



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

FOUNDATION
FOR THE
National Institutes of Health

Data Sources: Landscape and OMOP Distributed Partners Summary of OMOP Data Provider Survey

Abraham G. Hartzema, PharmD, MSPH, PhD, FISPE
Professor and Eminent Scholar,

Perry A. Foote Chair in Health Outcomes Research and Pharmacoeconomics;
Departments of Pharmaceutical Outcomes & Policy, and
Epidemiology and Biostatistics, University of Florida
OMOP Principal Investigator

This is pre-published data and not for citation or reproduction without permission.

The results of the survey have been submitted for publication.

PARTNERS FOR INNOVATION, DISCOVERY, LIFE



Our Enthusiastic Collaborators

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

FOUNDATION
FOR THE
National Institutes of Health

Patrick Ryan, MEng

Emily Welebob, RN MS

Paul Stang, PhD FISPE

J. Marc Overhage, MD PhD



Survey Overview

- Purpose: To assess the availability and content of observational data and data holders interest in participating in drug safety research
- Two-stage online survey conducting from January - June 2009
- First stage: 10-question survey, less than 10 minutes
 - 70 organizations contacted
 - Three follow-up reminders
 - 37 US sources responded, 2 international sources
- Second stage: 70-question questionnaire
 - Survey required quantitative descriptive summary of data elements
 - Organizations offered \$1,000 subsidy
 - 21 of 37 first-stage respondents completed second-stage survey



Characteristics of respondents

First-stage respondents

- 26/37 (70%) of US organizations have access to data for >1M persons
- 9 databases with access over 10M persons
- 4 databases with access to over 50M persons

Second-stage respondents

- 21 US organizations, representing 540M lives in aggregate
- 257M persons were currently active at the time of response

This is pre-published data and not for citation or reproduction without permission. The results of the survey have been submitted for publication.



Structure of Observational Data

- Integrated Health Care Systems
- Fiscal Intermediaries
- Data aggregators
 - Niche data aggregators, e.g. hospital bench markers

This is pre-published data and not for citation or reproduction without permission. The results of the survey have been submitted for publication.



Second stage survey respondents by population size and data type

Population size (in 000's)		Data type	
Total	Current	Claims	EHR
160,000	90,000	X	
115,500	50,000	X	
81,000	35,000	X	
50,000	14,000	X	
37,197	17,619	X	
20,000	10,000	X	
14,229	5,710	X	
12,920	9,043	X	X
11,100	6,000		X
9,454	2,000	X	X
8,000	6,300	X	
7,835	5,300		X
4,980	1,829	X	X
3,100	2,400	X	X
1,720	771	X	X
1,375	218	X	
763	218	X	
479	138		X
34	34		X

