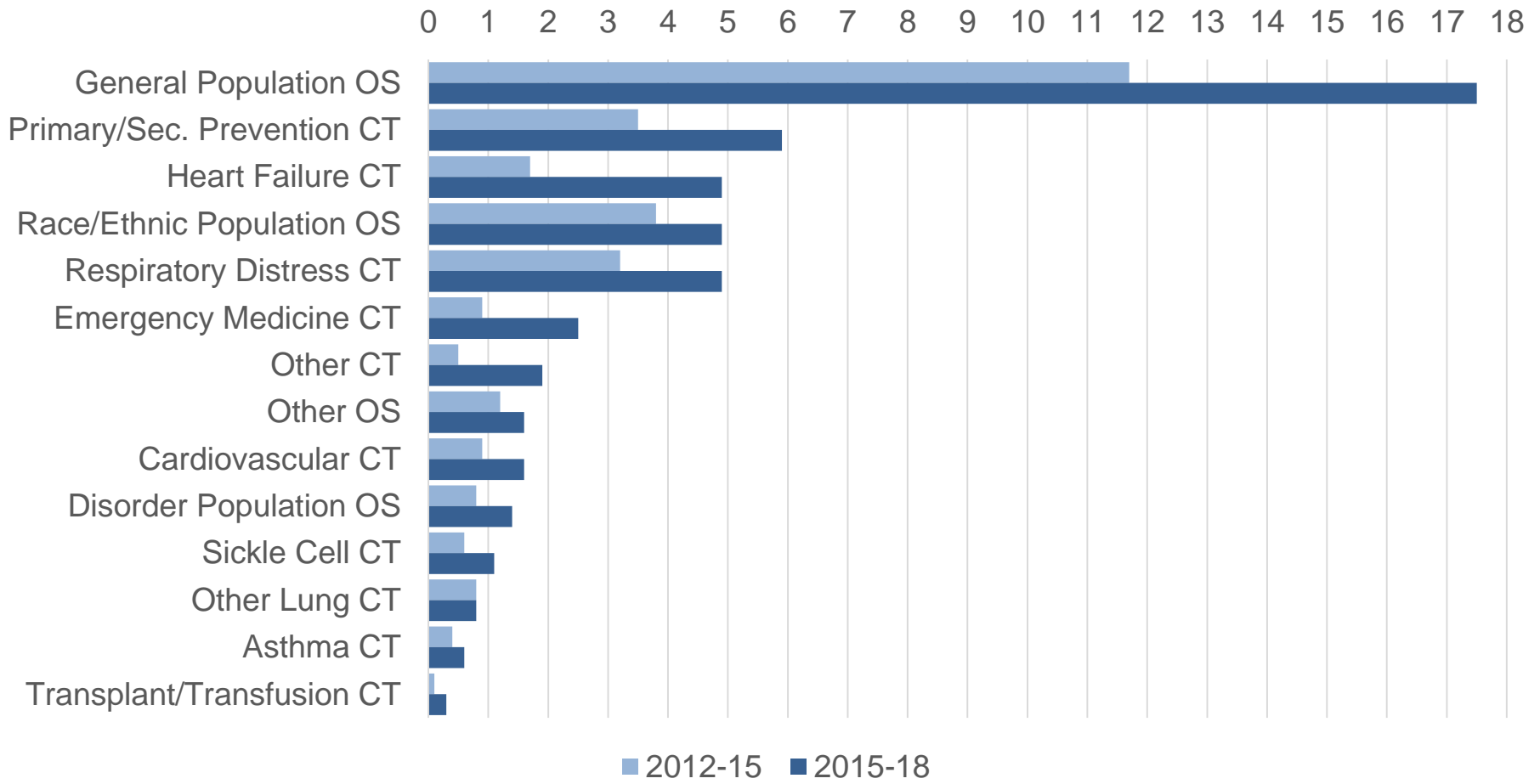


Cumulative access requests for trial and observational study data (thru Mar 15, 2018)

