

# NIH-Funded Primary and Secondary Prevention Research in Humans: A Portfolio Analysis

David M. Murray, Ph.D.

Associate Director for Prevention

Director, Office of Disease Prevention

Council of Councils Meeting

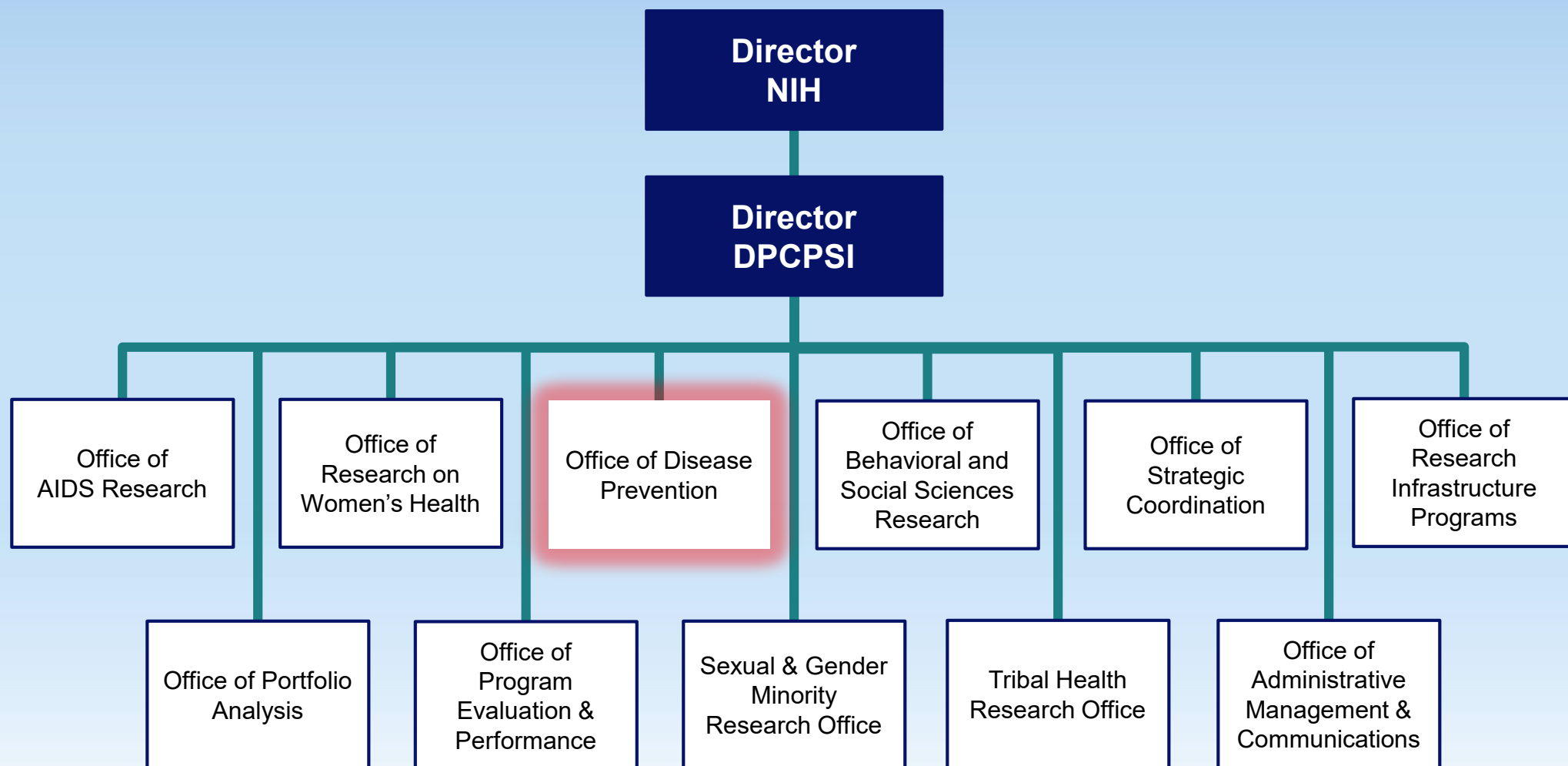
January 25, 2019



National Institutes of Health  
*Office of Disease Prevention*

# The Office of Disease Prevention

Division of Program Coordination, Planning, and Strategic Initiatives



# Mission

- To improve the public health by increasing the scope, quality, dissemination, and impact of prevention research supported by NIH.
- Provide leadership for the development, coordination, and implementation of prevention research in collaboration with ICs and other partners.





## The Office of Disease Prevention Strategic Plan FY 2019–2023

# ODP Strategic Priorities (2019-2023)



**Conduct Portfolio  
Analysis & Impact  
Assessment**



**Identify Research  
Gaps**



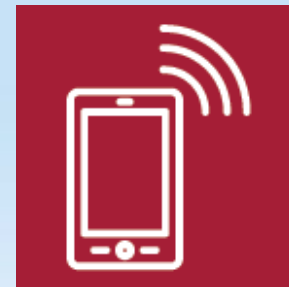
**Improve Research  
Methods**



**Promote  
Collaborative  
Research**



**Advance Tobacco  
Regulatory  
Science**



**Communicate  
Efforts &  
Findings**

# Strategic Priority I

- Systematically monitor NIH investments in prevention research and assess the progress and results of that research.
  - ODP defines prevention research to include primary and secondary prevention in humans, together with relevant methods development.
  - ODP's definition includes research designed to promote health; to prevent onset of disease, disorders, conditions, or injuries; and to detect, and prevent the progression of, asymptomatic disease.
  - Prevention research includes studies for:
    - Identification and assessment of risk and protective factors,
    - Screening and identification of individuals and groups at risk,
    - Development and evaluation of interventions to reduce risk,
    - Translation, implementation, and dissemination of effective, preventive interventions into practice, and
    - Development of methods to support prevention research.

# Strategic Priority I

- Systematically monitor NIH investments in prevention research and assess the progress and results of that research.
  - RCDC defines prevention research more broadly.
  - For ODP, RCDC methods...
    - Have unknown sensitivity and specificity for prevention research.
    - May not accurately characterize levels or trends for awards or dollars.
    - May not accurately identify areas with inadequate support.
    - In addition, RCDC methods provide inadequate detail on features like outcome, exposure, study type, design, etc.
- RCDC cannot be used to address this ODP priority.



