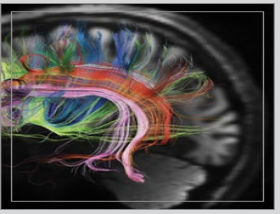



# NIH Rehabilitation Research Plan: The Way Ahead

Jennifer Jackson, Ph.D.

Alison Cernich, Ph.D.

Trans-NIH Medical Rehabilitation Coordinating Committee

National Institutes of Health

# Disclaimer

- All data presented for consideration are draft until financial management concurs
  - These data are close to final; review already occurred with the Trans NIH Medical Rehabilitation Coordinating Committee
  - Feedback is needed to determine if these are the types of data needed to track progress on the research plan, especially as the plan is being revised
  - Challenges and caveats are presented for consideration

# Agenda

- Overarching Strategy and Coding Process
- Algorithm
  - Goals and Development
  - Validation/Results
  - Conclusions and Future Directions
- 2018 Rehab Portfolio Analysis
- Clinical Impact and Interdisciplinary Collaborations
- Year-to-Year Trends (2015-2018)
- Challenges and Future Directions

# Overarching Strategy

- Baseline data taken from 2015 portfolio
  - Prior to plan publication
  - Allows for the year prior to serve as an “as is” for the rehabilitation portfolio
- Using only the Rehabilitation Research, Condition and Disease Category (RCDC)
  - Lists of projects available to the public (NIH RePORTER)
  - Official categories that are verified by the Institutes and Centers
  - Official dollars verified by Financial Management at NIH
  - Contains the Physical Rehabilitation Category
- Removed intramural projects

# Overarching Strategy (cont.)

- Each project categorized in two “Tiers”
  - Tier I – Based on a keyword approach for the 6 categories within the Rehabilitation Research Plan; primary and secondary codes are based on the primary and additional aims of each project
    - A: Rehabilitation Across the Lifespan
    - B: Community and Family
    - C: Technology Use and Development
    - D: Research Design and Methodology
    - E: Translational Science
    - F: Research Capacity and Infrastructure
  - Tier II – This is the phase of research for each project: basic, disease-related basic, applied (translational or clinical), infrastructure, and unable to categorize

# Coding Process

- NCMRR coded each IC's portfolio using the coding rules agreed upon by the group
  - Each project was coded separately by two coders and the codes were reconciled for agreement
  - Statistics computed for level of agreement for primary, secondary, and tier II
- Each IC received the reconciled portfolio and confirmed or revised NCMRR's proposed coding
- All ICs were integrated for the final analysis
- Each subsequent year will be coded for new grants only

# The Data: 2015-2018 (intramural projects excluded)

| Funding Institute/Center | 2015 Projects | New 2016 Projects | New 2017 Projects | Total Data for Auto Coding | New 2018 Projects |
|---|---|---|---|---|---|
| NINDS | 251 | 88 | 80 | 419 | 97 |
| NIDCD | 247 | 97 | 81 | 425 | 75 |
| NICHD | 224 | 61 | 57 | 342 | 74 |
| NIA | 127 | 56 | 74 | 257 | 71 |
| NCI | 101 | 52 | 39 | 192 | 55 |
| NIMH | 96 | 24 | 19 | 139 | 19 |
| NIAMS | 50 | 18 | 16 | 84 | 17 |
| NHLBI | 41 | 6 | 12 | 59 | 18 |
| NIGMS | 37 | 8 | 4 | 49 | 2 |
| NIBIB | 36 | 12 | 6 | 54 | 8 |
| NEI | 32 | 17 | 13 | 62 | 12 |
| NINR | 27 | 20 | 10 | 57 | 12 |
| NIDA | 23 | 8 | 9 | 40 | 13 |
| NCCIH | 18 | 11 | 10 | 39 | 12 |
| NIAAA | 12 | 5 | 3 | 20 | 8 |
| NIDDK | 11 | 1 | 2 | 14 | 8 |
| NIDCR | 9 | 2 | 1 | 12 | 6 |
| NIMHD | 5 | 2 | 5 | 12 | 5 |
| OD | 2 | 1 | 3 | 6 | 1 |
| FIC | 2 | 0 | 0 | 2 | 1 |
| NLM | 1 | 0 | 0 | 1 | 1 |
| NIEHS | 1 | 0 | 0 | 1 | 1 |
| NIAID | 1 | 1 | 0 | 2 | 0 |
| NCATS | 1 | 1 | 0 | 2 | 1 |
| Total | 1355 | 491 | 444 | 2290 | 517 |

# Algorithm Development: Minimum Agreement Goals

- Percentages of interrater agreement between two manual coders
- Agreement typically improved each year
- Automatic code should be at least as accurate (overall) as highest %
- Agreement goals:
  - 80% for Primary and Tier II categories
  - 60% for Secondary category

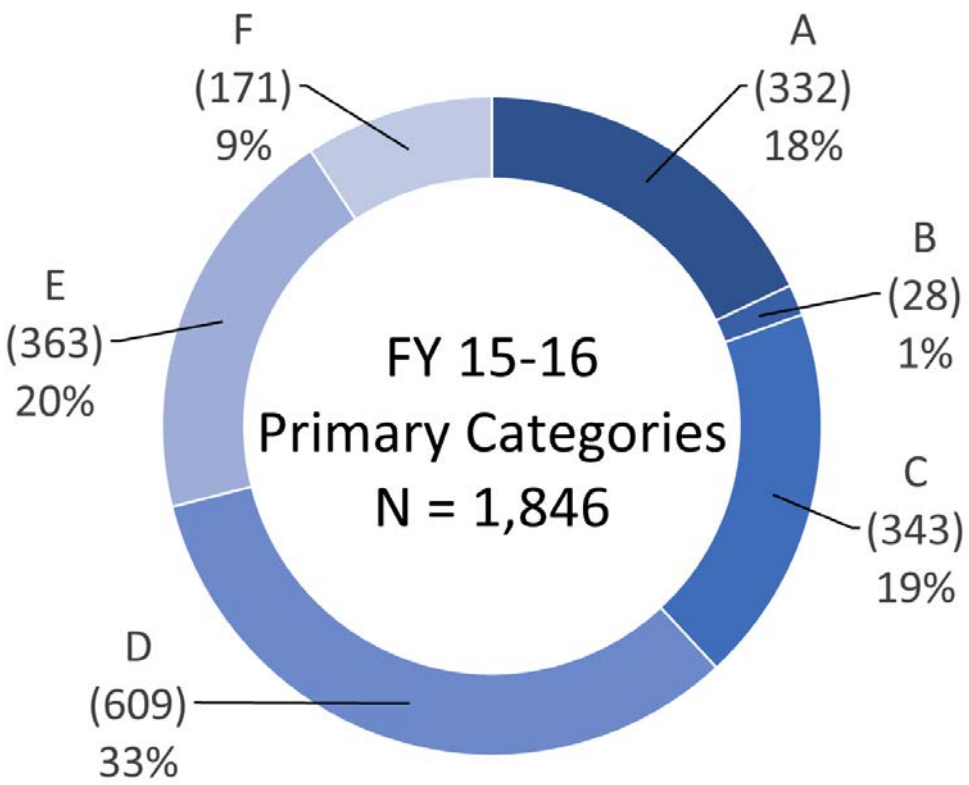
| | 2015 | 2016 | 2017 |
|---|---|---|---|
| Primary Agreement | 70.8% | 76.0% | 73.9% |
| Secondary Agreement | 44.3% | 48.7% | 52.0% |
| Tier II Agreement | 67.9% | 73.9% | 79.1% |
| Number of grants | 1355 | 491 | 444 |

