Data Analysis & Analytics at GAO

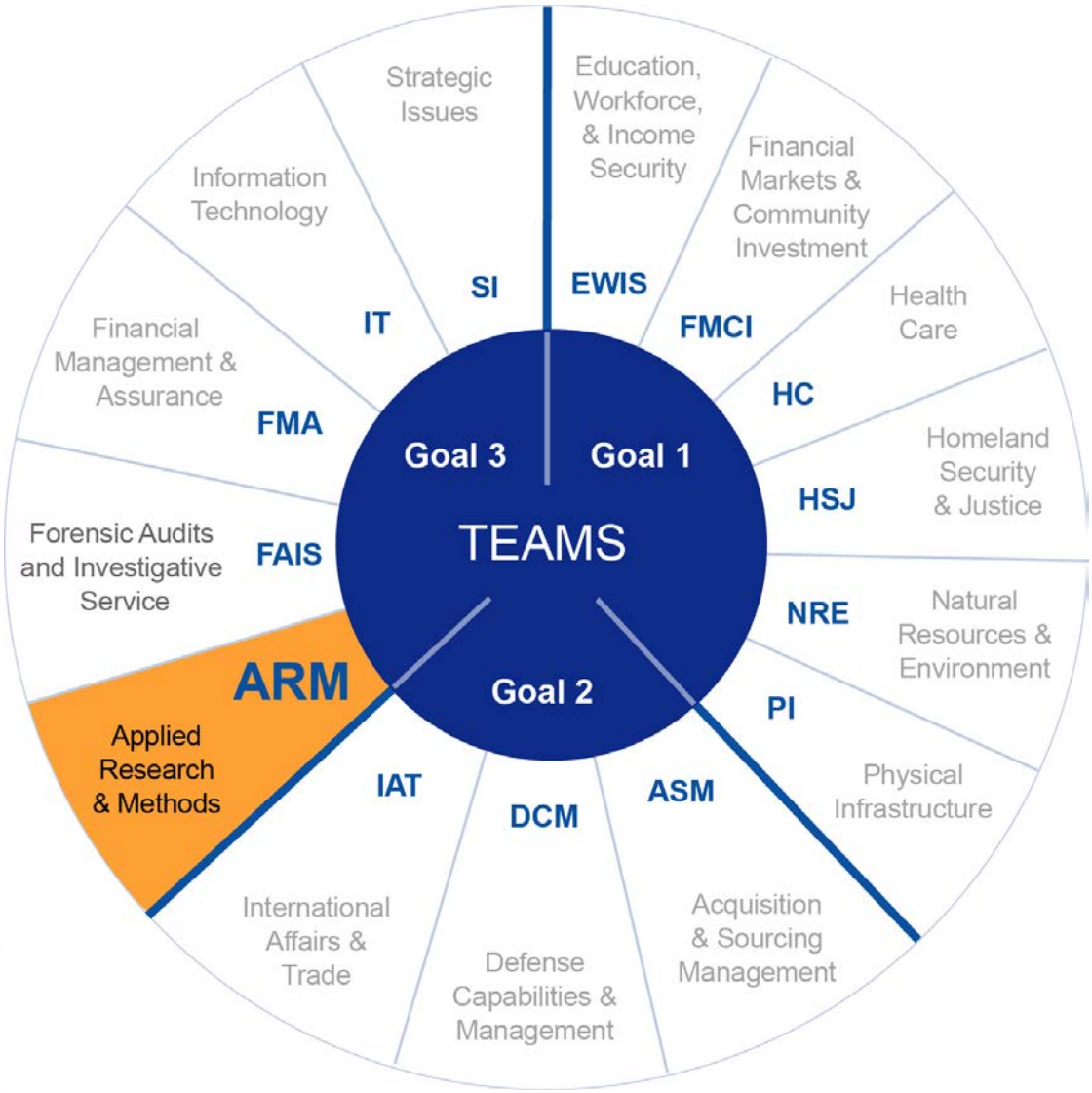
Melinda L. Cordero, Assistant Director
U.S. Government Accountability Office
Center for Enhanced Analytics

