



Report of NIH Cohort Inventory Findings

Dave Kaufman, Ph.D.

NHGRI

Division of Genetics and Society

Wednesday, February 11, 2015

Cohort inventory



Cohort inventory goal and methods

Characteristics for PM cohort

- Numbers enrolled
- Demography/geography
- Consent for broad uses of existing data
- Ability to follow up
- Biospecimens
- Willingness to share data
- Linked to EMRs

Inventory methods

Combined and triaged four cohort lists

Among 205 studies and sites

- Commonly cited, established
- Currently active
- Participate in collaborations, consortia
- Understudied populations
- Internet, PI contacts

Three major types of resources



Longitudinal research studies: (53)

5 million people

Research oriented healthcare systems (10)

7.3 million people

Hospital-based cohort studies (6)

1 million

Overall

69 studies and health systems

~13.3 million people

11 studies <1,000 people

22 studies >100,000

Median 51,000

Funding sources

NIH: 6.3 million participants

NCI, NHLBI, NIDDK, NIEHS, NIA, NICHD, NIAID, NIMH and others

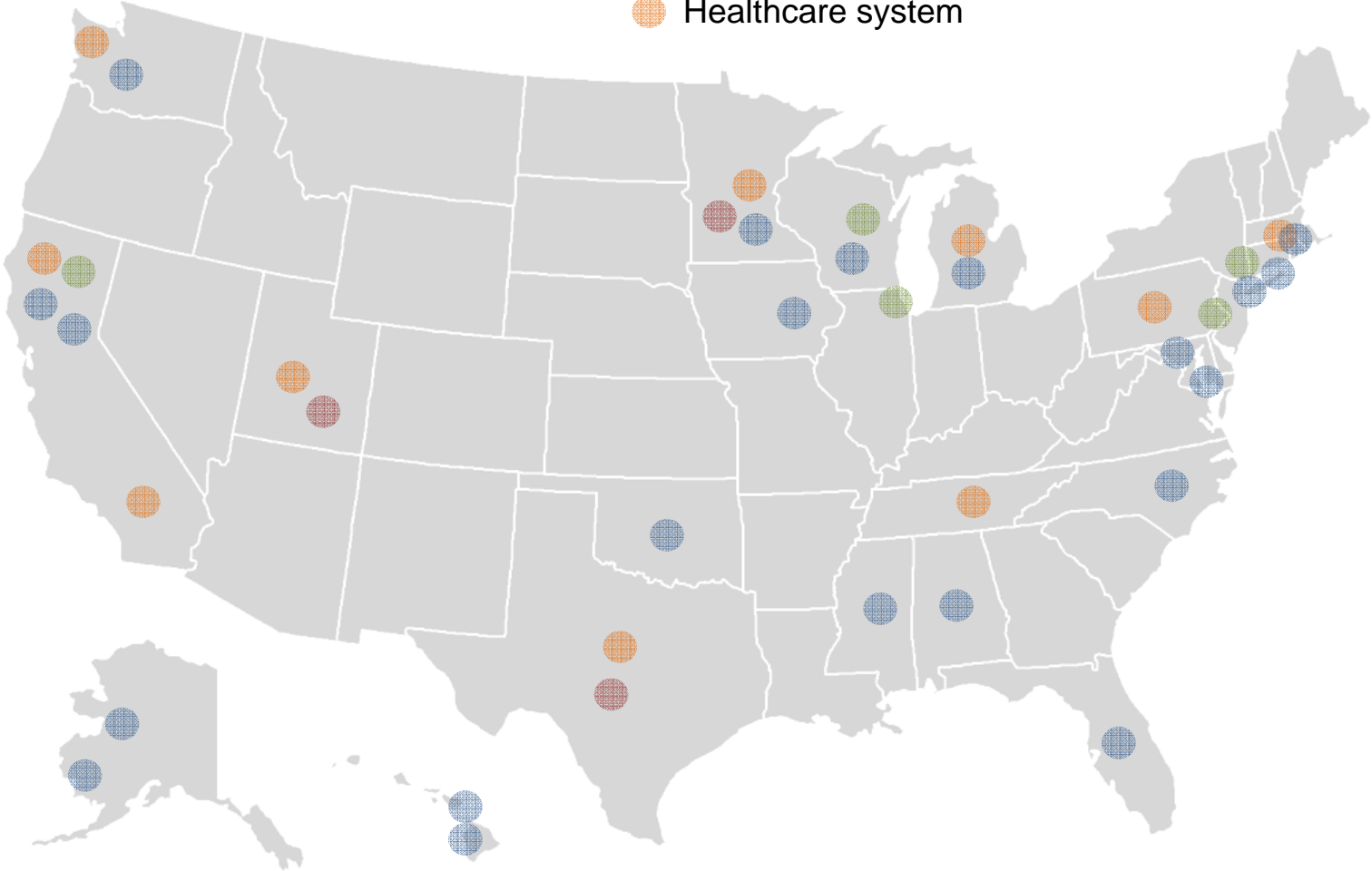
2.8 million people in longitudinal studies