Training Data

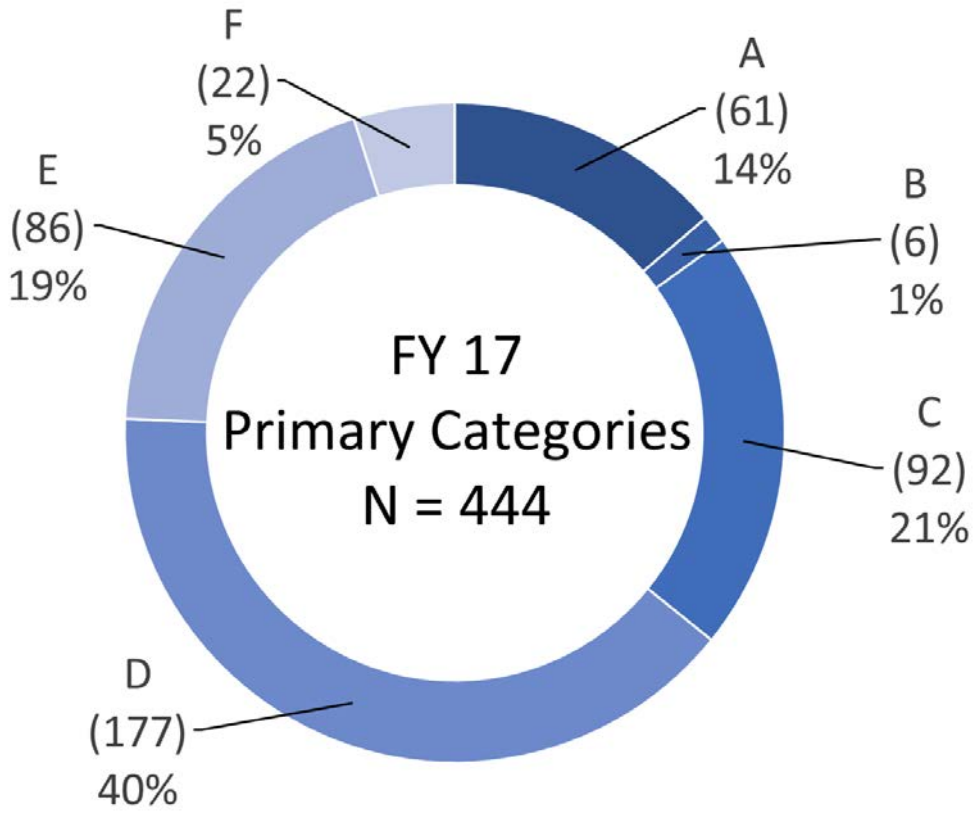
Testing Data

# Algorithm Development: Primary Data Sets

## Training Data Set

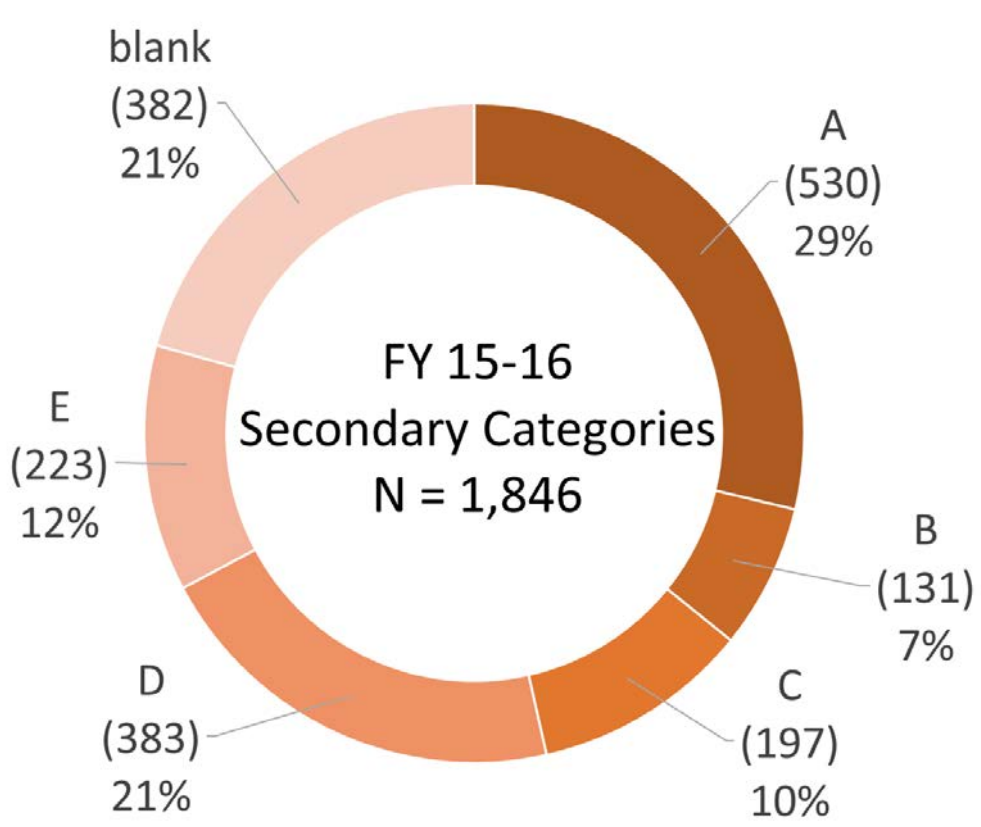



## Testing Data Set

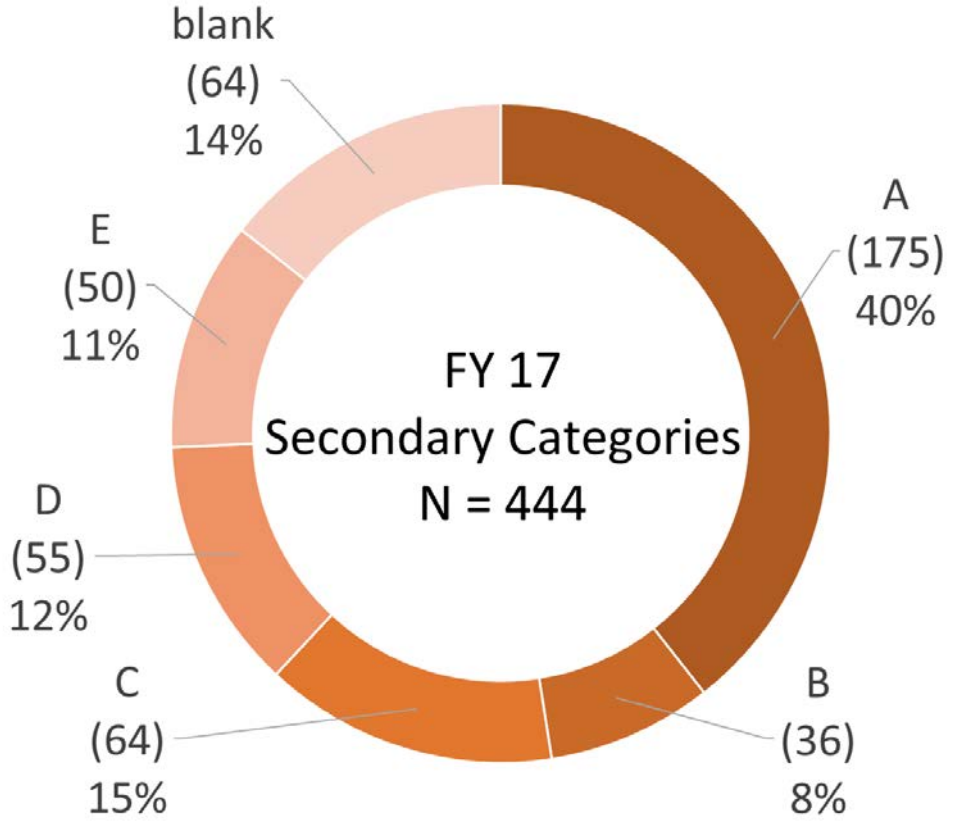


# Algorithm Development: Secondary Data Sets

## Training Data Set

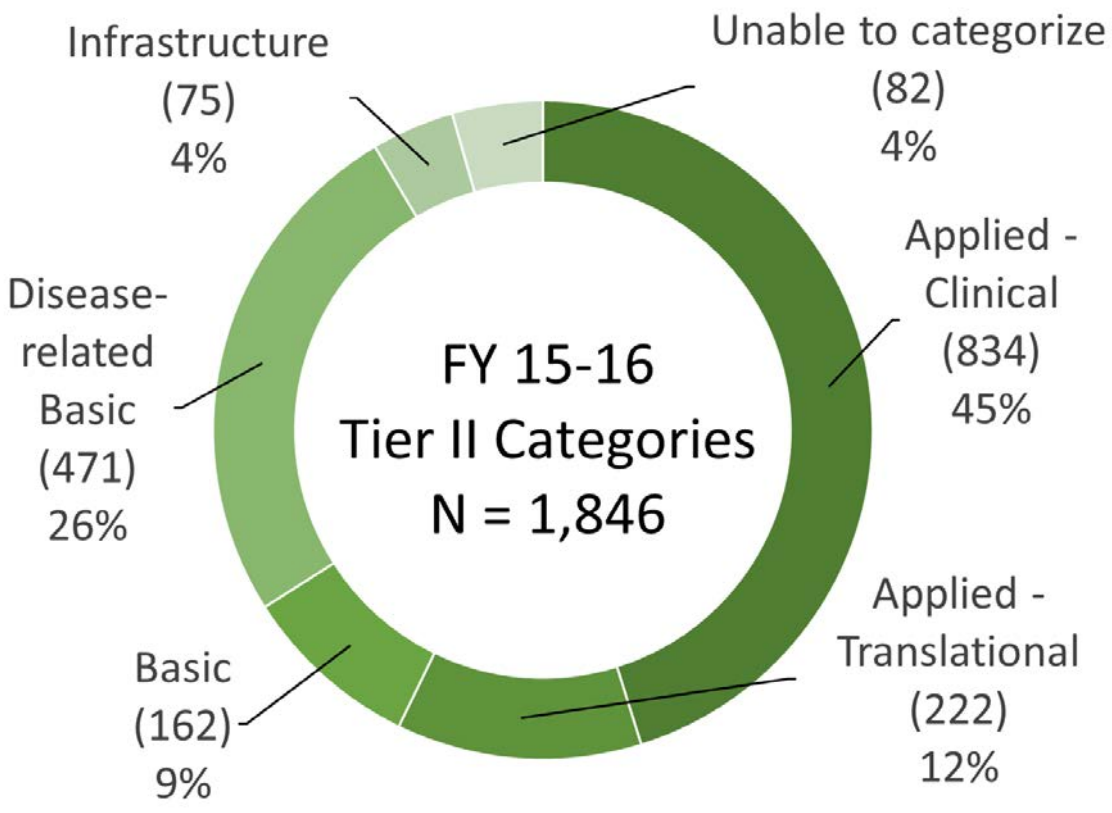



## Testing Data Set

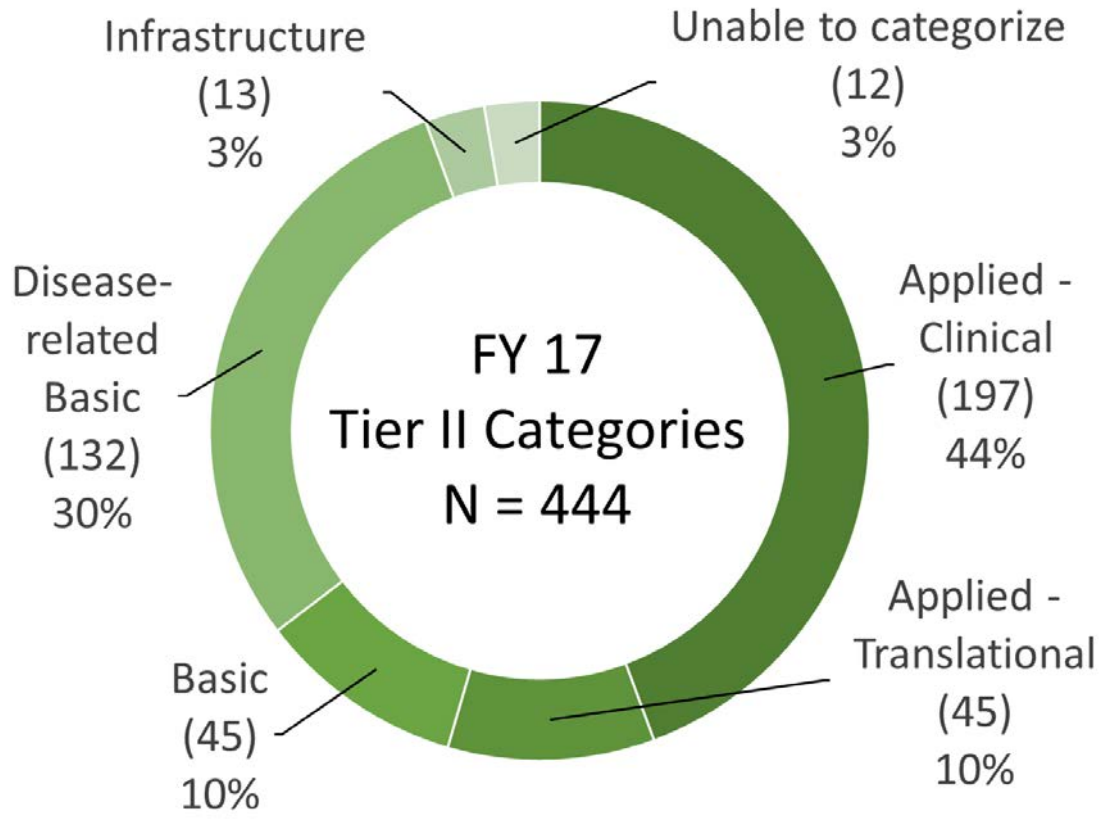


# Algorithm Development: Tier II Data Sets

## Training Data Set

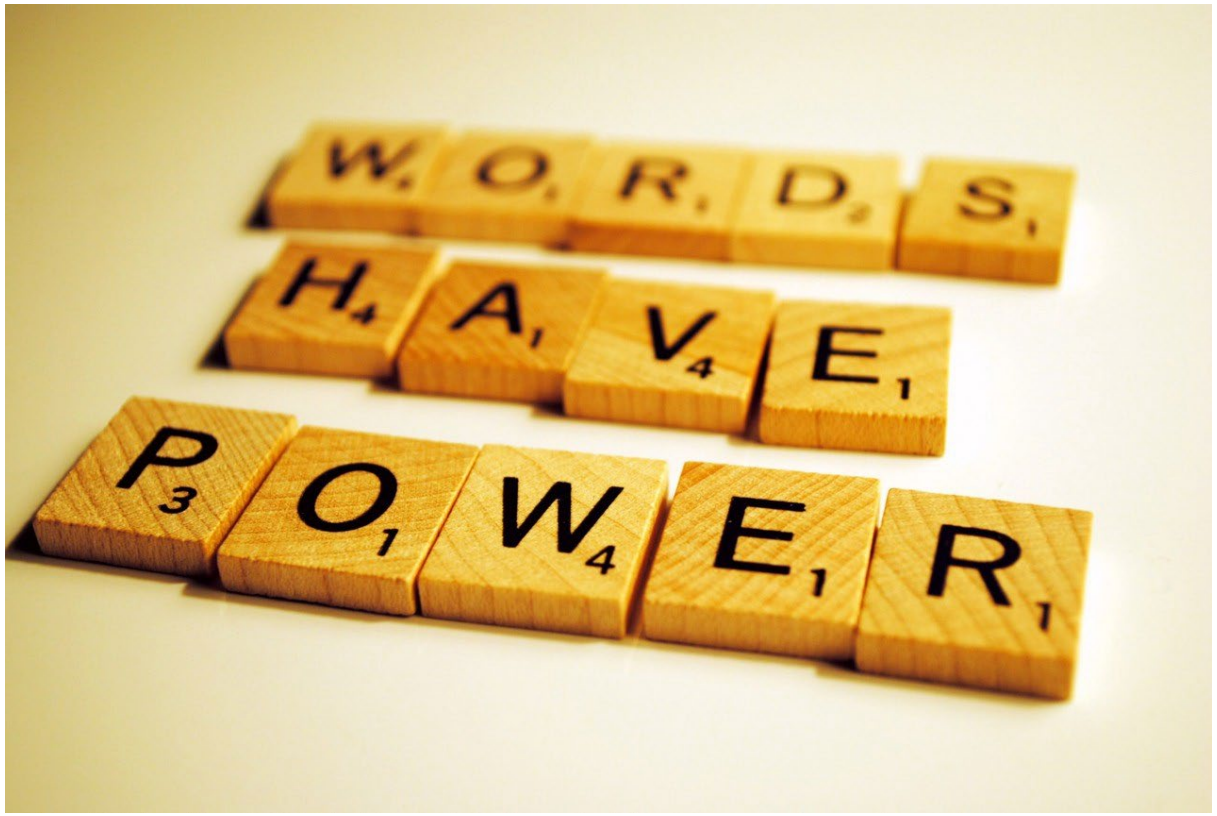



## Training Data Set




National Institutes of Health

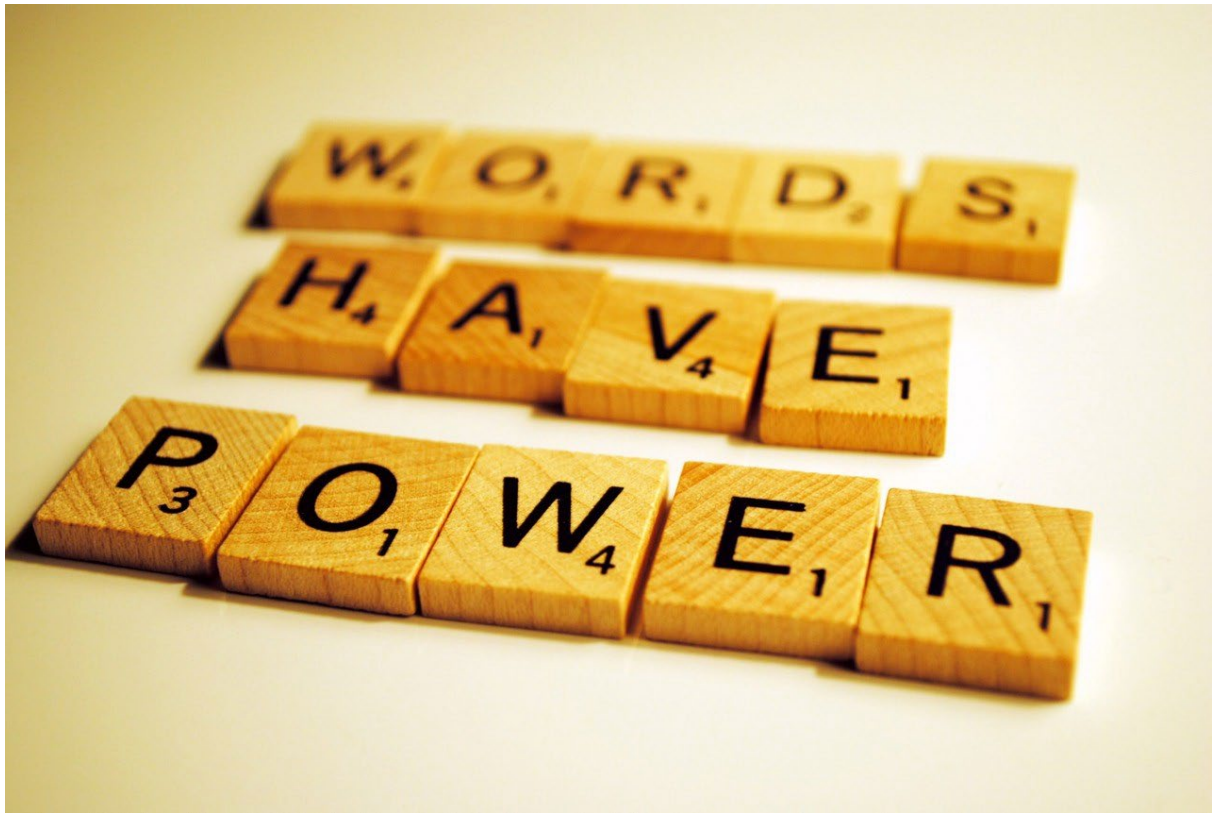
# “Bag of Words” Approach




- Text (i.e., grant specific aims text) is represented as a bag of its words
  - Grammar and word order disregarded
  - Multiplicity kept


- Commonly used to train classifiers when frequency of words is a desired feature
  1. Vocabulary of known words
  2. Measure of presence of known words

# “Bag of Words” Approach (Cont.)




- Documents are similar if content is similar
- Vectorization: process of converting text into numbers to apply mathematical principles to words
- Limitations:
  1) Vocabulary
  2) Sparsity
  3) Meaning

# “Bag of Words” Common Example

“It was the best of times”

“It was the worst of times”

“It was the age of wisdom”

“It was the age of foolishness”

Each sentence is a separate document in our corpus, much like each grant’s specific aims text is a document in our corpus (rehab portfolio)

First document:

“it” = 1

“was” = 1

“the” = 1

“best” = 1

“of” = 1

“times” = 1

“worst” = 0

“age” = 0

“wisdom” = 0

“foolishness” = 0

# “Bag of Words” Common Example (Cont.)

“It was the best of times” = [1,1,1,1,1,1,0,0,0,0]

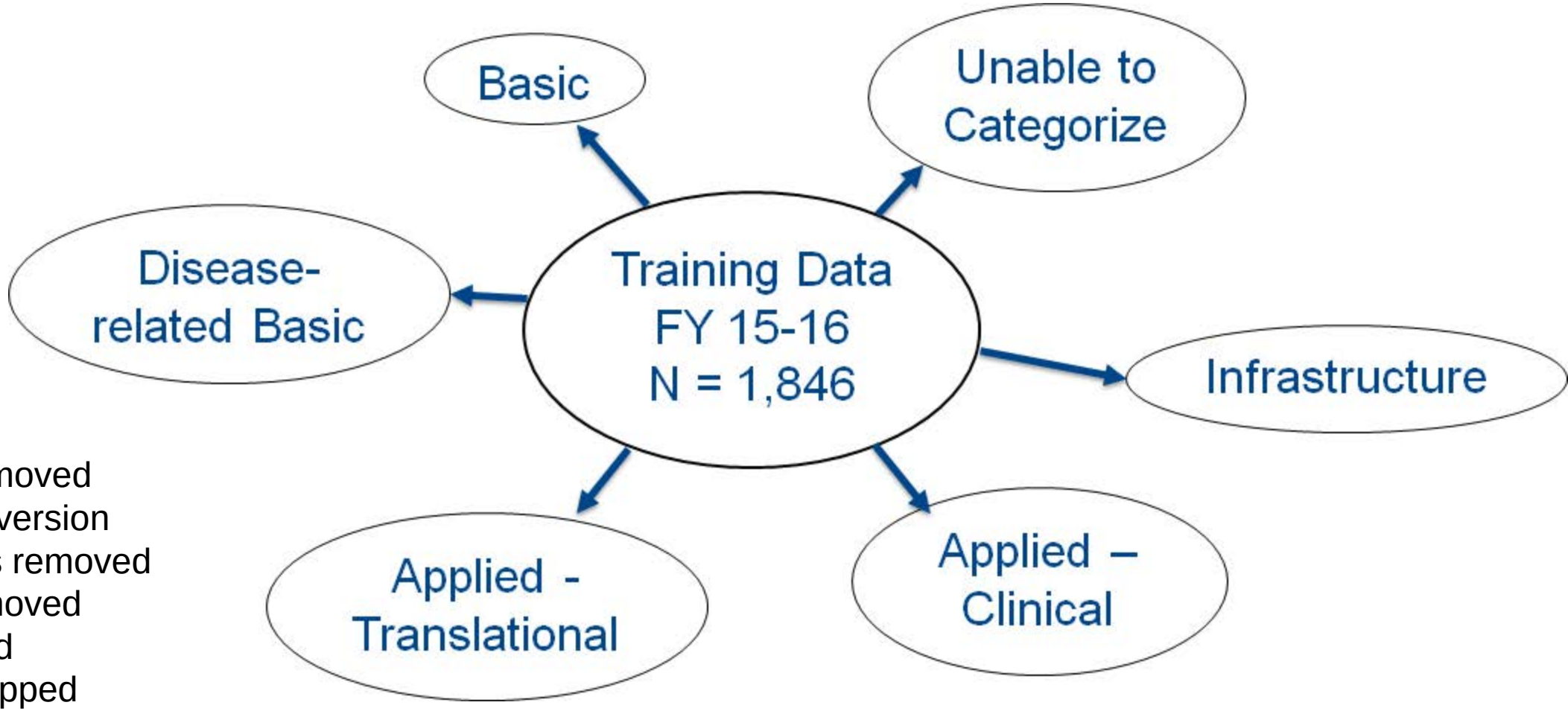
“It was the worst of times” = [1,1,1,0,1,1,1,0,0,0]

“It was the age of wisdom” = [1,1,1,0,1,0,0,1,1,0]

“It was the age of foolishness” = [1,1,1,0,1,0,0,1,0,1]

Each sentence is a separate document in our corpus, much like each grant’s specific aims text is a document in our corpus (rehab portfolio)

# Algorithm Development in R[1]




Each corpus:

- Punctuation removed
- Lowercase conversion
- Single numbers removed
- Stop words removed
- Words stemmed
- Whitespace stripped
- Matrix created that lists all occurrences of words by grant (document)

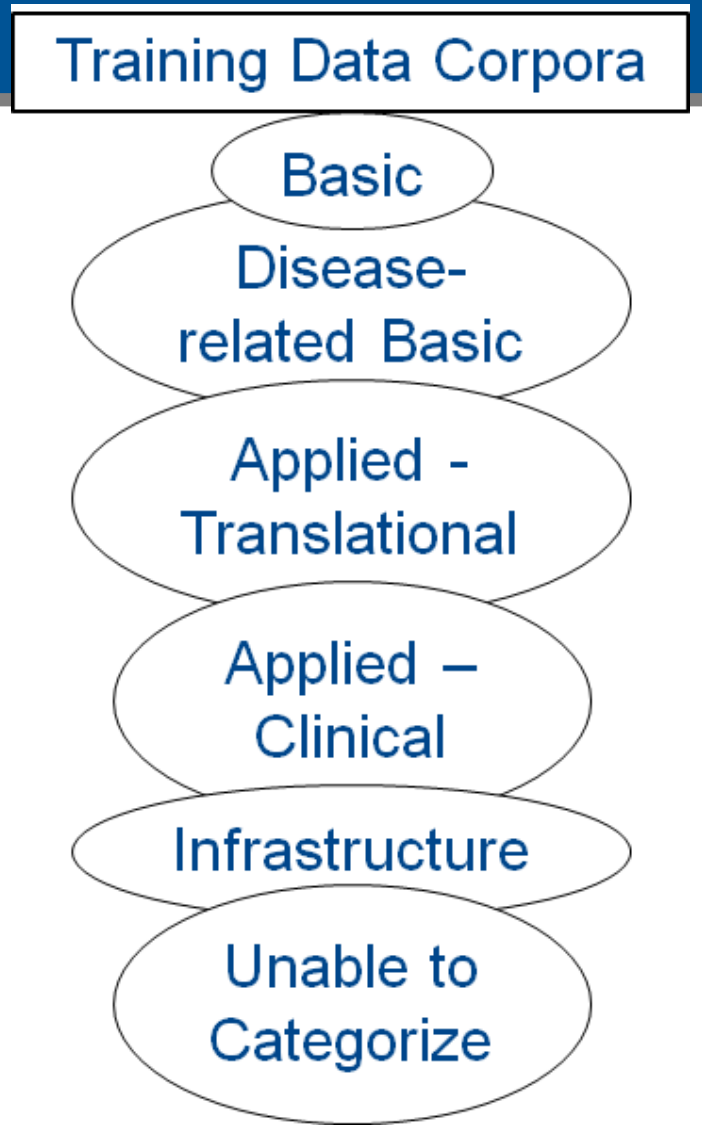
[1]R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

NIH National Institutes of Health

# Algorithm Development in R

Testing Data for FY 17 N = 444

- One grant compared to corpora at a time
- Each word compared to x number of most frequent corpora words, where x is optimized for highest agreement
  - Limited vocabulary controls sparsity (problem for words with small frequencies)
- Category determined based on number of unique words matched per category (i.e., not number of instances)



# Algorithm Development in R (cont.)

- Results were refined with automatic coding rules based on patterns in:
  - Funding mechanisms (e.g., K12s and T32s have specific codes)
  - Clinical trial and phase III clinical trial codes (e.g., CT code usually means applied clinical)
  - Animal and human codes (e.g., if neither code then not applied clinical)
  - Keywords in each grant's specific aims text (e.g., R01s with no animal or human codes with "mechanism" in specific aims text are basic grants)
- Same method was applied for primary and secondary coding (training corpora grouped by research plan categories)
- Secondary codes change if primary code matches
- Tertiary codes are automatic, based on funding mechanism or NIA RFA

# Algorithm Validation Results – Tier II Category

| Prediction \ Actual | AC | AT | BA | DR | IN | NC |
|---|---|---|---|---|---|---|
| AC | 175 | 3 | 3 | 20 | 1 | 0 |
| AT | 5 | 39 | 4 | 6 | 0 | 0 |
| BA | 3 | 2 | 32 | 10 | 2 | 0 |
| DR | 12 | 1 | 6 | 94 | 0 | 0 |
| IN | 1 | 0 | 0 | 1 | 10 | 0 |
| NC | 1 | 0 | 0 | 1 | 0 | 12 |

Agreement = 362 grants

| Prediction \ Actual | AC | AT | BA | DR | IN | NC |
|---|---|---|---|---|---|---|
| AC | 88.8% | 6.7% | 6.7% | 15.2% | 7.7% | 0% |
| AT | 2.5% | 86.7% | 8.9% | 4.6% | 0% | 0% |
| BA | 1.5% | 4.4% | 71.1% | 7.6% | 15.4% | 0% |
| DR | 6.1% | 2.2% | 13.3% | 71.2% | 0% | 0% |
| IN | 0.5% | 0% | 0% | 0.8% | 76.9% | 0% |
| NC | 0.5% | 0% | 0% | 0.8% | 0% | 100% |

Overall agreement = 81.5%

Tier II Categories

- AC: Applied – Clinical
- AT: Applied – Translational
- BA: Basic
- DR: Disease-related Basic
- IN: Infrastructure
- NC: Unable to Categorize

Diagonals are correct; other values are mismatches

Cohen's kappa = 0.74 (satisfactory agreement)

95% Confidence Interval = [0.1486,0.2208]

Starting agreement: 67.8%

# Algorithm Validation Results – Primary Category

| Prediction \ Actual | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| A | 48 | 0 | 6 | 14 | 9 | 1 |
| B | 2 | 5 | 0 | 2 | 0 | 0 |
| C | 2 | 0 | 73 | 5 | 2 | 0 |
| D | 9 | 0 | 9 | 154 | 5 | 1 |
| E | 0 | 0 | 4 | 1 | 69 | 0 |
| F | 0 | 1 | 0 | 1 | 1 | 20 |

Agreement = 369 grants

| Prediction \ Actual | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| A | 78.7% | 0% | 6.5% | 7.9% | 10.5% | 4.6% |
| B | 3.3% | 83.3% | 0.0% | 1.1% | 0% | 0% |
| C | 3.3% | 0% | 79.4% | 2.8% | 2.3% | 0% |
| D | 14.8% | 0% | 9.8% | 87.0% | 5.8% | 4.6% |
| E | 0.0% | 0% | 4.4% | 0.6% | 80.2% | 0% |
| F | 0.0% | 16.7% | 0% | 0.6% | 1.2% | 90.9% |

Overall agreement = 83.1%

Primary Categories

- A: Rehabilitation across the lifespan
- B: Community and family
- C: Technology use and development
- D: Research design and methodology
- E: Translational science
- F: Building research capacity and infrastructure

Diagonals are correct; other values are mismatches

Cohen's kappa = 0.77 (satisfactory agreement)

95% Confidence Interval = [0.1341,0.2038]

Starting agreement: 70.7%

# Algorithm Validation Results – Secondary Category

| Prediction \ Actual | Blank | A | B | C | D | E |
|---|---|---|---|---|---|---|
| Blank | 30 | 0 | 0 | 0 | 0 | 0 |
| A | 6 | 113 | 10 | 10 | 9 | 16 |
| B | 0 | 10 | 23 | 2 | 0 | 2 |
| C | 8 | 22 | 2 | 37 | 7 | 0 |
| D | 10 | 20 | 1 | 9 | 37 | 5 |
| E | 10 | 10 | 0 | 6 | 2 | 27 |

Agreement = 267 grants

| Prediction \ Actual | Blank | A | B | C | D | E |
|---|---|---|---|---|---|---|
| Blank | 46.9% | 0% | 0.0% | 0.0% | 0.0% | 0.0% |
| A | 9.4% | 64.6% | 27.8% | 15.6% | 16% | 32% |
| B | 0.0% | 6% | 63.9% | 3.1% | 0.0% | 4% |
| C | 12.5% | 13% | 5.6% | 57.8% | 12.7% | 0.0% |
| D | 15.6% | 11% | 2.8% | 14.1% | 67.3% | 10% |
| E | 15.6% | 5.7% | 0% | 9.4% | 3.6% | 54.0% |

Overall agreement = 60.1%

Secondary Categories

- Blank: no secondary category
- A: Rehabilitation across the lifespan
- B: Community and family
- C: Technology use and development
- D: Research design and methodology
- E: Translational science

Diagonals are correct; other values are mismatches

Cohen's kappa = 0.49 (moderate agreement)

95% Confidence Interval = [0.3531,0.4442]

Starting agreement: 34.7%

# Automatic Coding Process

- Algorithm uses 2015-17 data for training and 2018 data for testing
  - Optimized values remain the same
- 20% random sample chosen from each IC's new 2018 data (N = 115)
- NCMRR checked coding for the sample using the coding rules
  - Each project checked separately by two coders and the codes reconciled for agreement
  - Statistics computed for level of agreement for primary, secondary, and tier II
- Each IC received the reconciled portfolio and confirmed or revised the algorithm+NCMRR proposed coding, same as before

# Conclusions/Future Directions

- The algorithm was able to automatically code grants with better agreement compared to individual coders.
  - Might look into context and meaning to further improve agreement, like n-gram models
- NCMRR can use this classification algorithm to automatically code NIH awarded rehabilitation grants.
- This approach will be updated each year to include more data in the training corpora.
- Using prior manually coded and verified data, this algorithm/process can be modified to automatically code the type of science, as well as additional criteria, for other portfolios across the NIH.