Agenda

- About GAO
- Data Analysis and Analytics at GAO
- Case Study

Key Facts to Know About GAO

- GAO provides Congress, the heads of executive agencies, and the public with timely, fact-based, non-partisan information to improve government and save taxpayers billions of dollars.
- Our work resulted in $75 billion in financial benefits to the government in FY 2015
- Most GAO work is done at the request of Congress, or is required by legislation.
- GAO's staff at all levels are civil servants; none are political appointees


TEAMS
Goal 3
Goal 1
Goal 2
SI
EWIS
IT
FMCI
FMA
HC
FAIS
HSJ
ARM
NRE
IAT
PI
DCM
ASM
Strategic Issues
Education, Workforce, & Income Security
Information Technology
Financial Markets & Community Investment
Financial Management & Assurance
Health Care
Forensic Audits and Investigative Service
Homeland Security & Justice
Applied Research & Methods
Natural Resources & Environment
International Affairs & Trade
Physical Infrastructure
Defense Capabilities & Management
Acquisition & Sourcing Management

Data Analysis and Analytics at GAO

- Most data management, analysis and analytics work centralized in ARM
- Specialists in data analysis, statistics, economics and modeling
- Engagement teams are multidisciplinary, specialists assigned as necessary depending on methods used

Analysis: Process

1. Mission team starts engagement
2. Methodologists usually consult on project design
3. Team and methodologists identify appropriate data
4. Data analysts and other specialists are assigned as needed
5. Team assesses data for reliability (GAO-09-680G)
6. Specialists conduct analyses
7. Team synthesizes analyses
8. Specialists consult on the reporting of results

Analysis: Commonly Used Data

- Existing Large Data Sets
  - Social Security's Death Master File
  - Health Care Data
  - Postal address database
  - Personnel and Payroll data
  - Census Data
  - Tax Data
- Other data as needed

Analysis: Tools

- Extensive use of SAS and Stata
- Limited use of open-source tools and scripting to conduct analyses and automate data-gathering
    - R for predictive advanced statistics, graphics and text
    - Anaconda Python for:
        - web scraping
        - data extraction, formatting and transformation

Analysis: Text analysis

- Experimenting with text analysis tools
- Text-based document classification
- Computer aided analysis of themes and similarities in text

Comptroller General Forum on Analytics in Government

- Planned for September 2016
- Focusing on innovative use of analytics in government (both oversight and programmatic)
- Speakers have been invited from industry, government and academia
- Document case studies of innovative and effective uses of analytics to address management challenges such as fraud and improper payments

GAO

Case Study

Using Data Analyses to Identify Gender Disparities in Research Grant Awards

STEM and gender disparities

- In fiscal year 2014, U.S. universities received nearly $25 billion in federal grant funding for Science, Technology, Engineering and Math (STEM) research.
- Studies show women are often underrepresented in STEM fields.
- Title IX requires gender equity in education programs receiving any federal financial assistance—including federally funded research at universities.

Women in STEM Research – GAO-16-14

- This report examines:
  - the extent to which differences exist in federal grant awards between women and men in STEM fields,
  - the extent to which federal agencies enforce Title IX at universities they fund for STEM research, and
  - possible actions federal agencies could take to address the underrepresentation of women in STEM research.

Methodology

- Reviewed six federal agencies that make STEM research grants:
  - Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA),
  - Department of Defense (DOD),
  - Department of Energy (DOE),
  - National Aeronautics and Space Administration (NASA),
  - National Institutes of Health (NIH), and
  - National Science Foundation (NSF)

Analysis Challenges Encountered

- Data necessary for analysis was not always available
  - Gender information was not consistently available
- Administrative data did not include demographic information that might explain differences between award rates
- Collection of data from various agencies
  - Data were not delivered at the same time
  - Data elements were not consistent across agencies
  - Data needed clean-up and standardization to allow consistent analysis

Variety of Data Sources

Data Source	Description
Agency Databases	Records of the applications and grants awarded by the six agencies
Survey of Doctorate Recipients (SDR)	NSF Survey of those who received a science, engineering or health research doctorate from a US academic institution
Social Security Administration's Name Database	Record of names associated with Social Security Numbers and reported gender