Others: 7 million

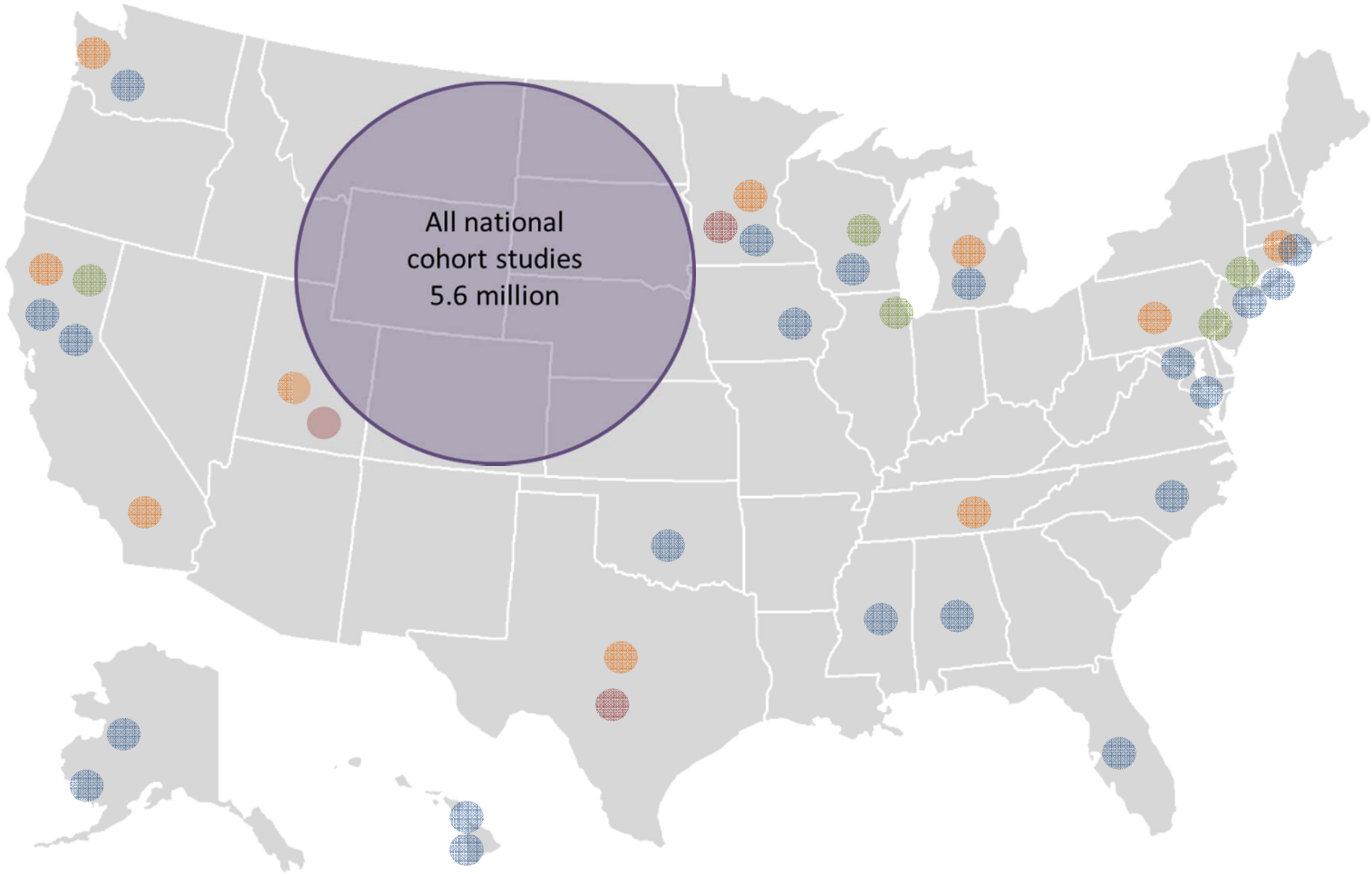
DoD:	220,000
VA:	330,000
Am. Cancer Society:	2.7 million
Health systems:	3.9 million