# Selection of Activity Codes

- ODP worked with staff from many ICs to identify activity codes likely to support NIH prevention research that met ODP's definition.
  - Basic and preclinical research were excluded.
  - Awards for community services, facilities, infrastructure, loan repayment, meetings, planning, and training were excluded.
  - Intramural research was excluded to focus on extramural research.
  - Contracts proved too difficult to code using our methods.
  - Methodological research was included only if it yielded products that were applicable to prevention research without additional development.
- We included all remaining R, P, and U activity codes with at least 500 awards across FY12-17 or at least \$500M awarded across FY12-17.
- Several of these activity codes involved awards with multiple subprojects; as a result, we sampled Application IDs (Appl IDs) instead of awards.

# 12 Activity Codes Included in the Portfolio Analysis

Code	Total Awards FY12-17	Total Appl IDs FY12-17
R01	32176	32190
R21	11992	11992
R43	3439	3439
R03	2932	2932
U01	2188	2187
R56	1943	1945
R44	1901	1902
P01	534	3755
U54	328	1939
P50	268	2143
U19	203	1328
<b>UM1</b>	<b>200</b>	<b>232</b>

Code	Total Costs FY12-17
R01	\$14500 M
R21	\$2600 M
U01	\$2000 M
R44	\$1200 M
P01	\$996 M
R56	\$815 M
R43	\$780 M
U54	\$747 M
UM1	\$742 M
P50	\$536 M
U19	\$527 M
<b>R03</b>	<b>\$259 M</b>

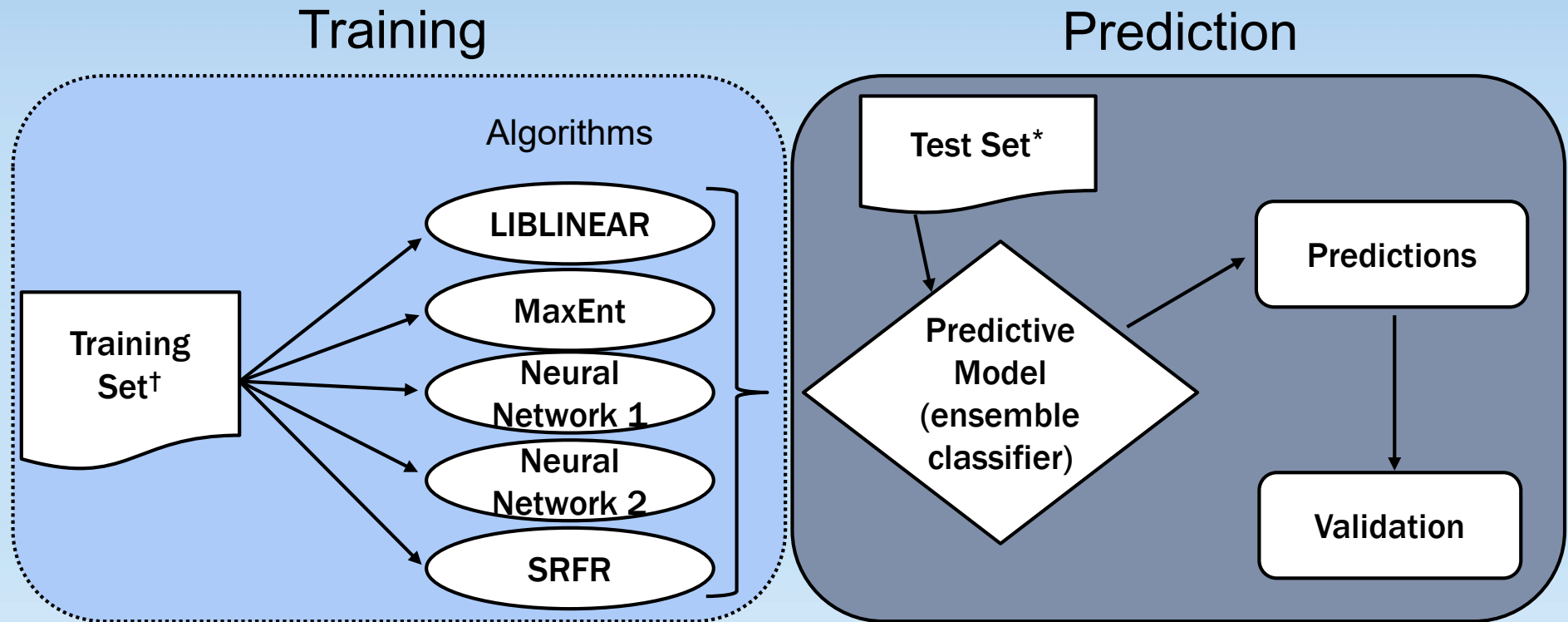
- We considered all Type I, 2, and 9 awards and Application IDs from FY12-17 made using these activity codes.

# Portfolio Coverage by These Activity Codes

	All Activity Codes	R, P, U Activity Codes	Research R, P, U Activity Codes	ODP's Selected Activity Codes	% Research R, P, U Activity Codes
Total Awards	111,626	68,757	63,381	58,104	91.7%
Total Costs	\$57.5 B	\$32.6 B	\$30.6 B	\$25.7 B	84.1%

- All figures based on Type I, 2, and 9 awards from FY12-17, excluding parent awards for projects with sub-awards to avoid double counting.