Researcher Information Collected

Table 4: Researcher Information Collected by Selected Agency Administrative Data Systems

Variables collected	U.S. Dept. of Agriculture[a]	Dept. of Defense[b]	Dept. of Energy[c]	NASA	National Institutes of Health	National Science Foundation
Sex	X				X	X
If tracked, approx. % of records with sex recorded[d]	86%				90-95%	85-90%
Race	X				X	X
Age/year of degree	X	1 of 8 components			X	X
Institution	X	X	X	X	X	X
Highest degree	X	2 of 8 components			X	X
Discipline (or subject area of grant)	X	4 of 8 components	3 of 4 components	X	X	
Co-Investigator information[e]	X	2 of 8 components		X	X	X

Source: GAO analysis of agency documents. | GAO-16-14

Results of Gender Assignment

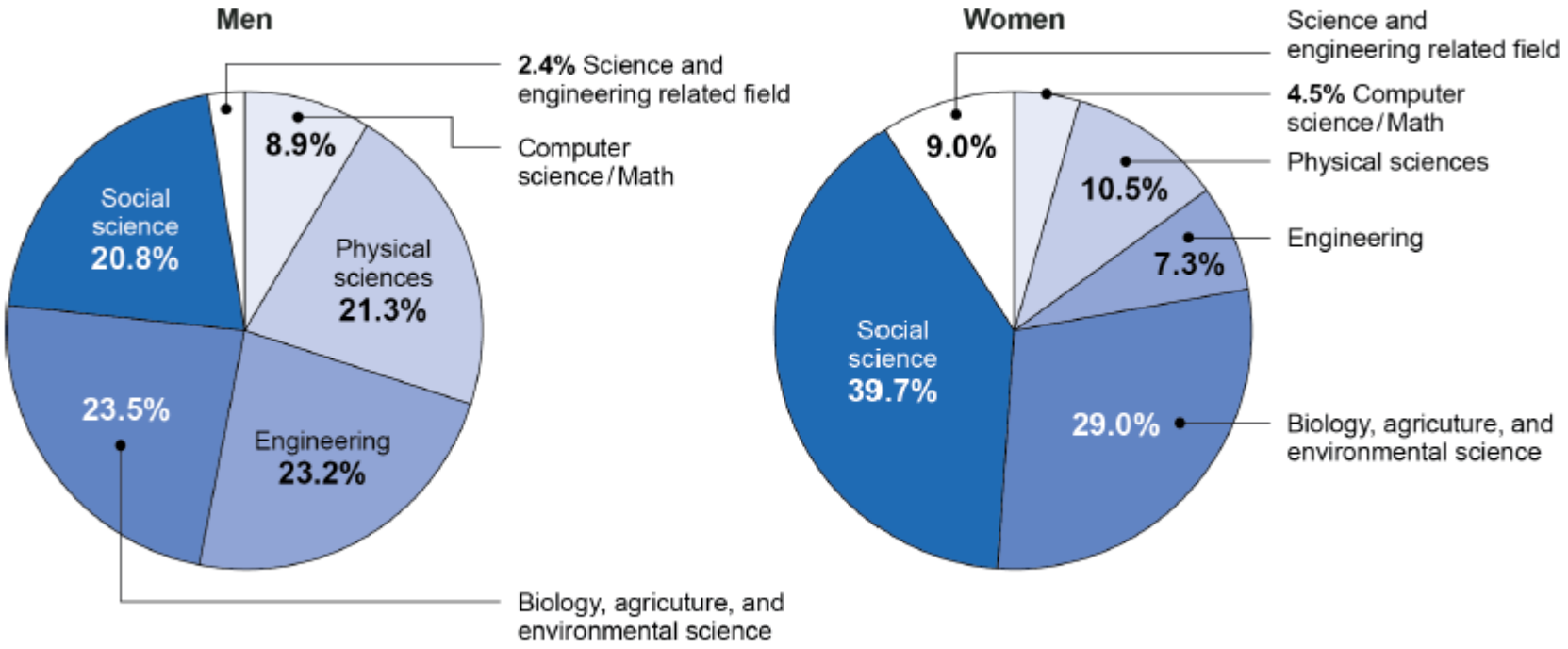
Table 5: Gender Name Match Percentages For Grant Proposal and Award Data at Department of Energy (DOE) and Department of Defense (DOD) Components