# Algorithm Versatility Example: NINDS

- Full Dataset
  - 20,158 grants (2002-2019) coded for type of science
  - Quartile percentage coding (adding up to 100%) of four categories (basic (BA), disease-related basic (DR), and applied (translational (AT) or clinical (AC))
- Usable Dataset
  - 15,737 grants (2002-2019)
  - Only grants coded 100% of one category
  - Testing set: 11,802 grants (75%)
  - Training set: 3,934 grants (25%)
  - 72.0% agreement (bag of words only; 2,835 grants accurately coded)

| | Actual | | | |
|---|---|---|---|---|
| Predictions | AC | AT | BA | DR |
| AC | 299 | 84 | 34 | 169 |
| AT | 17 | 487 | 48 | 160 |
| BA | 8 | 32 | 885 | 305 |
| DR | 39 | 99 | 105 | 1164 |

| | Actual | | | |
|---|---|---|---|---|
| Predictions | AC | AT | BA | DR |
| AC | "82.37" | "11.97" | " 3.17" | " 9.40" |
| AT | " 4.68" | "69.37" | " 4.48" | " 8.90" |
| BA | " 2.20" | " 4.56" | "82.56" | "16.96" |
| DR | "10.74" | "14.10" | " 9.79" | "64.74" |

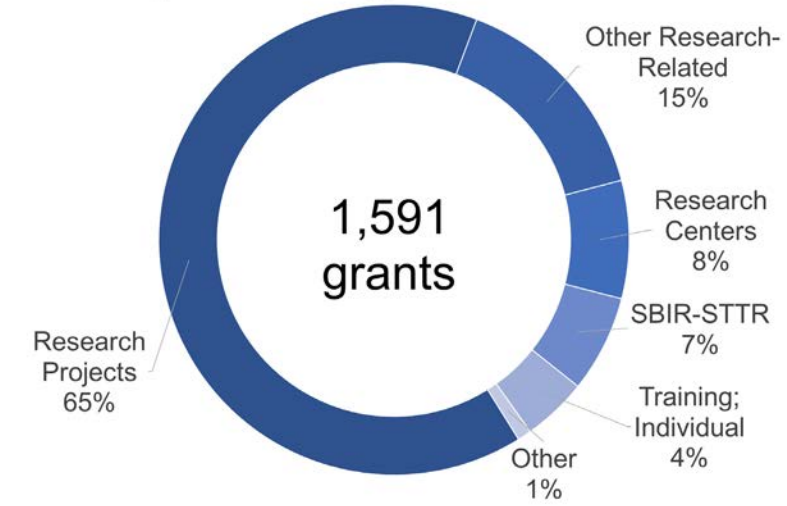
# The Portfolio: 2018 (excluding intramural projects)

| Funding Institute/Center | Total Projects per Admin IC | New Projects per Admin IC | Total Funding per Funding IC |
|---|---|---|---|
| NINDS | 288 | 97 | $123,884,006 |
| NIDCD | 285 | 75 | $94,120,088 |
| NICHD | 240 | 74 | $74,688,089 |
| NIA | 203 | 71 | $92,260,394 |
| NCI | 168 | 55 | $59,271,914 |
| NIMH | 74 | 19 | $29,727,346 |
| NIAMS | 54 | 17 | $17,756,096 |
| NHLBI | 41 | 18 | $22,356,389 |
| NIGMS | 38 | 2 | $9,370,764 |
| NEI | 34 | 12 | $12,668,736 |
| NINR | 30 | 12 | $10,620,734 |
| NCCIH | 30 | 12 | $8,681,706 |
| NIDA | 28 | 13 | $11,672,968 |
| NIBIB | 24 | 8 | $8,866,205 |
| NIAAA | 15 | 8 | $6,749,791 |
| NIMHD | 13 | 5 | $4,976,574 |
| NIDCR | 11 | 6 | $6,948,867 |
| NIDDK | 9 | 8 | $4,006,178 |
| OD | 2 | 1 | $6,175,051 |
| NCATS | 1 | 1 | $747,350 |
| NIEHS | 1 | 1 | $485,439 |
| NLM | 1 | 1 | $229,914 |
| FIC | 1 | 1 | $214,181 |
| RMAP | 0 | 0 | $2,641,954 |
| **2018 Total** | **1591** | **517** | **$609,120,734** |
| **2017 Total** | **1447** | **446** | **$541,712,921** |
| **2016 Total** | **1396** | **491** | **$517,579,011** |
| **2015 Total** | **1360** | **1360** | **$499,126,720** |