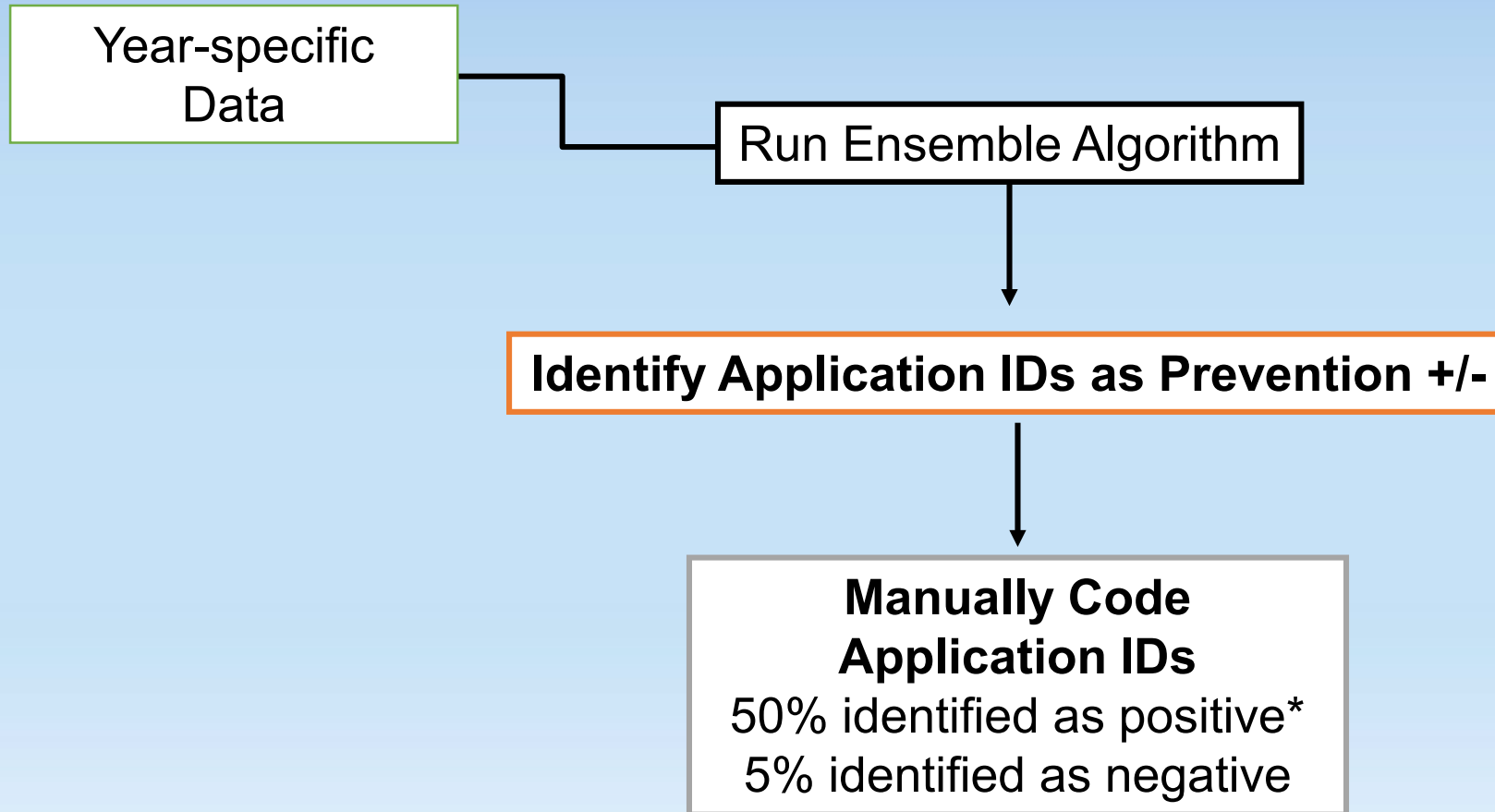
# Machine Learning Used to Select Application IDs



† Each training set was derived from manually coded 1R01s from the previous fiscal year (FY)

\* Each test set included the Application IDs from activity codes of interest for that specific FY

# Sampling of Application IDs



\* 100% of 1R01s identified as positive were manually coded

# Coding Based on a Prevention Research Taxonomy

- A classification system to characterize projects or subprojects on:
  - Study Focus
    - Rationale
    - Exposures
    - Outcomes
  - Population focus
  - Study design/purpose
  - Prevention research category
- 128 topics, 29-page protocol
  - Applied to title, abstract, public health significance
- Input from the PRCC

Rater: \_\_\_\_\_  
 Date: \_\_\_\_\_

**Prevention Taxonomy Form**  
**CHECK ALL THAT APPLY IN EACH COLUMN**  
**(TOPICS ARE NOT MUTUALLY EXCLUSIVE)**  
**See accompanying protocol for definitions and examples**

Appl ID: \_\_\_\_\_ PI Last Name: \_\_\_\_\_ Project Title: \_\_\_\_\_

Study focus	Rationale	Exposure	Outcome
1. Alcohol			
2. Alzheimer's disease			
3.			
4. Blood disorder			
5. Blood pressure			
6. Cancer			
7. Chemical/toxin			
8. Cholesterol			
9. Diabetes			
10. Diet/nutrition			
11. Education/counseling			
12. Firearms			
13. Gastrointestinal disease			
14. Genetics			
15. Healthcare delivery			
16. Heart disease			
17. HR quality of life			
18. Infectious disease			
19. Kidney disease			
20. Lung disease			
21. Maternal/paternal/child health			
22. Medication/device			
23. Mental health			
24. Microbiome			
25. Mortality			
26. Motor vehicle crash			
27. Musculoskeletal disease			
28. Neurological disease (not Alzheimer's)			
29. Obesity			
30. Physical activity			
31. Policy/built environment			
32. Pneumonia/influenza			
33. Sexual behavior			
34. Stress			
35. Stroke			
36. Substance abuse			
37. Suicide			
38. Surgery			
39. Tobacco			
40. Unintentional injuries			
41. Vaccine			
42. Violence			
43. Other or unclear			

Population focus	
1. Incarcerated/institutionalized	
2. LGBTI	
3. Low income	
4. Military/veterans	
5. Older adults/elderly	
6. People with disabilities	
7. Pregnant and/or post-partum women	
8. Rural	
9. Urban	
10. Youth (infants, children, adolescents)	
11. Other or unclear	

Study design/purpose	
1. Analysis of existing data	
2. Methods research	
3. Non-randomized intervention study	
4. Observational study	
5. Pilot/feasibility/proof-of-concept/safety	
6. Randomized intervention study	
7. Other or unclear	

Prevention research category	
1. Preventing new health condition, promoting health in the general population, or identifying risk factors for a new health condition	
2. Screening for risk factor	
3. Screening for early disease	
4. Preventing progression of disease, preventing recurrence in those with a known health condition, identifying risk factors for progression or recurrence	
5. Methods research	
6. Other or unclear	