	Total Number of Records	Name assigned female	Name assigned male	No assignment
DOD-Army Research Office	**4567**	10%	62%	29%
DOD-Defense Threat Reduction Agency	**1560**	11%	62%	26%
DOD-Office of Naval Research	**9413**	12%	64%	24%
DOE-Advanced Research Projects Agency-Energy	**2078**	7%	61%	32%
DOE-Energy Efficiency and Renewable Energy	**1213**	10%	63%	27%
DOE-Office of Science	**13651**	13%	63%	25%

Source: GAO. | GAO-16-14

SDR: Gender Representation in Disciplines

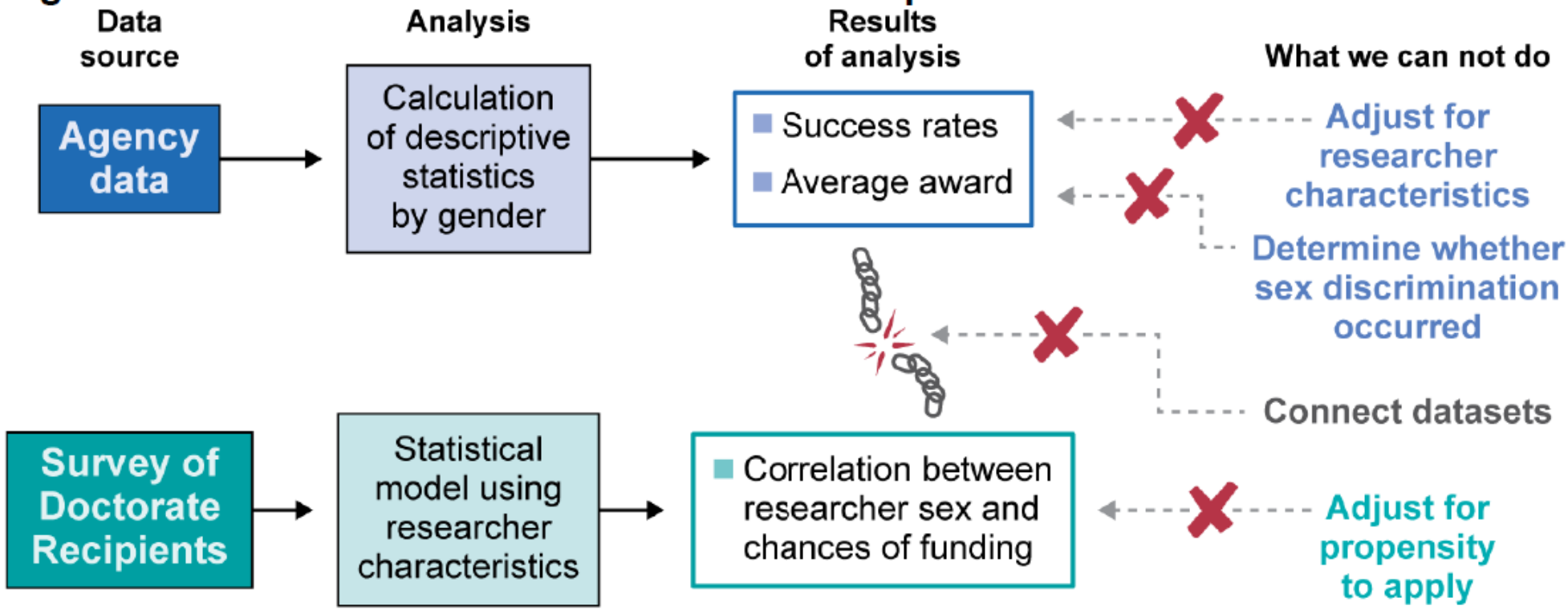
Figure 10: Male and Female PhDs in STEM Fields, by Discipline