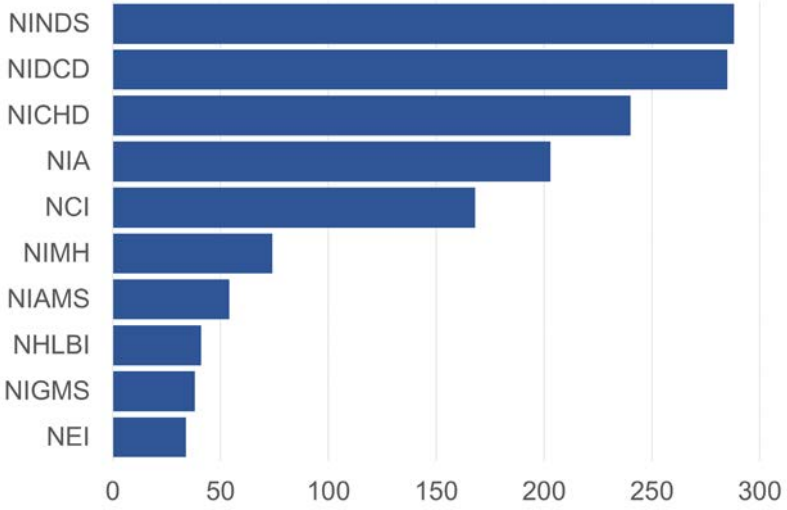
National Institutes of Health

# 2018 Number of Projects

## Grants per funding mechanism




## Grants per IC (top 10 ICs)

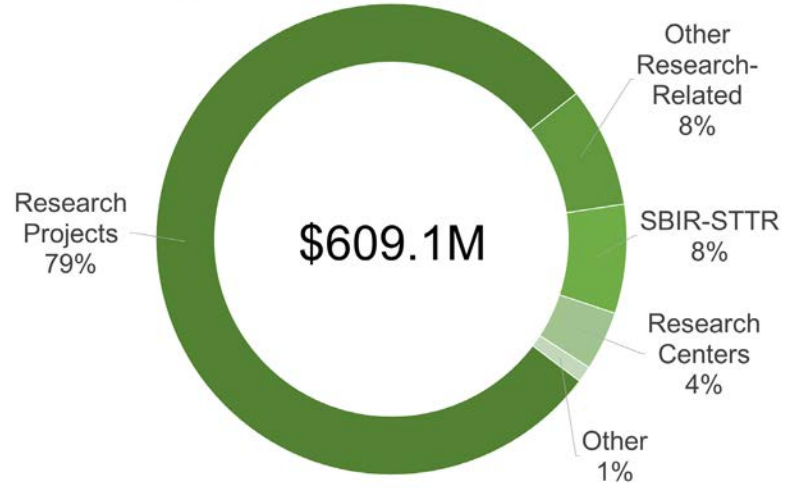



## Grants per RCDC term (top 30 terms)

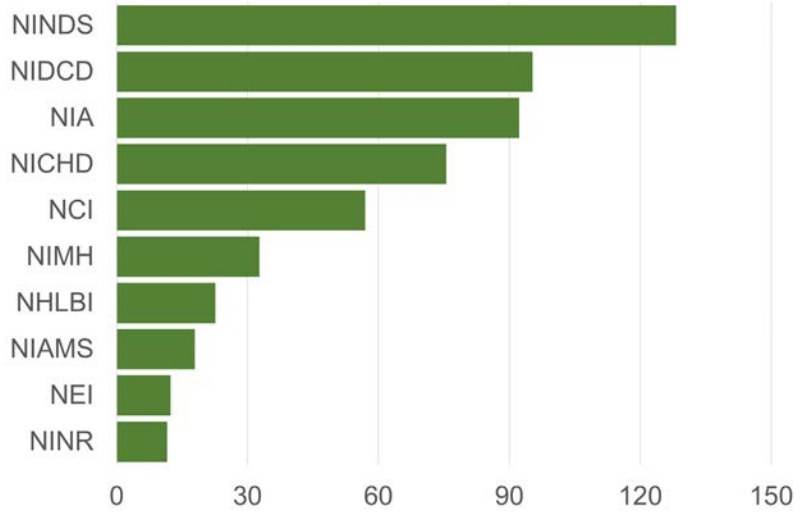

NIH National Institutes of Health

# 2018 Total Funding ($M)

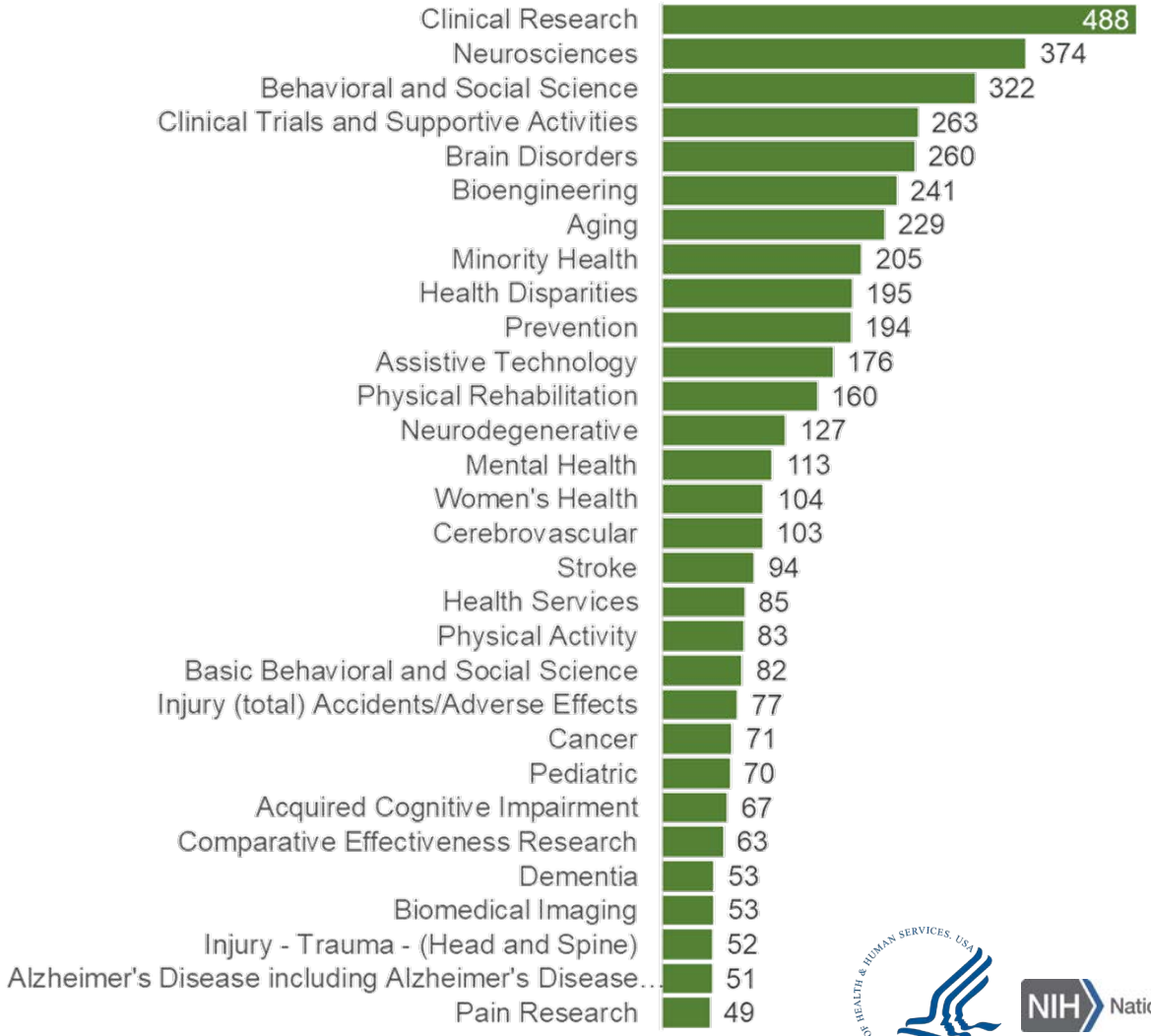
## Funding per funding mechanism




## Grants per IC (top 10 ICs)

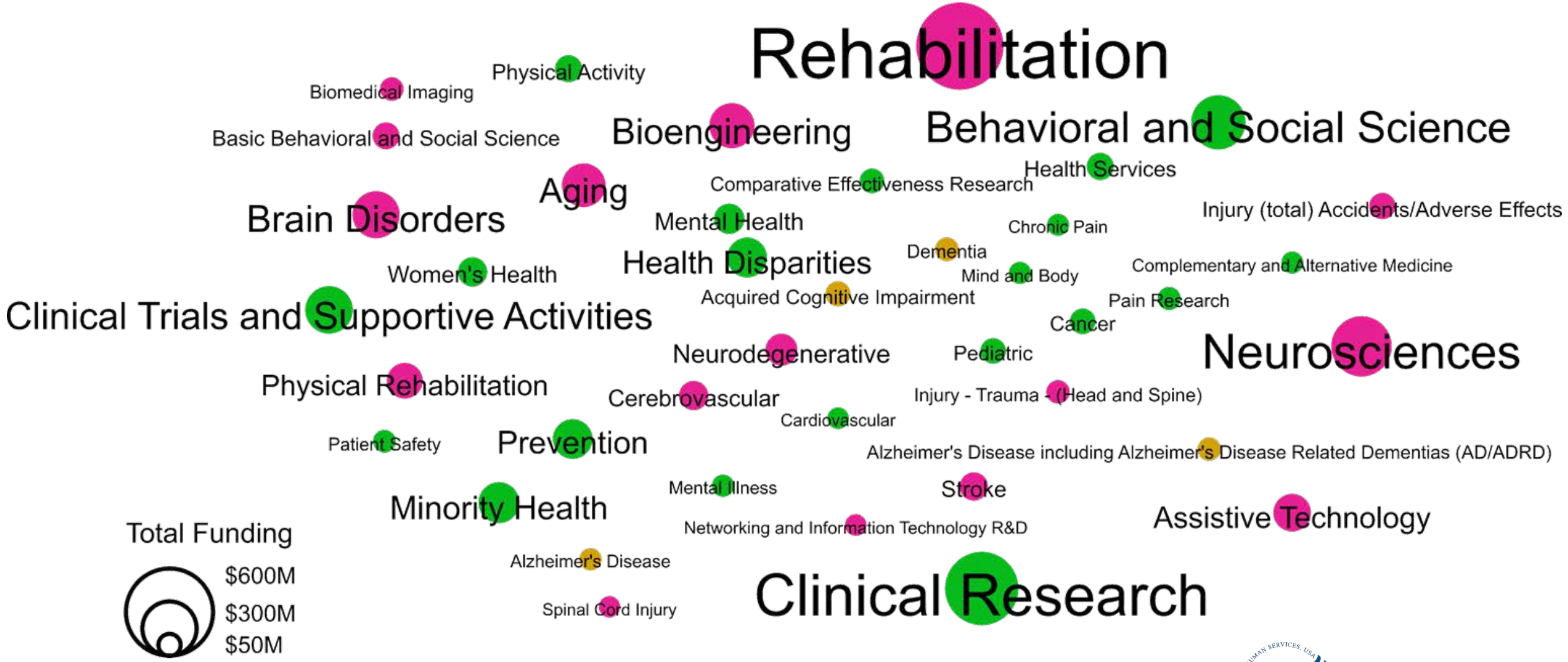