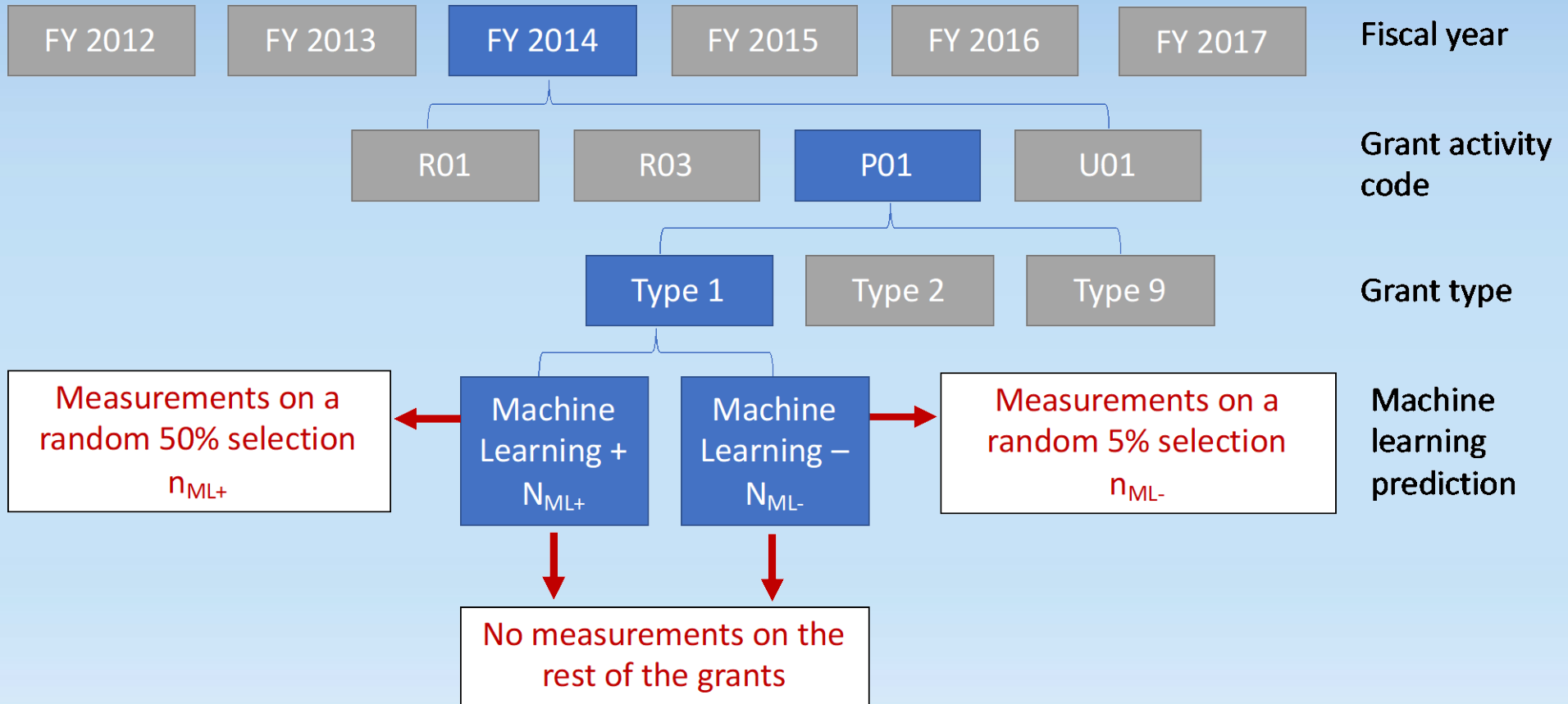
# Team-Based Coding



## ■ Coders

- Contract staff: MPH grads led by a PhD epidemiologist
- 2 months training in groups of 3-4
- Overseen by ODP staff
- 3-person teams coded abstracts using iPads
  - Each person coded independently, then the team resolved disagreements to generate a set of consensus codes for each Application ID
  - ODP coded 10-20% of the abstracts weekly for QC using the same methods
  - ODP reconciled discrepancies with the contract coding teams
- Average interrater agreement was 0.86.

# Weighting



Weights for FY14, P01, type 1, Machine Learning + =  $N_{ML+} / n_{ML+}$

Weights for FY14, P01, type 1, Machine Learning - =  $N_{ML-} / n_{ML-}$

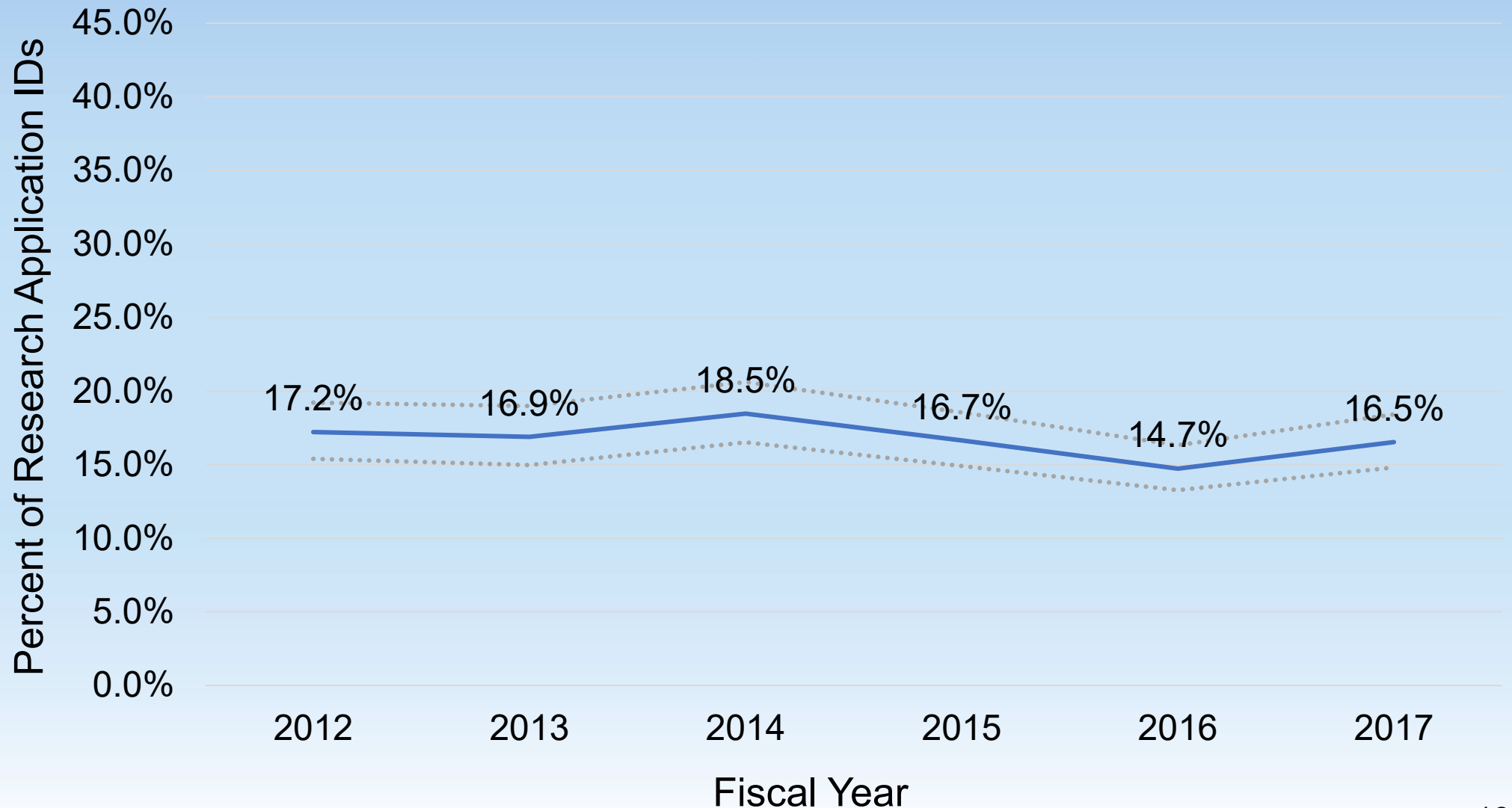
# Sensitivity and Specificity of Machine Learning Ensemble for Identifying Prevention Research