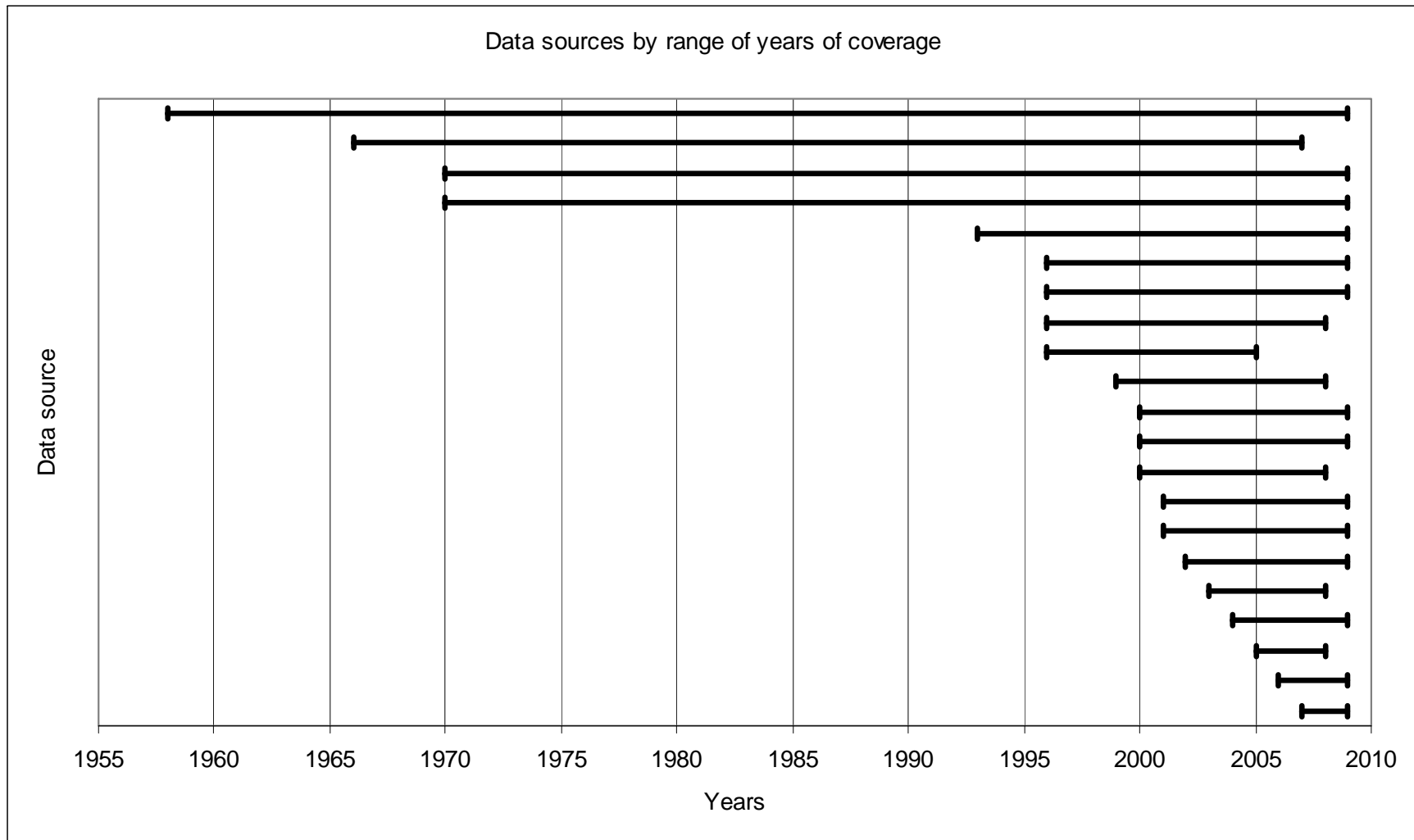


Capture of data elements

Data Element		% of respondents
Demographics	Year of Birth	100%
	Gender	100%
	Race	90%
Conditions	Diagnoses, outpatient	100%
	Diagnoses, inpatient	90%
	Symptoms not represented in diagnostic codes	43%
Drugs	Vaccine administration	86%
	Drug dispensings	76%
	Prescriptions written	62%
	Inpatient pharmacy records	62%
	Infusion administrations	67%
	Other drugs reported	67%
Observations	Laboratory values, outpatient	81%
	Vital signs	62%
	Laboratory values, inpatient	52%
Visits and Procedures	Outpatient visits	100%
	Inpatient visits	90%
	Procedures, outpatient	95%
	Procedures, inpatient	90%