## Funding per RCDC term (top 30 terms)



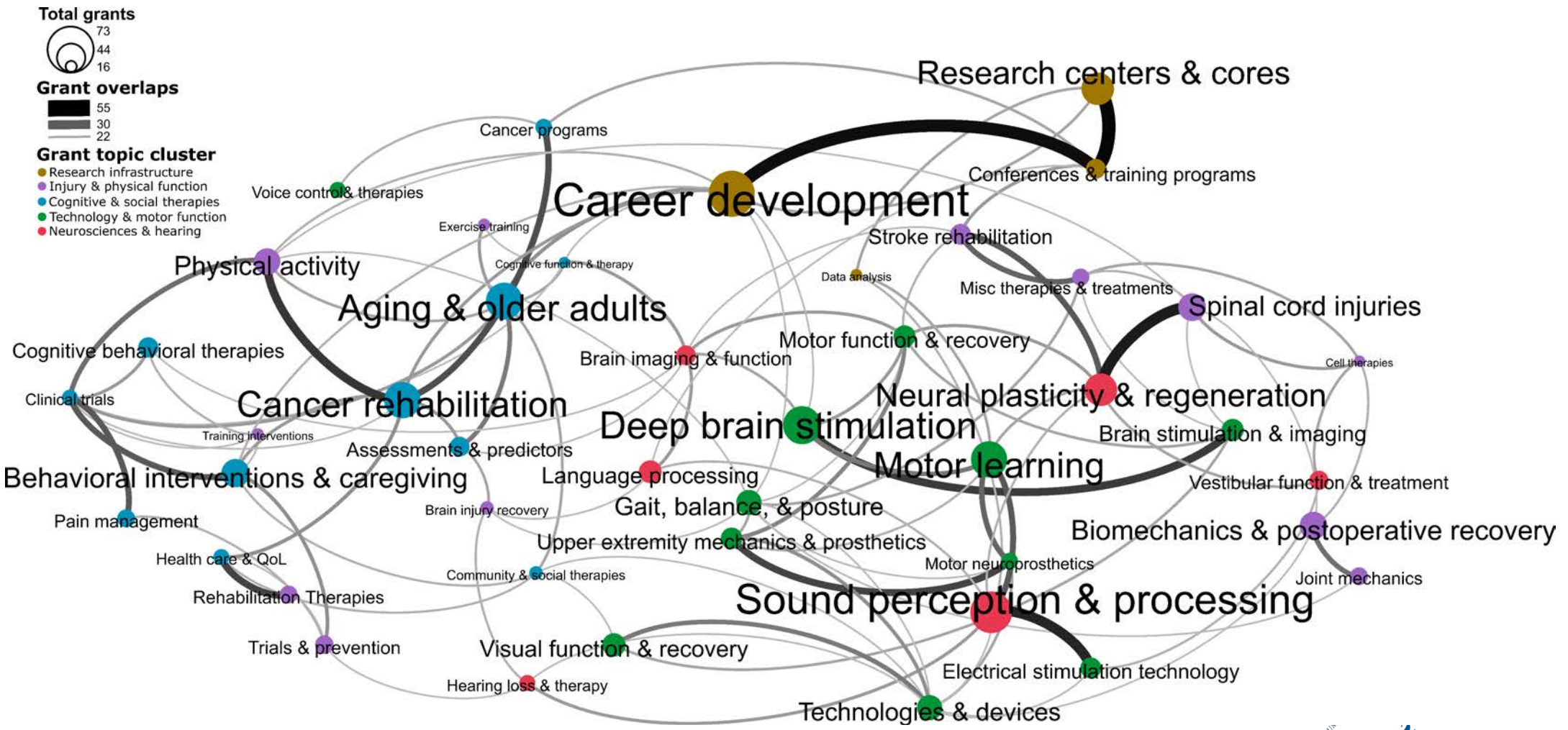

National Institutes of Health

# 2018 RCDC Term Co-occurrence Networks

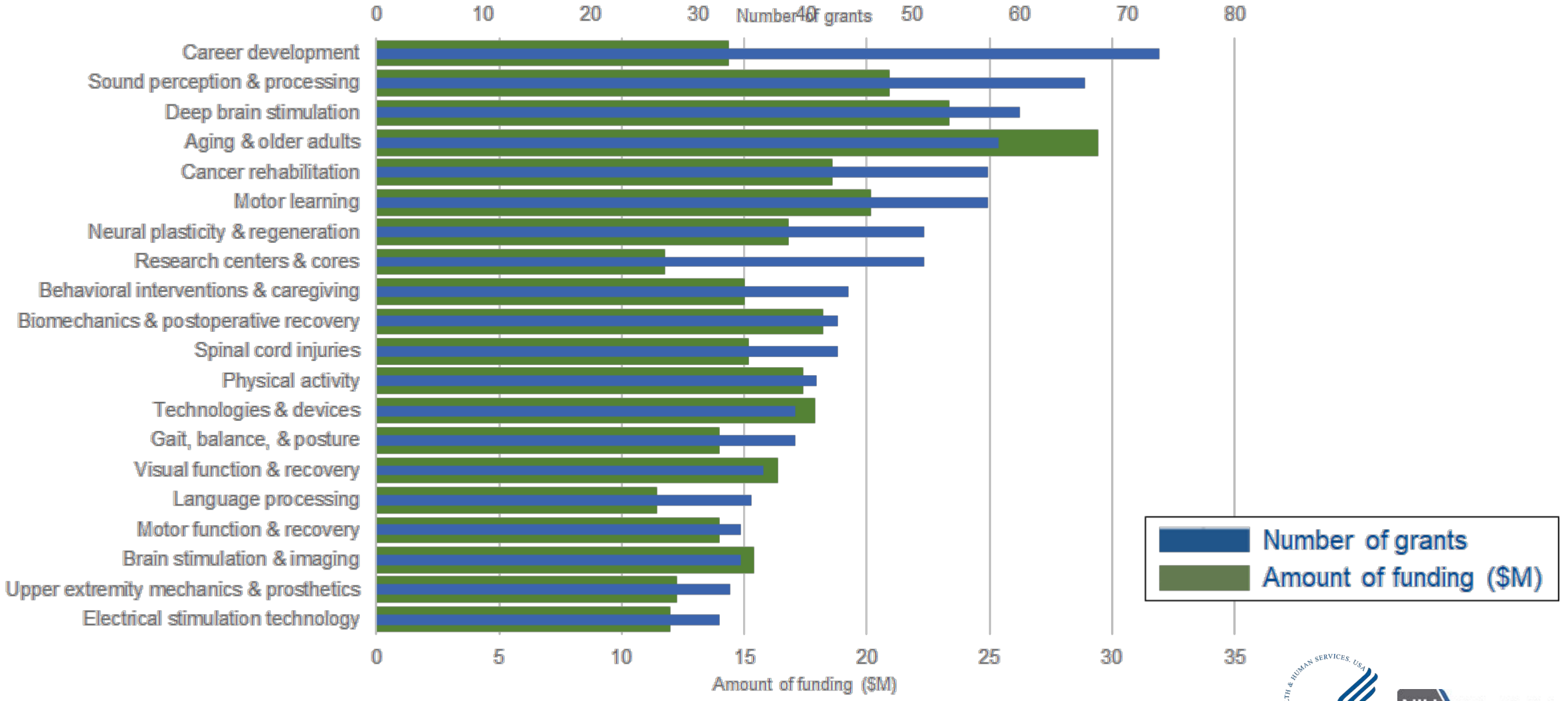



National Institutes of Health

# 2018 LDA (Latent Dirichlet Allocation) Topic Map

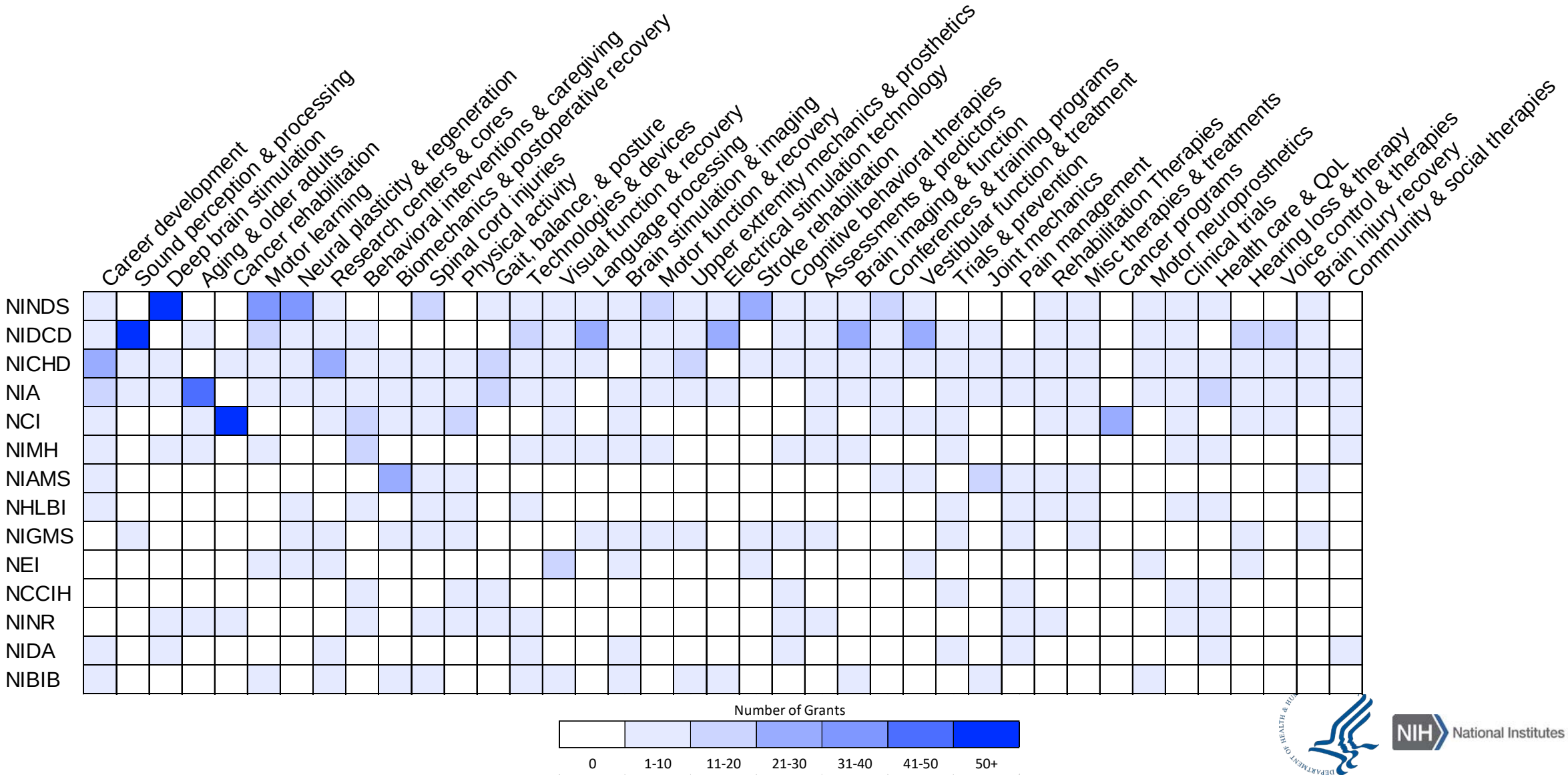


# 2018 Grants and Funding Per Topic (Top 20 by # of Grants)



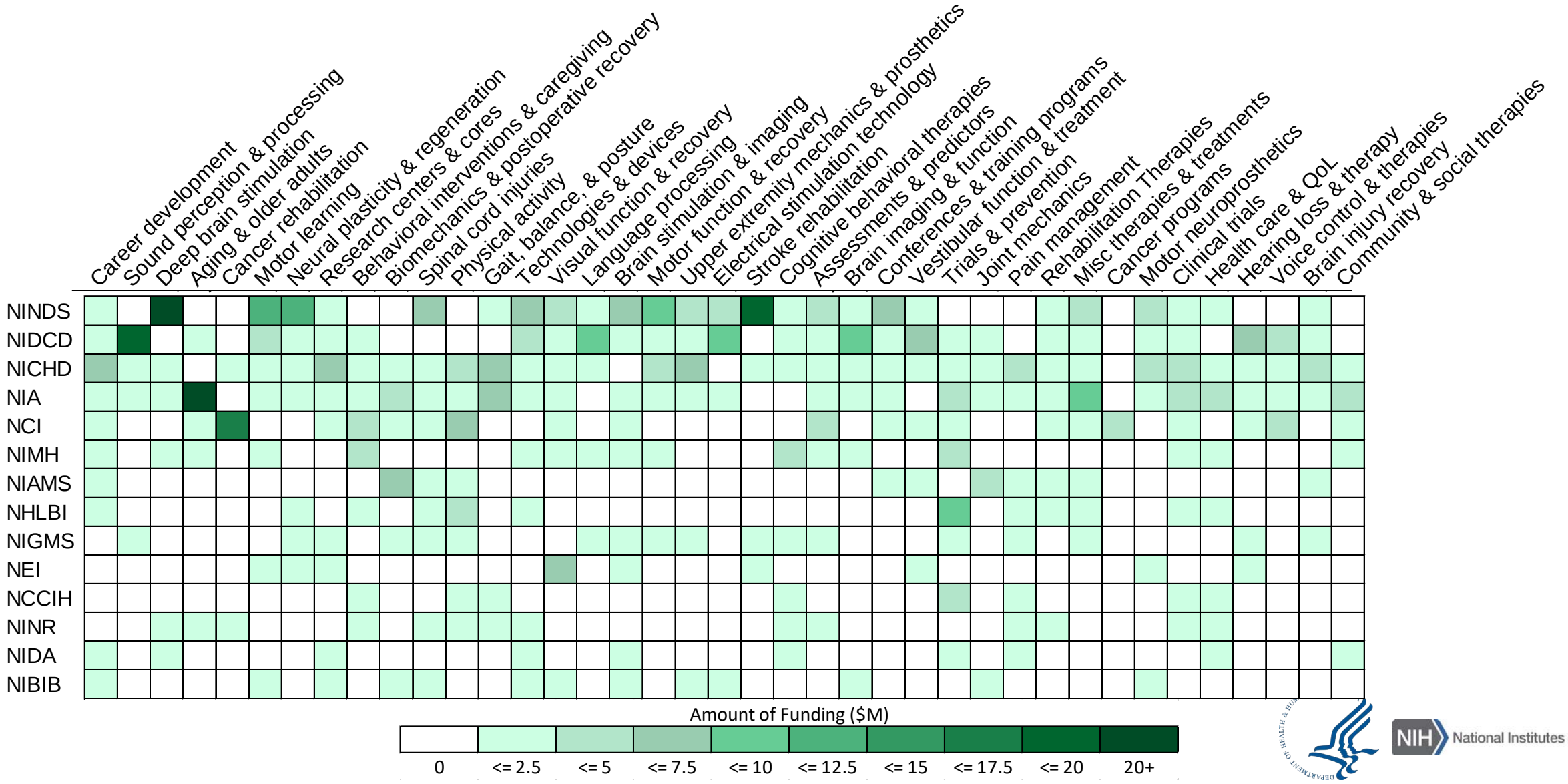

National Institutes of Health

# 2018 Number of Grants per IC per Topic



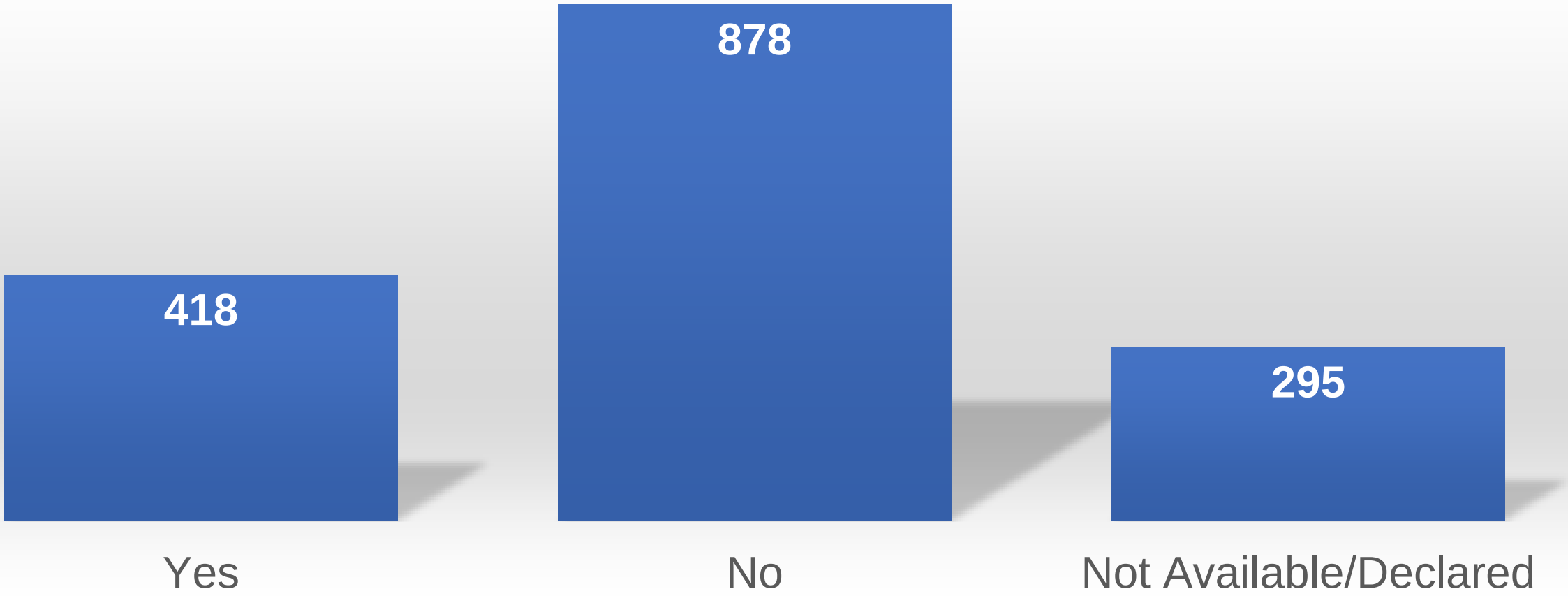

National Institutes of Health