	Sensitivity	Specificity
Machine Learning (ML): 1R01s	75.6%	93.3%
ML: All Activity Codes	70.3%	91.1%
ML: All Activity Codes Excluding Cores	72.7%	92.9%

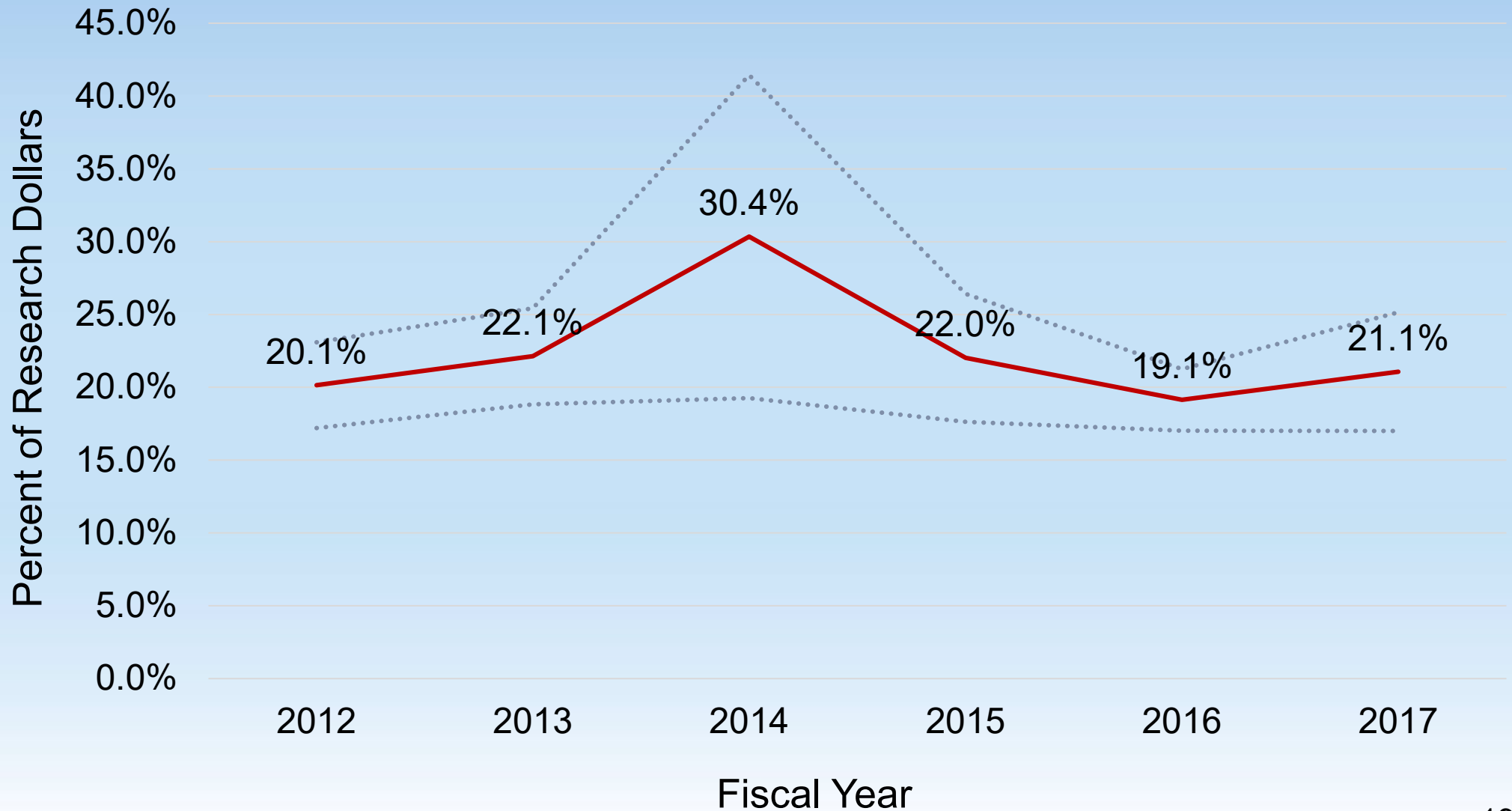
- The most recent iteration of the machine learning ensemble, developed after this work was completed, was much improved:
  - Sensitivity 95.8%
  - Specificity 93.0%
- This new ensemble will be used going forward.