Geography

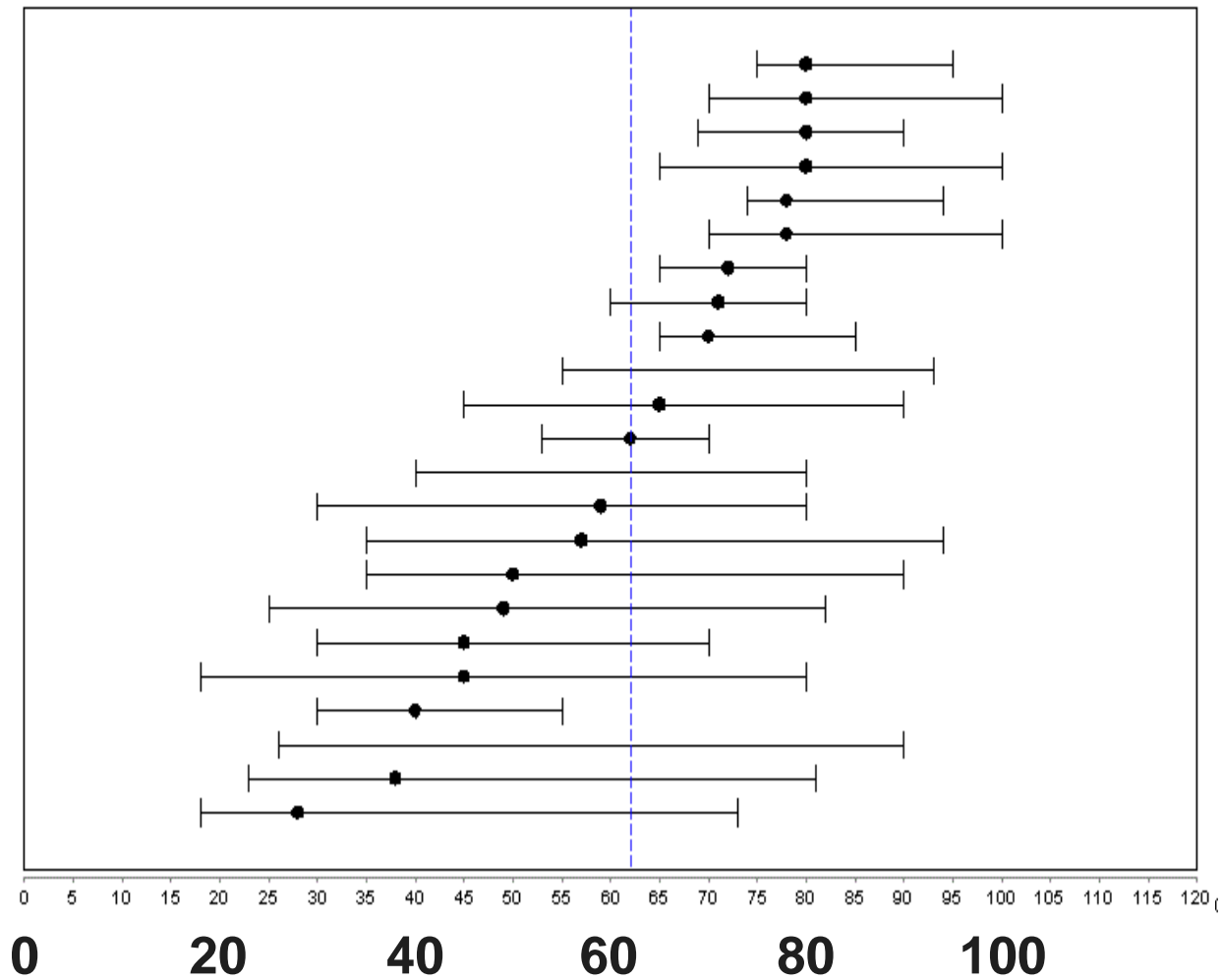
- NIH-funded longitudinal study
- Other longitudinal study
- Hospital based cohort
- Healthcare system