# 2018 Funding per IC per Topic



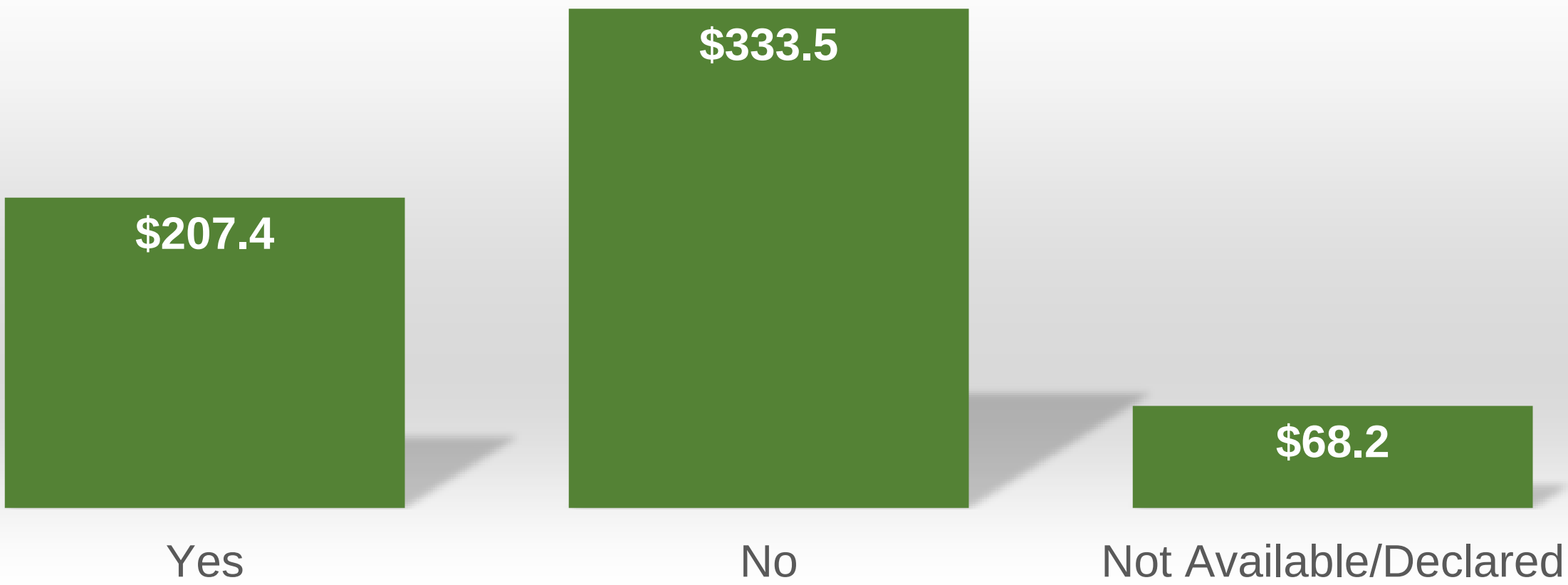

National Institutes of Health

# 2018 Clinical Trials (Numbers)

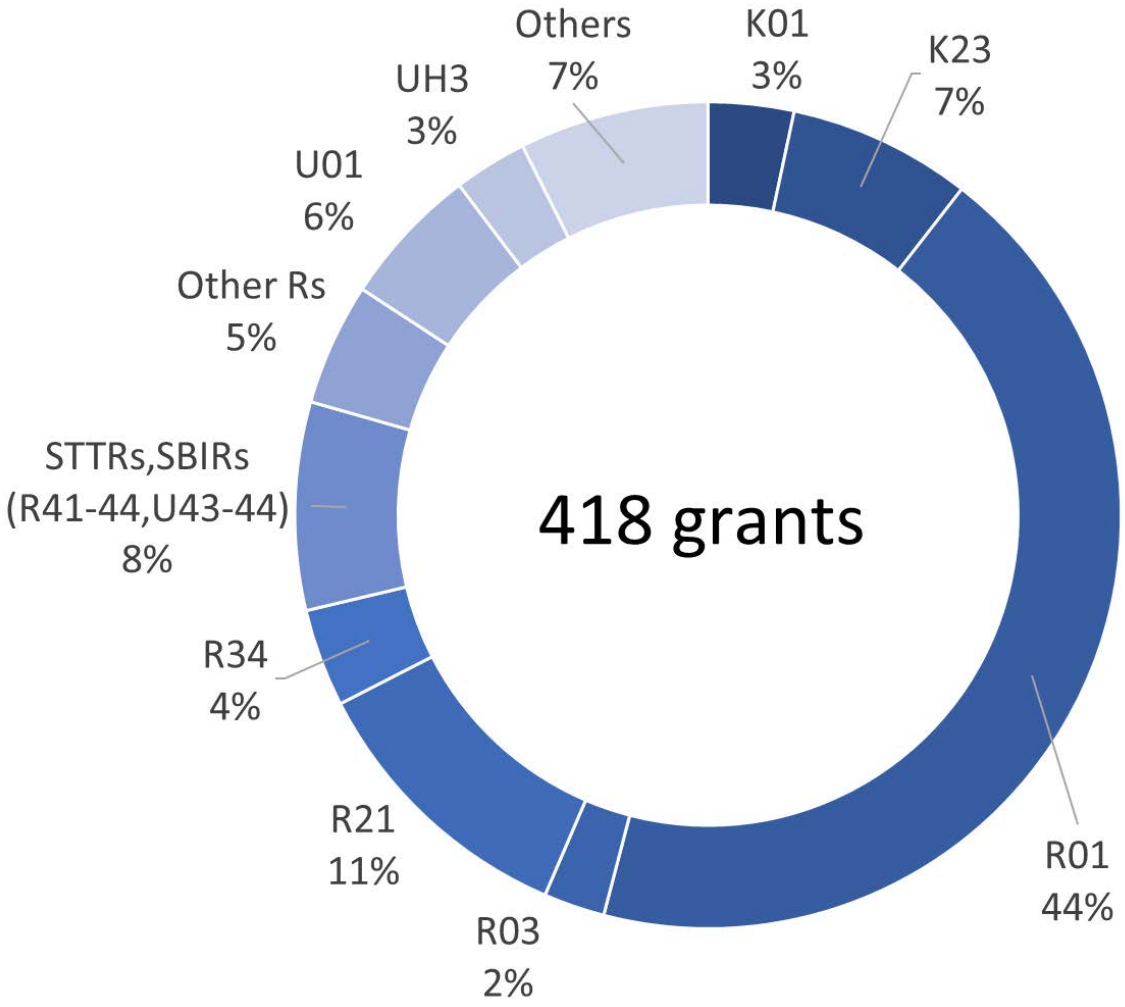


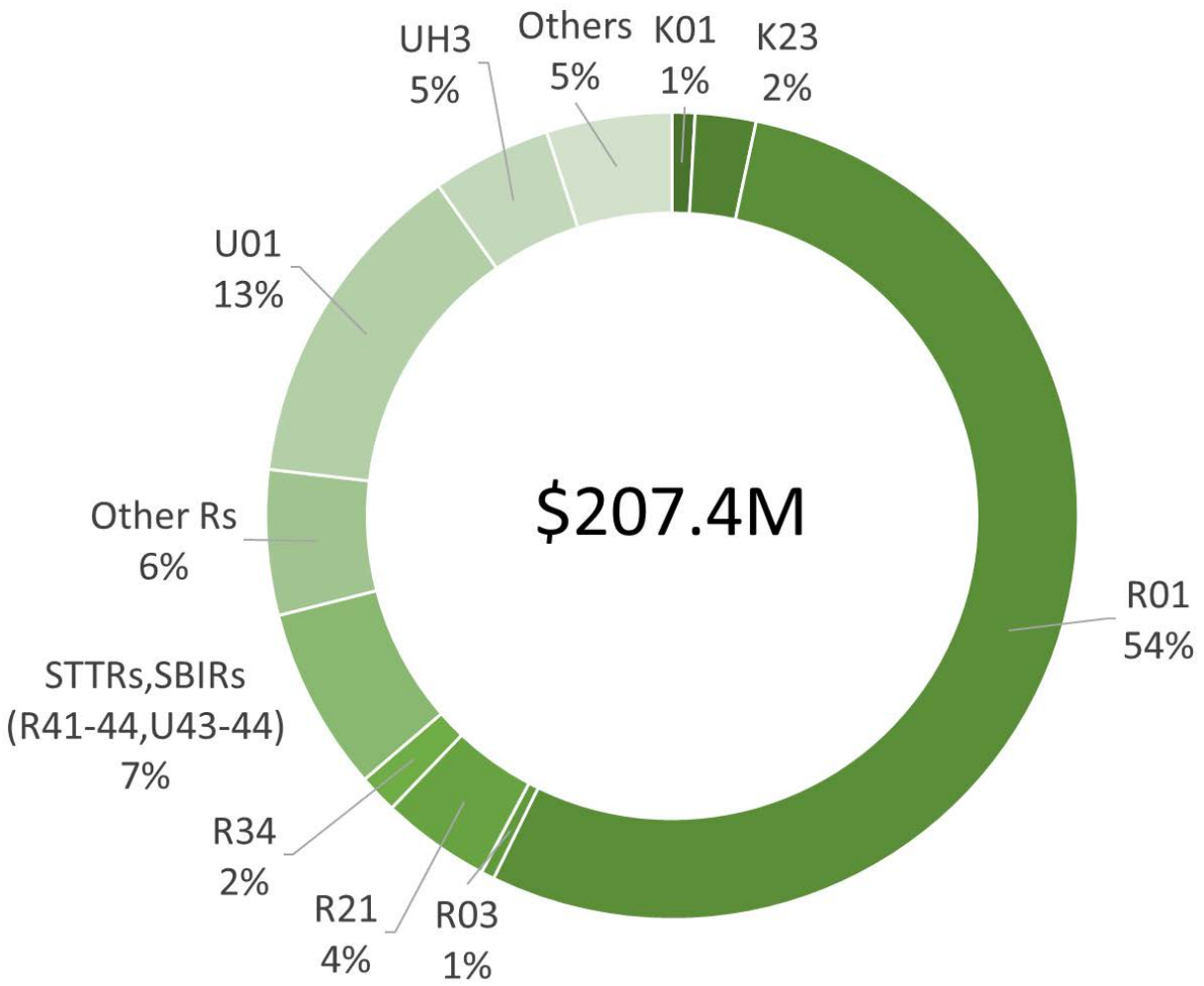

National Institutes of Health

# 2018 Clinical Trials ($M)



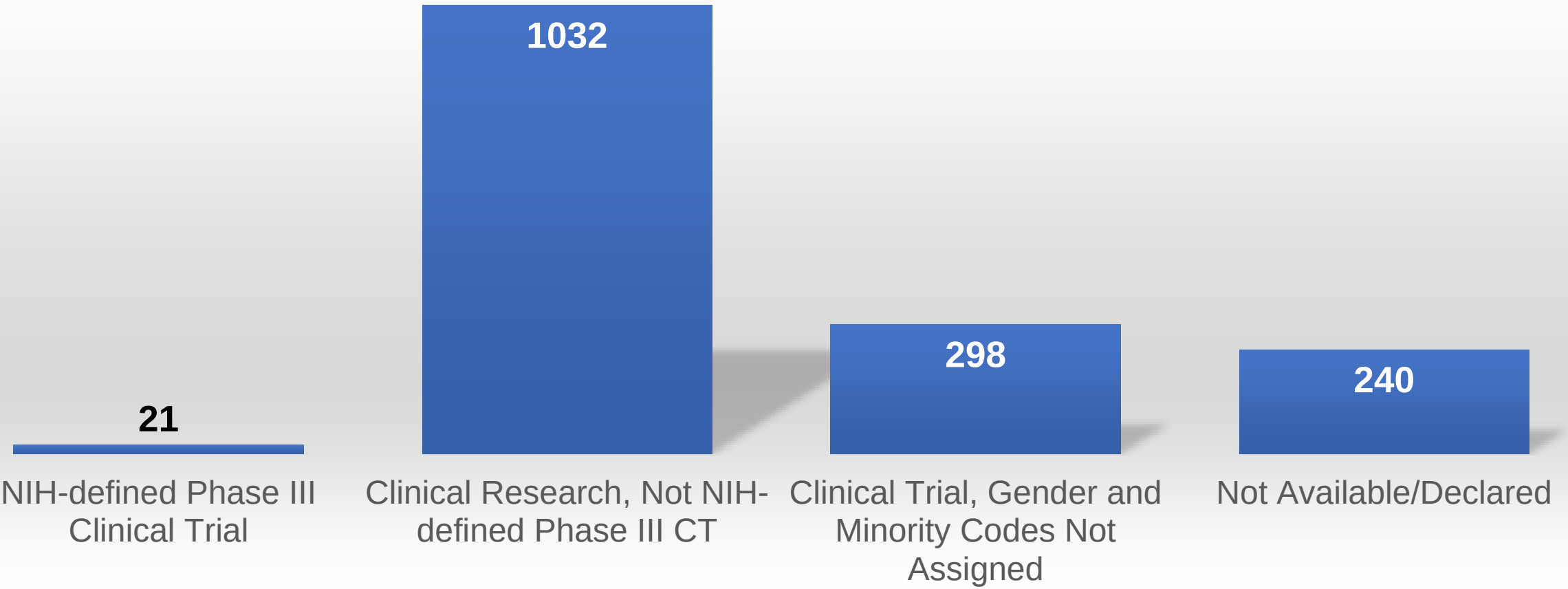
# 2018 Clinical Trial Funding Mechanisms



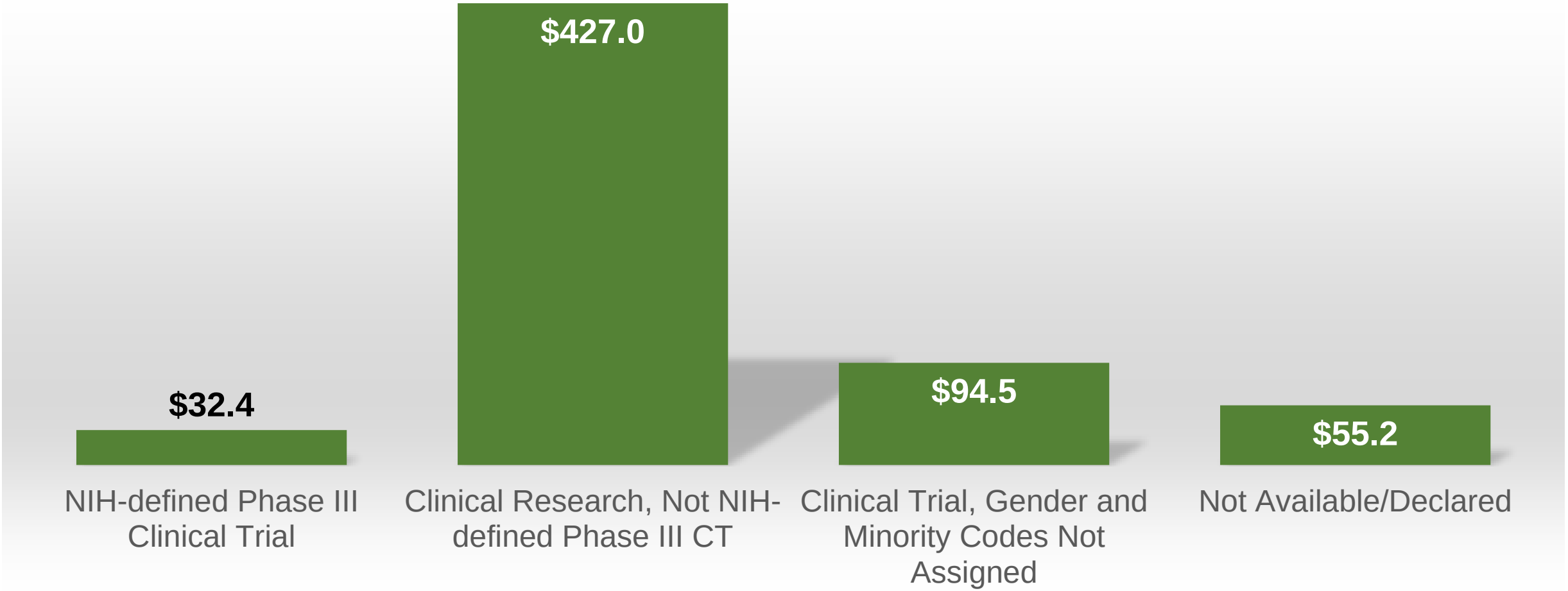




National Institutes of Health

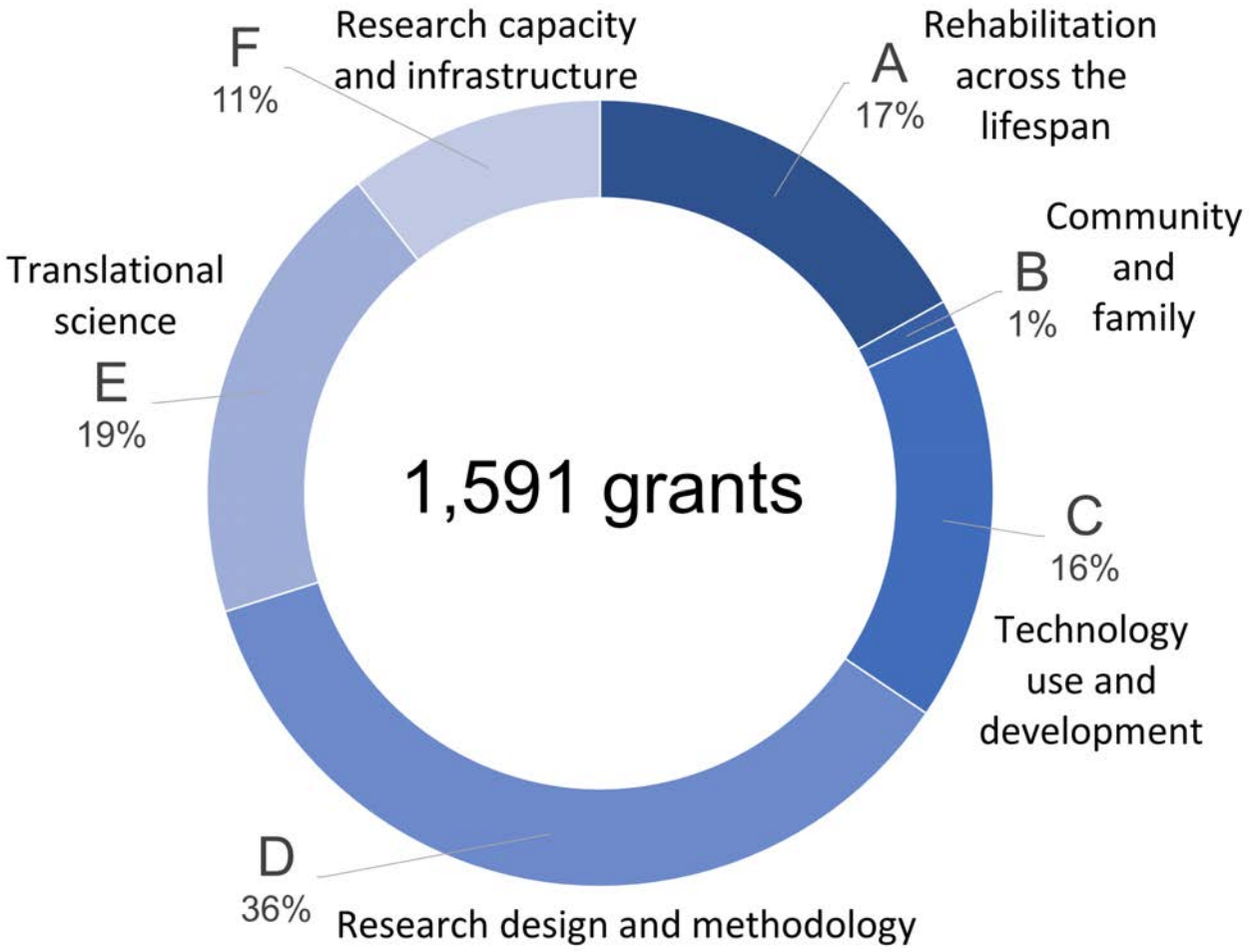
# 2018 Phase III Clinical Trials (Numbers)