Source: GAO analysis of 2013 Survey of Doctorate Recipients (SDR) data. | GAO-16-14

How Data are Used

Figure 9: How Data Sources Are Used In This Report



Source: GAO. | GAO-16-14

Grant Award Success Rates

Table 1: Summary of Federal Grant Award Success Rates for Women and Men at Selected Agencies

Agency	Women[a]-Average Success Rate Fiscal Year 2009-2013	Men[a]-Average Success Rate Fiscal Year 2009-2013
National Institutes of Health	25%	25%
National Science Foundation	26%	25%
Department of Agriculture-National Institute of Food and Agriculture	22%	21%

Source: GAO analysis of agency data. | GAO-16-14

Grant Award Success Rates

Table 2: Summary of Federal Grant Award Success Rates for Women and Men by Agency and Component at the Department of Defense (DOD), the Department of Energy (DOE), and NASA

Agency	Women-Average Success Rate Fiscal Year 2009-2013	Men-Average Success Rate Fiscal Year 2009-2013
Department of Defense		
Air Force Office of Scientific Research	Could not assess due to data limitations	
Air Force Research Labs[a]	Could not assess due to data limitations	
Army Research Office	48%	49%
Army Medical Command	Could not assess due to data limitations	
Basic Research Office	Could not assess due to data limitations	
Defense Advanced Research Projects Agency	Could not assess due to data limitations	
Defense Threat Reduction Agency	17%	17%
Office of Naval Research[b]	42%	48%
Department of Energy		
Advanced Research Projects Agency – Energy[c]	10%	7%
Energy Efficiency and Renewable Energy[d]	7%	5%
Nuclear Energy	Could not assess due to data limitations	
Office of Science	34%	41%
National Aeronautics and Space Administration	Could not assess due to data limitations	

Source: GAO analysis of agency administrative data. | GAO-16-14

Findings

- No evidence of disparity in grant application success rates between Women and Men at NIH, NSF, and USDA/NIFA
- Evidence of disparities varied or data were insufficient to analyze success rates at DOD, DOE, NASA

Success Rates between Women and Men at Federal STEM Grant-making Agencies

Source: GAO analysis of STEM grant data for fiscal years 2009 through 2013 from Health and Human Services (HHS) National Institutes of Health (NIH); National Science Foundation (NSF); Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA); Department of Defense (DOD); Department of Energy (DOE); and NASA. | GAO-16-14

Recommendations

- DOD, DOE, and NASA collect additional data;
  - *Retain complete records of pre-proposal, proposal, and award data, including a record of proposal disposition, in linked electronic files to facilitate aggregate, statistical analysis of the grant-making process, including the calculation of success rates.*
  - *Collecting demographic, education, and career information from applicants, on a voluntary basis, that is not available to proposal reviewers but is used for analysis of success rates.*
- DOD and HHS conduct Title IX compliance reviews; and
- DOJ facilitate information sharing among STEM agencies.

Conclusion

- Our collaborative approach was key to the success of this project. Policy experts, economists, a methodologist, a statistician and a data analyst all worked together.

- The diverse skills and abilities of the team allowed us to make the best possible use of data that had limitations and still make recommendations to the agencies to improve their data collection.

Questions?

- Full report is available at: http://gao.gov/products/GAO-*16-14*
- Email: Melinda Cordero - corderom@gao.gov

GAO on the Web
Web site: http://www.gao.gov/

Congressional Relations

Katherine Siggerud, Managing Director, siggerudk@gao.gov
(202) 512-4400, U.S. Government Accountability Office
441 G Street, NW, Room 7125, Washington, DC 20548

Public Affairs
Chuck Young, Managing Director, youngc1@gao.gov
(202) 512-4800, U.S. Government Accountability Office
441 G Street, NW, Room 7149, Washington, DC 20548

Strategic Planning and External Liaison
James-Christian Blockwood, Managing Director, blockwoodjc@gao.gov
(202) 512-4707 , U.S. Government Accountability Office
441 G Street, NW, Room 7814, Washington, DC 20548