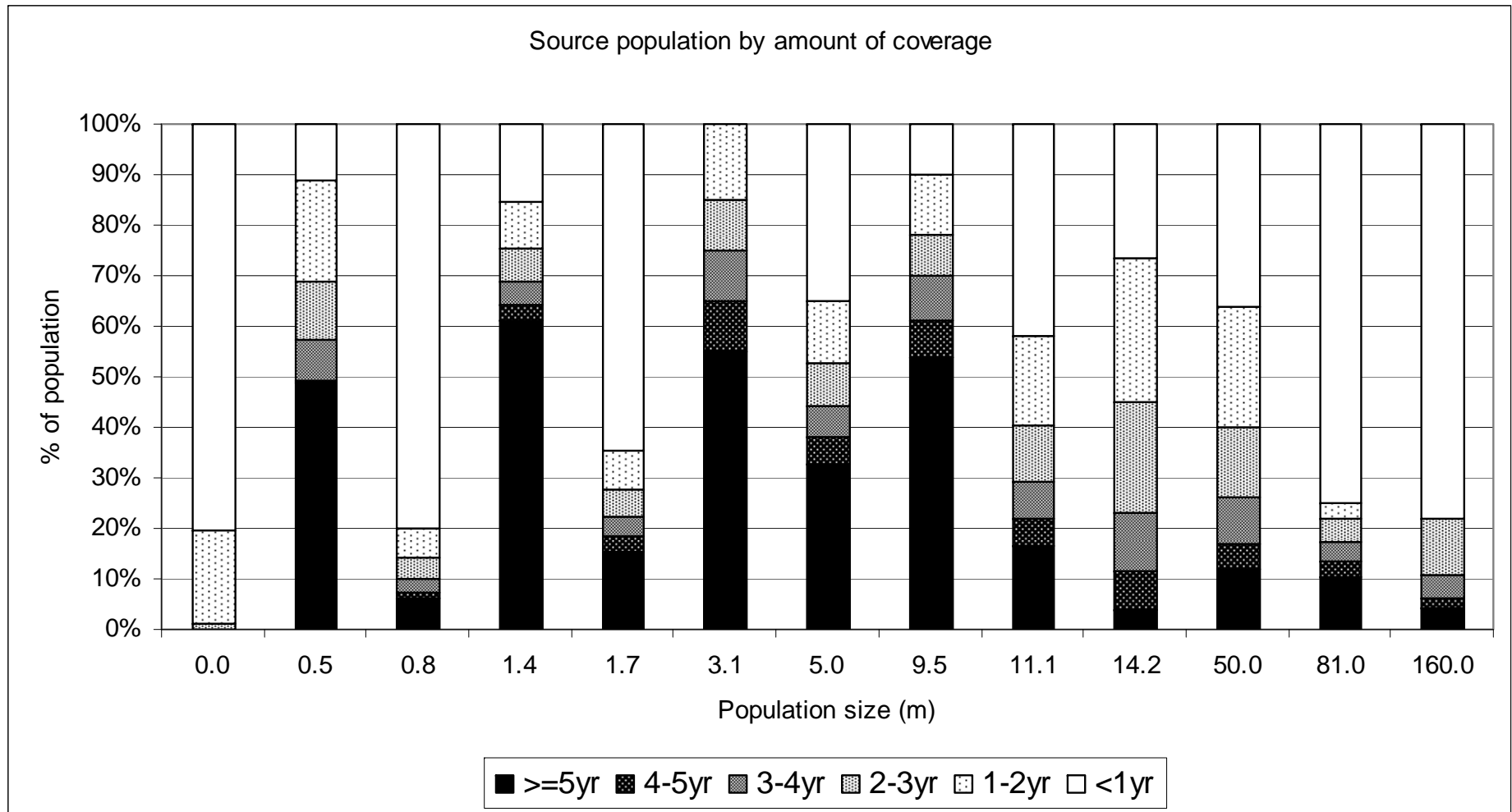


Data sources by years of coverage





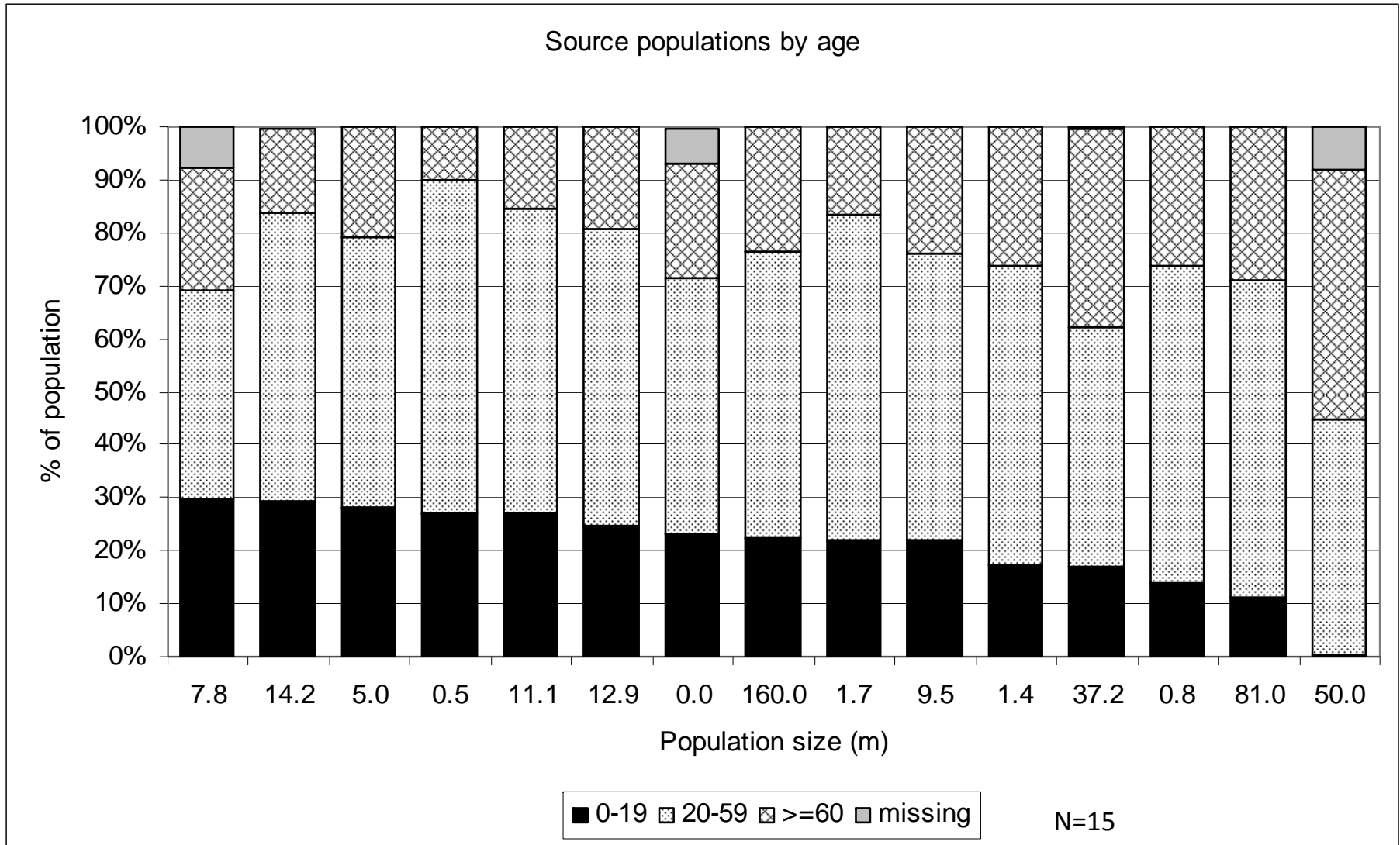
Source population by amount of coverage



*This is pre-published data and not for citation or reproduction without permission.
The results of the survey have been submitted for publication.*



Source population by age





Provider willingness for data access models

	Organizations (n=21)	Total population (m)
Centralized model: Provide your data externally to load into the Central Research Core IT environment	7	297
Federated model: Facilitate OMOP researchers access to execute queries directly (through firewall)	4	252
Distributed CDM Model: OMOP queries run locally by your research staff	17	470
Distributed protocol model: Develop and run your own queries locally	19	413

Each access model would have access to over 250m lives in aggregate, indicating 100m persons is achievable under all alternative infrastructures without full participation of potential data sources

*This is pre-published data and not for citation or reproduction without permission.
The results of the survey have been submitted for publication.*



Key points

- OMOP has undertaken this two-stage survey to better understand the extent of data assets available for active drug surveillance, the characteristics of the data, and the willingness of data organizations to participate in any of a number of data access models.
- There is great diversity in the currently available data for active drug safety surveillance; no single source is likely to meet all requirements, and each has its unique set of limitations.
- Significant variability in characteristics of different data sources suggests consistency and performance of drug safety analyses across sources requires empiric evaluation.
- Data organizations' willingness to participate in drug safety research was apparent across a number of data access models, including providing de-identified data into a centralized environment, granting secure access through a federated model, or conducting standardized analysis as part of a distributed partnership



Questions?

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

FOUNDATION
FOR THE
National Institutes of Health