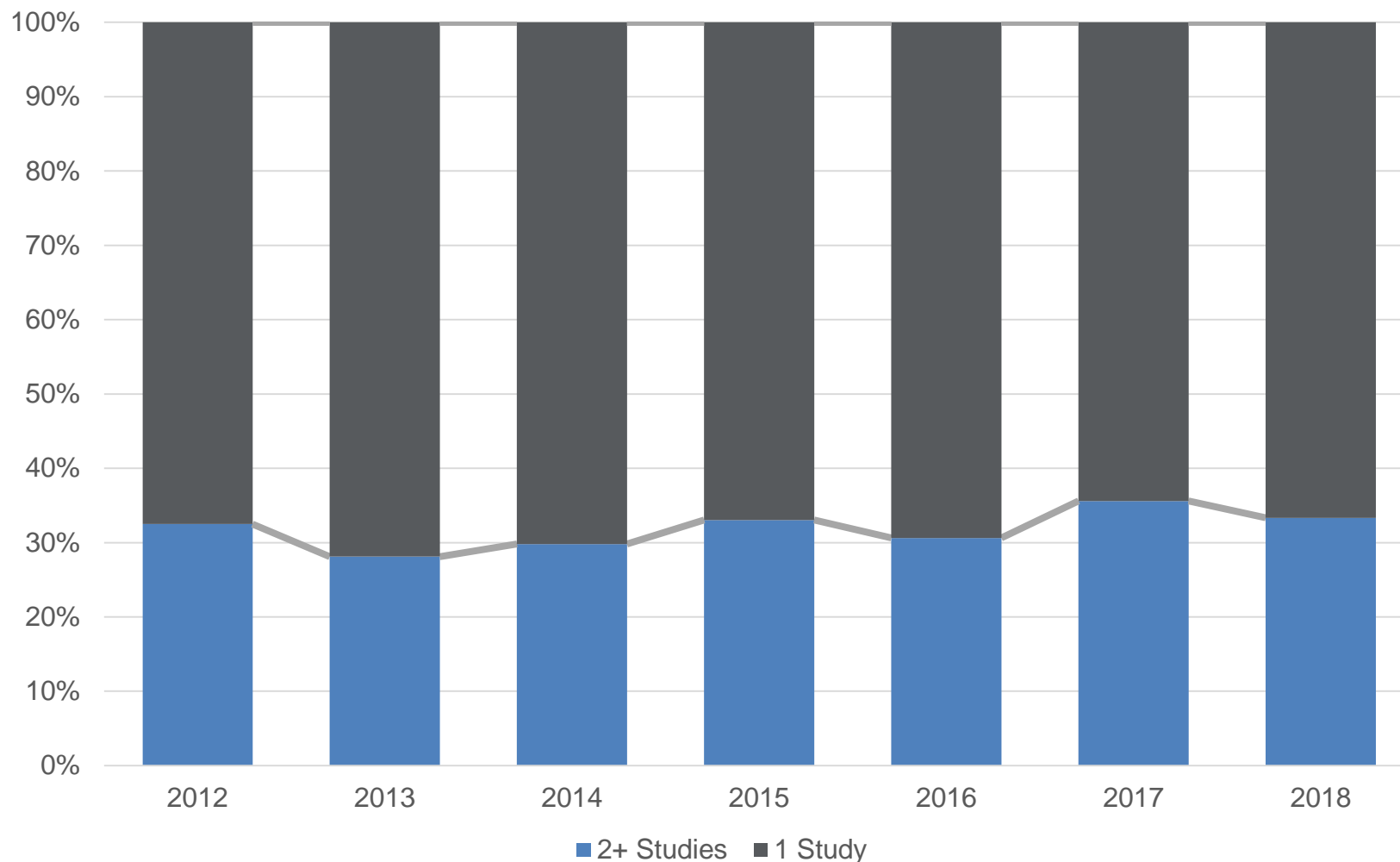


Annual average number of data access requests for study databases per study per year available (3 year rates)