# Overall Results

# Primary and Secondary Prevention Research in Humans: FY12-17



# Primary and Secondary Prevention Research in Humans: FY12-17



# Primary and Secondary Prevention Research in Humans by Activity Code FY12-17

Activity Code	Research Appl IDs	% of Research Appl IDs (95% CI)
P01	3,226	8.4% (6.0-11.8)
P50	1,896	18.3% (13.3-24.6)
R01	32,190	16.8% (15.8-17.9)
R03	2,932	26.9% (23.1-31.2)
R21	11,992	15.0% (13.4-16.7)
R43	3,439	9.6% (7.4-12.3)
R44	1,902	11.4% (8.6-15.0)
R56	1,945	13.3% (9.6-18.2)
U01	2,187	43.8% (37.7-50.1)
U19	1,130	12.9% (8.2-19.8)
U54	1,680	13.8% (10.2-18.3)
UM1	225	33.6% (18.5-53.1)

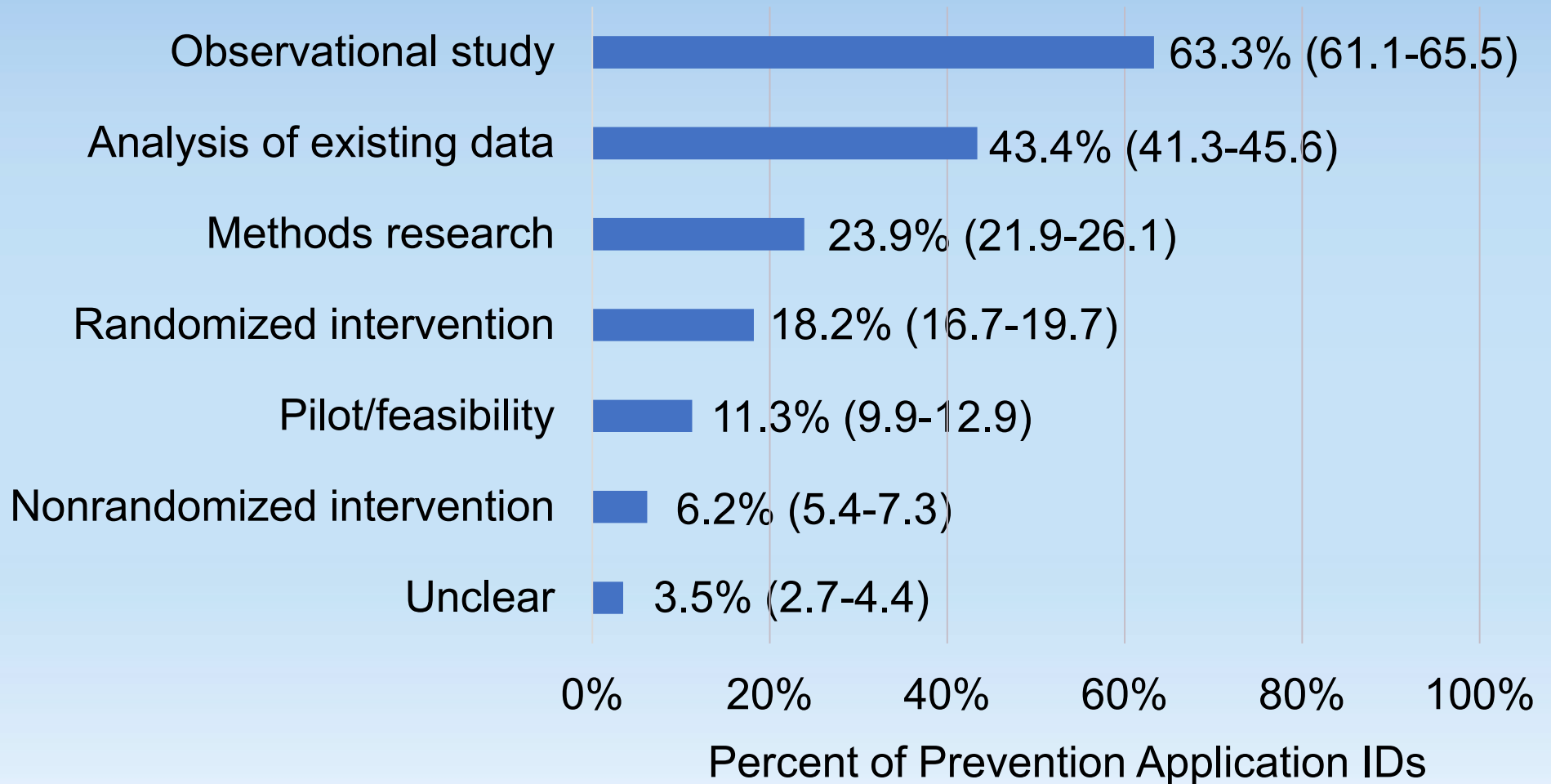
# Results for Application IDs Coded as Primary or Secondary Prevention Research in Humans

# Distribution of Gender and Minority Inclusion Codes

Gender Inclusion Code	% Prevention Appl IDs (95% CI)
Both Genders	65.7% (53.6-67.8)
Only Women	6.3% (5.5-7.2)
Only Men	3.2% (2.5-4.0)
Gender Unknown	0.4% (0.2-0.9)
Missing	24.4% (22.5-26.5)