Geography



Current age distributions of 29 epi cohorts



Under-served and under-studied groups

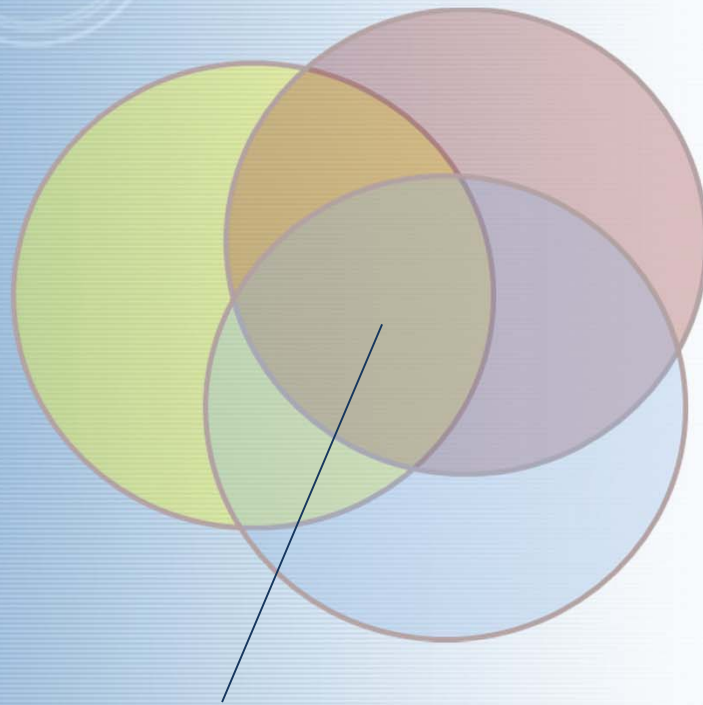


African Americans:	317,000+
Latinos:	116,000+
Asian Americans:	90,000+
Native Hawaiians & Pacific Islanders	18,000+
American Indians & Alaska Natives	12,700+
Rural	90,000+

Criteria for incorporation: Active Follow-Up or Ability to Recontact

- 58 of 69 studies
- 9.4 m people total
- 6.7 m from hospital studies & health systems

Follow-up/
recontact



Biospecimens available

Broad sharing of data

**best case:
2.8m - 6.1m people with samples
32 studies and systems**

Value-added characteristics

Broad consent

36% have it

Genetics

2 in 3 have done some genotyping

One quarter have done exome or genome sequencing

EMR/Medicare linkage

one-third of longitudinal cohorts; half linked to Medicare
most/all health systems

Additional potential resources

Other studies

Other Large Scale Consortia

HMO Research Network: 12 million pts

NCI Cohort Consortium: 4 million pts

PCORNet goal: 11 million pts

Independent large scale efforts

PatientsLikeMe, 23andMe, Sage Bionetworks, Google+
1-2 million people

Biobanks- Henderson GE and Cadigan RJ, UNC

120+ w/ broad consent, ability to recontact

1.6 million people

From Cohort PIs:

- **Enthusiasm for PM Cohort**
- **Re-consent is possible but impractical**
- **Already experienced harmonizing**
- **Engagement will be needed**

Thanks to:

Rebecca Baker

Allison Lea

Stephanie Devaney

Teri Manolio

Joanne Elena

Scott Rogers

Study PIs