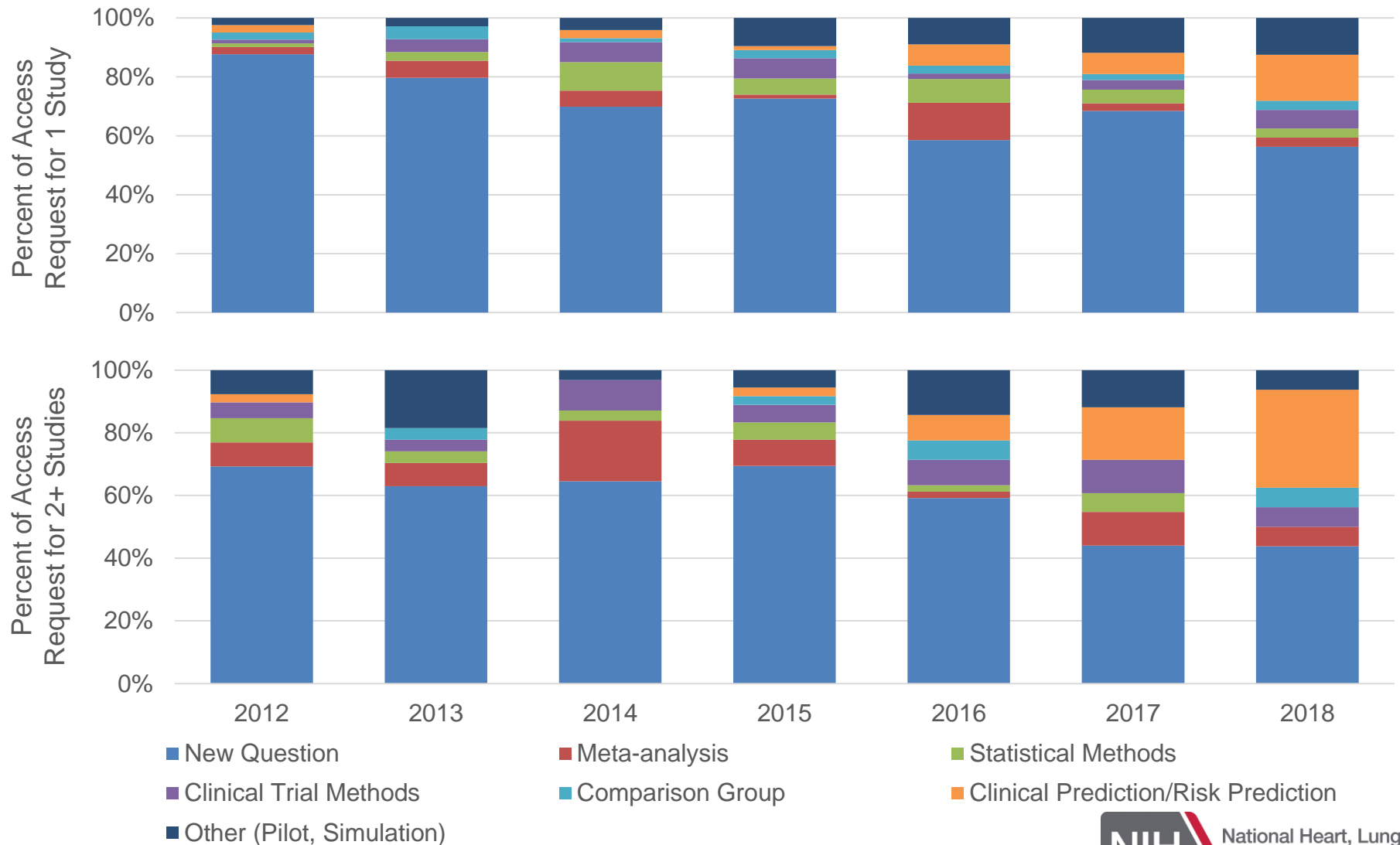
Average number of data access requests per study per year



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




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

Search for Study Datasets and/or Biospecimens

Search Term(s) 

Common Properties | Material Types | Diseases | Study Type

- Hypertension
- Hypertension, Pulmonary
- Hypertension, Renovascular
- Idiopathic Pulmonary Fibrosis
- Immunization

New Study: :
2018-03-06

A new study, S
Failure (STICH
from this study

New Study: I
2018-03-05

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

Current
search is
“study”
based

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”

The screenshot shows a web interface with a 'Description' tab. It displays search results for a study filter. The results are organized into several sections, each highlighted with a red oval:

- Basic study information:** Includes fields for 'study_filter2' (Cardiovascular Trial), 'acronym' (AVID), 'study_name_noacronym' (Antiarrhythmics Versus Implantable Defibrillators), and 'current_accession' (HLB00160202a).
- Basic data element info:** Includes fields for 'name' (SEX00), 'label' (Patient gender), 'dataset' (RT00), and 'folder' (Files).
- Data distribution info:** Includes a table with columns for 'levels', 'n', 'mean', 'std', and 'median'. The 'levels' column shows 'Female' and 'Male'. The 'n' column shows counts of 209 for Female and 807 for Male. The other columns are empty.

At the bottom of the table, there is a note: *S denotes suppressed field due to small counts*.

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

Pilot results – “sex” / “gender”

Description

x

study_filter2 Race/Ethnic Specific Population

acronym HCHS-SOL study_name_noacronym Hispanic Community Health Study-Study of Latinos

current_accession HLB01141416a specimens

name GFRSCYS_SCR label eGFR based on serum cystatin C, serum creatinine, gender, age,

dataset PART_DERV_LAD1 folder HCHS_LAD_Data

levels 13368 n 13911 mean 102.767131 std 20.422526 median 104.74300



S denotes suppressed field due to small counts

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

SORT & FILTER		RECORDS
Description	StudyCounts	
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 straightforward programming
- To be usable, need mapping to an ontology designed for clinical research data