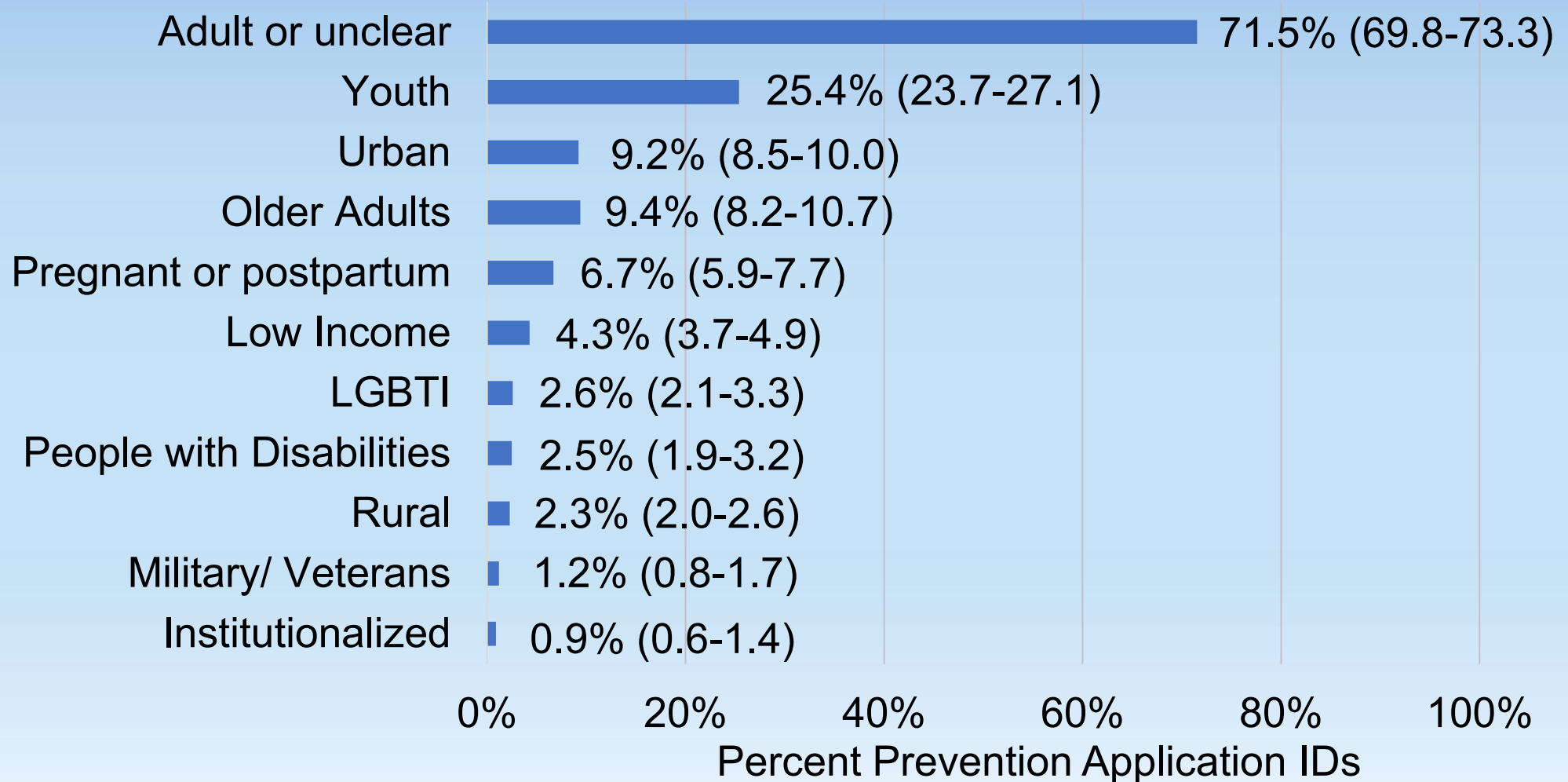
Minority Inclusion Code	% Prevention Appl IDs (95% CI)
Minorities and Non-minorities	64.3% (62.2-66.4)
Only Minorities	4.1% (3.6-4.8)
Only Non-minorities	1.3% (0.8-2.0)
Minority Representation Unknown	0.5% (0.3-1.0)
Missing	29.8% (27.8-31.8)

# Study Designs FY12-17



■ Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Populations Studied FY12-17



■ Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Study Rationales FY12-17

Topic	% of Prevention Appl IDs (95% CI)	Topic	% of Prevention Appl IDs (95% CI)
Mortality	28.0% (26.0-30.1)	Diabetes	5.5% (4.7-6.4)
Cancer	17.8% (16.1-19.7)	Alcohol	5.7% (4.9-6.6)
Infectious Disease	17.8% (16.2-19.6)	Lung Disease	3.7% (2.9-4.8)
MPCH	13.2% (11.9-14.7)	Alzheimer's Disease	3.1% (2.3-4.2)
Heart Disease	10.2% (9.0-11.6)	Kidney Disease	2.8% (2.0-3.8)
Mental Health	10.2% (9.0-11.5)	Musculoskeletal Disease	2.6% (1.9-3.4)
Stroke	8.6% (7.5-9.8)	Gastrointestinal Disease	2.7% (1.9-3.9)
Substance Abuse	9.1% (8.0-10.2)	Unintentional Injuries	2.0% (1.5-2.6)
Neurological Disease	8.0% (6.7-9.4)	Suicide	1.3% (1.0-1.8)
Obesity	7.3% (6.5-8.2)	Pneumonia/Influenza	0.6% (0.3-1.1)
Tobacco	6.5% (5.5-7.6)	Blood Disorder	0.5% (0.4-0.7)

- Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Study Exposures FY12-17

Topic	% of Prevention Appl IDs (95% CI)	
Genetics	26.9%	(24.8-29.0)
Education/Counseling	14.3%	(13.2-15.4)
Medication/Device	9.8%	(8.4-11.5)
Diet/nutrition	5.2%	(4.4-6.1)
Healthcare Delivery	4.3%	(3.6-5.0)
Infectious Disease	3.8%	(3.0-4.6)
Chemical/Toxin	3.4%	(2.7-4.1)
Physical Activity	2.4%	(2.1-2.8)
Tobacco	2.6%	(2.1-3.4)
Substance Abuse	2.2%	(1.7-2.6)
Stress	2.1%	(1.6-2.6)
Microbiome	1.8%	(1.3-2.6)
Mental Health	1.7%	(1.3-2.1)
Alcohol	1.6%	(1.4-2.0)

- Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Study Outcomes FY12-17

Topic	% of Prevention Appl IDs (95% CI)	Topic	% of Prevention Appl IDs (95% CI)
Cancer	11.8% (10.4-13.4)	Mortality	4.9% (4.1-5.8)
Infectious Disease	9.9% (8.7-11.2)	Alcohol	4.8% (4.1-5.6)
Healthcare Delivery	8.2% (7.3-9.2)	Obesity	4.5% (4.0-5.1)
Mental Health	7.3% (6.3-8.5)	Stroke	3.6% (3.0-4.4)
HRQOL	6.6% (5.6-7.7)	Genetics	3.6% (2.7-4.7)
Substance Abuse	6.2% (5.5-7.0)	Diet/Nutrition	3.5% (3.0-4.0)
Medication/Device	5.9% (4.7-7.3)	Physical Activity	3.3% (2.8-4.0)
Neurological Disease	5.6% (4.5-6.9)	Diabetes	3.1% (2.6-3.8)
Heart Disease	5.4% (4.5-6.6)	Lung Disease	2.8% (2.1-3.8)
Tobacco	5.1% (4.3-6.0)	Sexual Behavior	2.3% (2.0-2.6)

- Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Comparison to CDC and GBD

#	2015 Leading Causes of Death	% <sup>+</sup>	#	2016 Leading Risk Factors	% <sup>+</sup>
1	Heart Disease	10.9%	1	Tobacco	7.7%
2	Cancer	18.3%	2	Obesity	7.8%
3	Chronic Lower Respiratory	-	3	Dietary Factors	7.8%
4	Accidents	2.7%	4	Alcohol or Drug Use	14.2%
5	Stroke	9.2%	5	High Fasting Plasma Glucose	-
6	Alzheimer's Disease	3.9%	6	High Blood Pressure	2.7%
7	Diabetes	6.4%	7	High Total Cholesterol	1.8%
8	Influenza and Pneumonia	0.6%	8	Impaired Kidney Function	-
9	Kidney Disease	2.9%	9	Occupational Risks	-
10	Suicide	1.4%	10	Air Pollution	-

Based on CDC data

Based on GBD data

<sup>+</sup>Percent of prevention Application IDs calculated using coding on Rationales, Exposures, and Outcomes

# Proportion of Prevention Application IDs That Used Each Study Design By Outcome

#	Outcome	Analysis of Existing Data	Methods Research	Nonrandomized Intervention	Observational Study	Pilot / Feasibility	Randomized Intervention	Unclear
1	Heart disease	69.8%	16.0%	1.7%	81.0%	6.9%	20.0%	0.8%
2	Cancer	56.2%	41.5%	1.8%	68.4%	6.2%	10.5%	4.1%
5	Stroke	76.6%	12.8%	3.3%	73.0%	4.8%	29.3%	2.0%
6	Alzheimer's disease	75.8%	38.1%	0.5%	89.2%	2.8%	3.2%	0.9%
7	Diabetes	60.3%	7.6%	4.7%	57.8%	7.3%	39.5%	2.3%
8	Pneumonia or flu	72.8%	62.7%	0.0%	39.4%	0.0%	10.1%	0.0%
9	Kidney disease	69.9%	9.6%	0.4%	94.1%	1.7%	7.2%	2.1%
10	Suicide	42.1%	9.5%	9.3%	61.2%	9.4%	30.9%	1.4%

- For example, among prevention Application IDs with Cancer as an outcome, 10.5% included a randomized trial, while 68.4% included an observational study.
- Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Proportion of Prevention Application IDs That Used Each Study Design By Outcome

#	Outcome	Analysis of Existing Data	Methods Research	Nonrandomized Intervention	Observational Study	Pilot / Feasibility	Randomized Intervention	Unclear
1	Tobacco	23.4%	9.5%	16.3%	34.6%	13.4%	38.4%	4.2%
2	Obesity	37.1%	9.6%	7.1%	40.8%	11.6%	50.8%	1.4%
3	Diet/nutrition	21.8%	16.0%	14.7%	31.2%	17.4%	52.7%	4.0%
4	Alcohol	37.9%	38.1%	17.1%	42.0%	15.3%	34.8%	2.5%
4	Substance use	38.6%	12.8%	14.7%	49.9%	19.8%	32.6%	3.0%
6	Blood pressure	42.4%	41.5%	17.2%	61.1%	9.2%	32.0%	2.0%
7	Cholesterol	55.7%	7.6%	22.4%	64.6%	9.0%	29.6%	0.0%
11	Physical activity	21.1%	62.7%	10.4%	27.1%	17.5%	57.3%	2.5%

- For example, of prevention Application IDs with Tobacco as an outcome, 38.4% included a randomized trial, while 34.6% included an observational study.
- Coders selected all categories that applied to each Application ID; percentages do not sum to 100%.

# Summary and Conclusions

- ODP coded 11,082 projects from 12 activity codes for FY12-17.
- Those codes represent 91.7% of all projects and 84.1% of all dollars used for research in NIH extramural grants and collaborative agreements.
- For those activity codes, primary and secondary prevention research in humans represented 16.7% of projects and 22.6% of dollars.
- 63.3% of the prevention projects included an observational study, 43.4% included an analysis of existing data, 23.9% included methods research.
- Only 18.2% included a randomized intervention, suggesting that only 3% of NIH resources for research are used for preventive intervention trials.
- Given that 74% of the variability in county-level life expectancy across the US is explained by established risk factors, it seems appropriate to devote a larger proportion of the NIH research portfolio to randomized prevention trials to address those risk factors.

# Published Results

## Am J Prev Medicine 2018;55(6)

### American Journal of Preventive Medicine

#### RESEARCH METHODS

## A Machine Learning Approach to Identify NIH-Funded Applied Prevention Research



Jennifer Villani, PhD, MPH,<sup>1</sup> Sheri D. Schully, PhD,<sup>1</sup> Payam Meyer, MS,<sup>2</sup>  
Ranell L. Myles, PhD, MPH, CHES,<sup>1</sup> Jocelyn A. Lee, PhD, MPH,<sup>1</sup> David M. Murray, PhD,<sup>1</sup>  
Ashley J. Vargas, PhD, MPH, RDN<sup>1</sup>

## NIH Primary and Secondary Prevention Research in Humans During 2012–2017



David M. Murray, PhD,<sup>1</sup> Jennifer Villani, PhD, MPH,<sup>1</sup> Ashley J. Vargas, PhD, MPH,<sup>1</sup>  
Jocelyn A. Lee, PhD, MPH,<sup>1</sup> Ranell L. Myles, PhD, MPH, CHES,<sup>1</sup> Jessica Y. Wu, PhD,<sup>2</sup>  
Patricia L. Mabry, PhD,<sup>3</sup> Sheri D. Schully, PhD<sup>1</sup>

# Next Steps for ODP

- We will work with colleagues across the ICOs to examine our data for their portfolio and to consider the implications of those findings for their prevention research going forward.
- We will make IC-specific data available to interested ICs.
- We will continue to share our data with RCDC, as it provides a rich source of validated results that can be used to improve their system.
- We will extend the application of the machine learning algorithms to many of the 128 topics and assess sensitivity and specificity.
  - We hope this will allow us to reduce the level of manual coding.
- We will assess the progress and results of primary and secondary research in humans using new tools available from OPA.
- We will continue to code new awards from FY18 and beyond.

# Acknowledgments

## **ODP SP1 Team**

Sheri Schully, Team Lead  
Kat Schwartz  
David Tilley  
Ashley Vargas  
Jennifer Villani

## **OPA Team**

George Santangelo  
Kirk Baker  
Paula Fearon  
Payam Meyer

## **IQ Solutions**

Shahina Akter  
Luis Ganoza Caballero  
Adeola Olufunmilade  
Richard Panzer  
Priti Patel  
Agnieszka Roman  
Taylor Walter  
Lamyaa Yousif  
Sundeeep Vikraman

## **Past ODP Members**

Patty Mabry  
Jocelyn Lee  
Ranell Myles  
Jessica Wu



**National Institutes of Health**  
*Office of Disease Prevention*