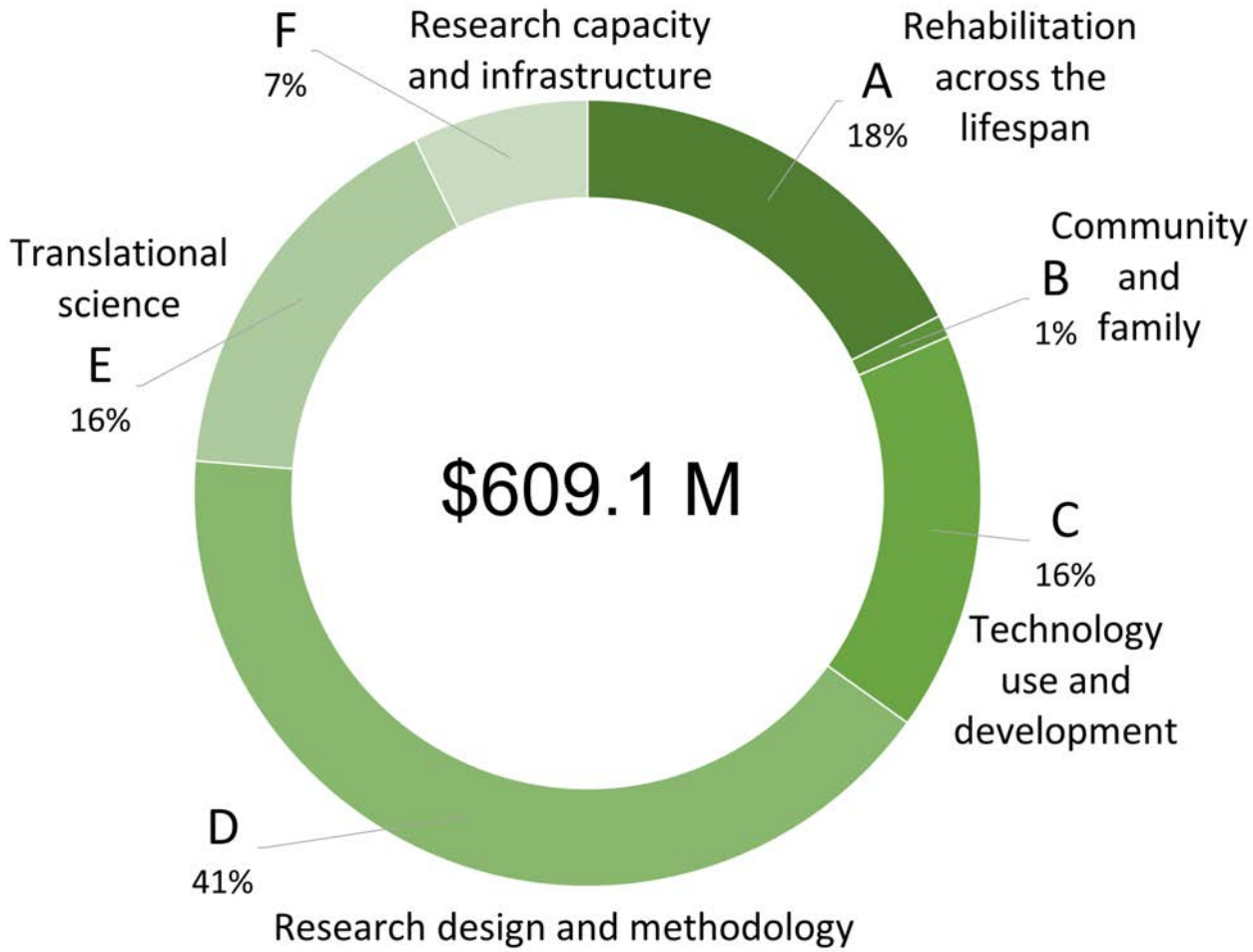


# 2018 Phase III Clinical Trials ($M)



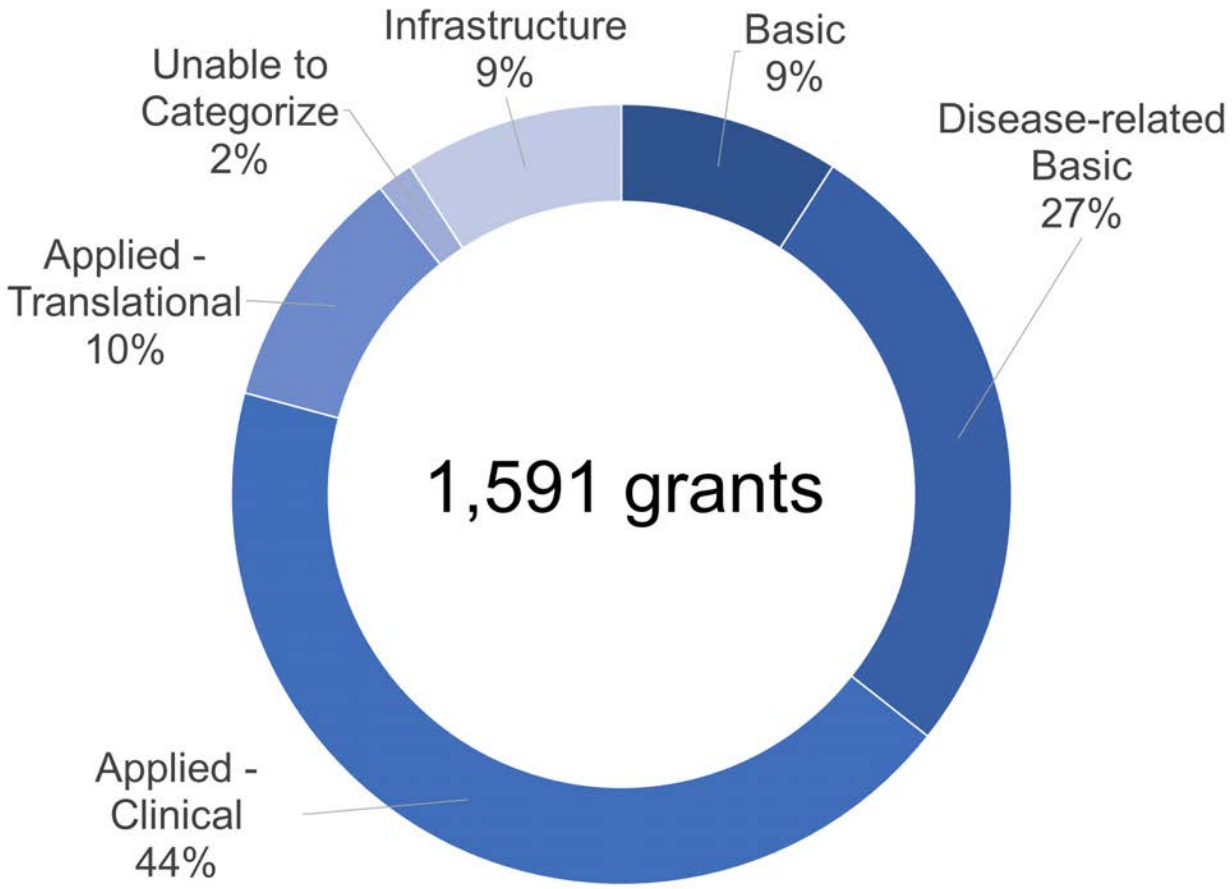
# 2018 Research Category Analysis (Primary Categories)

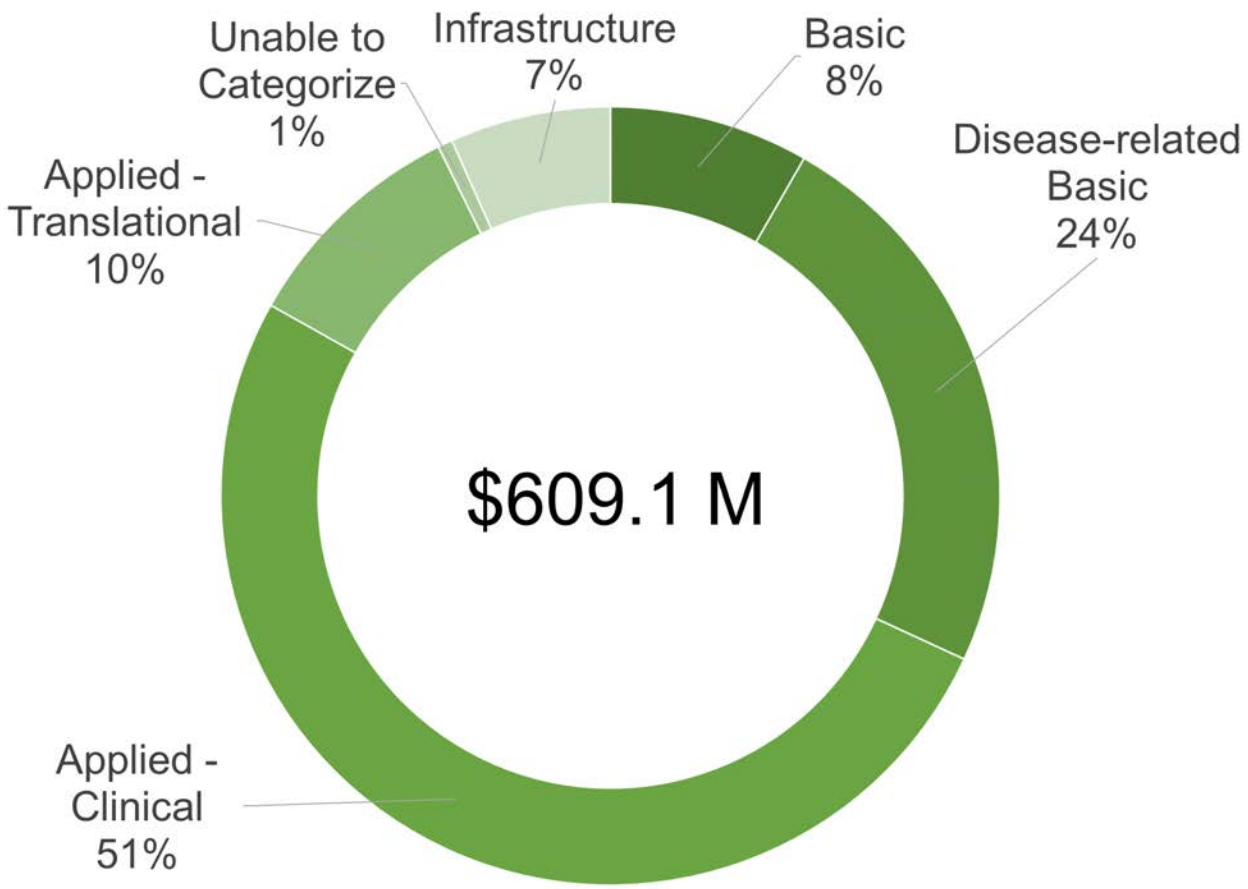



National Institutes of Health

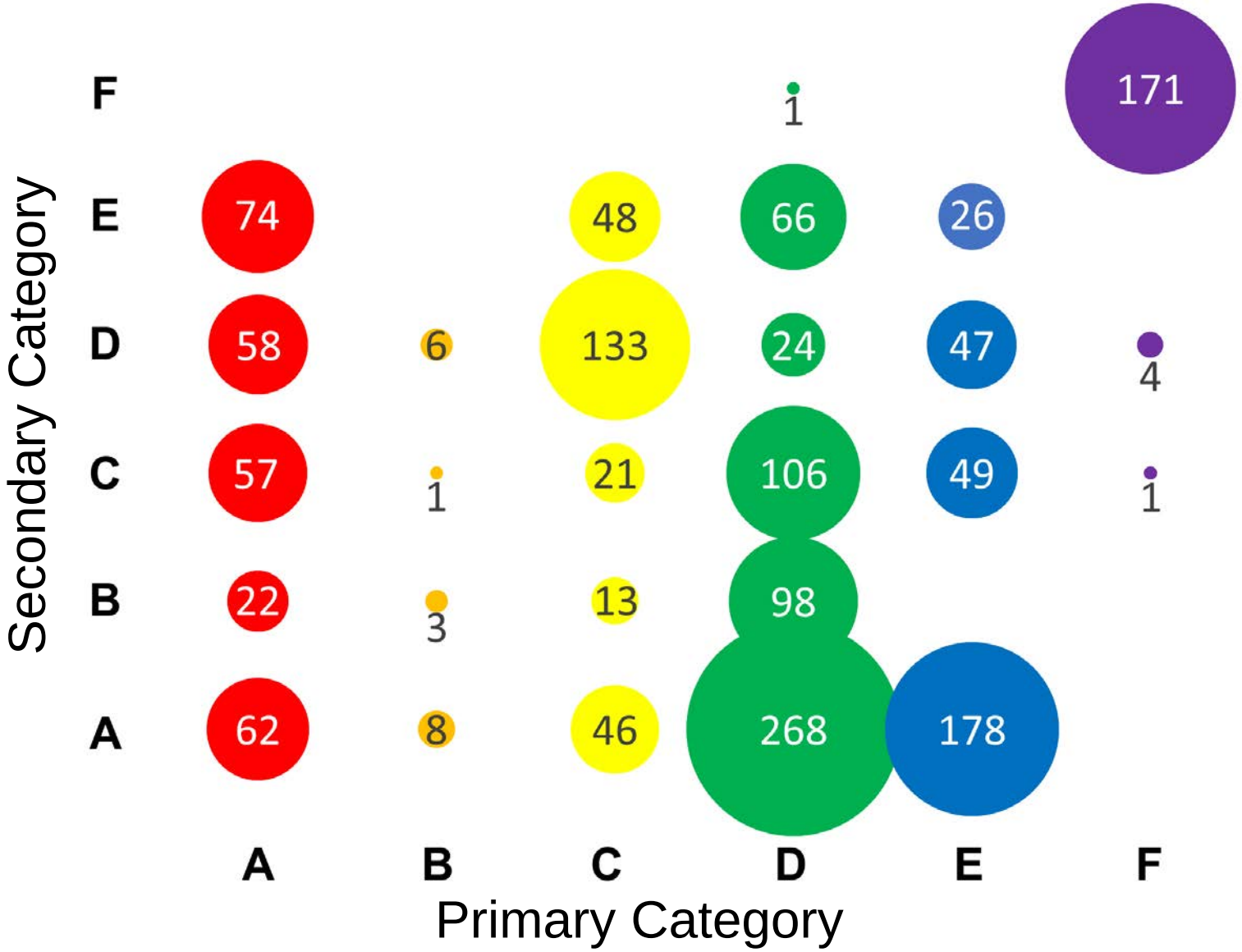
# 2018 Research Type Analysis (Tier II Categories)




National Institutes of Health

# 2018 Primary and Secondary Categories




A – Rehabilitation Across the Lifespan

B – Community and Family

C – Technology Use and Development

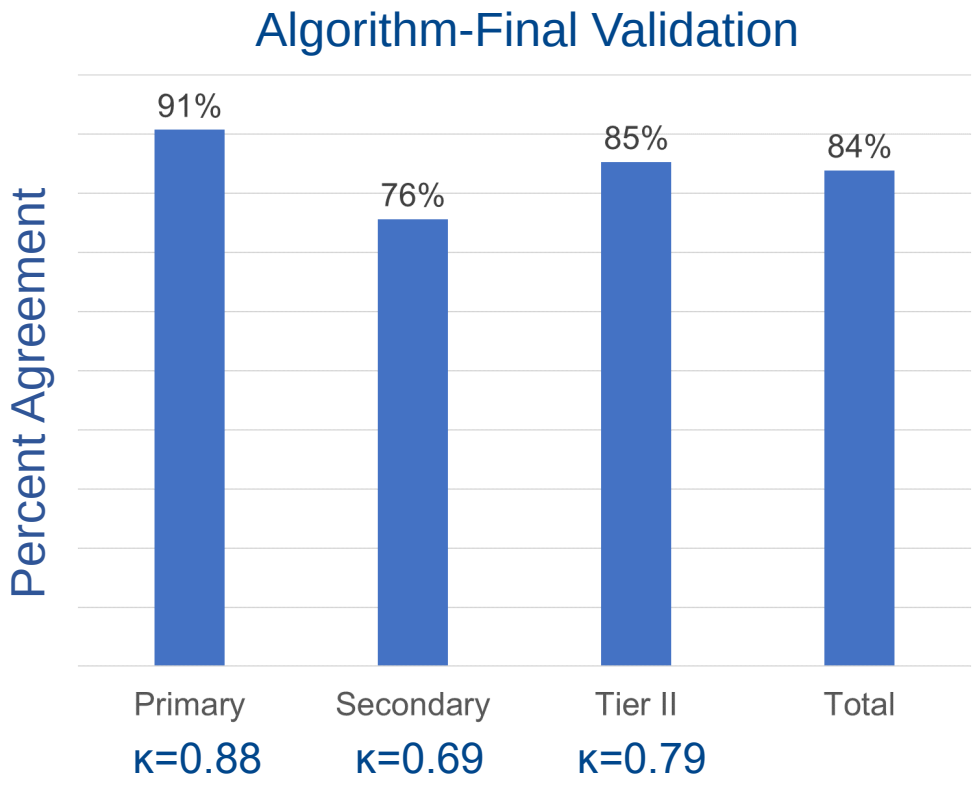
D – Research Design and Methodology

E – Translational Science

F – Research Capacity and Infrastructure

National Institutes of Health

# 2018 Overall Agreement (Algorithm and ICs): New Grants




| | Algorithm-NCMRR (20% sample) | | | NCMRR-Final (all new grants) | | |
|---|---|---|---|---|---|---|
| | # Incorrect | % Difference | % Agreement | # Incorrect | % Difference | % Agreement |
| Primary | 20 | 3.9% | 96% | 28 | 5.4% | 95% |
| Secondary | 55 | 10.6% | 89% | 72 | 13.9% | 86% |
| Tier II | 28 | 5.4% | 95% | 56 | 10.8% | 89% |
| Total | 103 | 6.6% | 93% | 156 | 10.1% | 90% |

## Disagreement Across Years

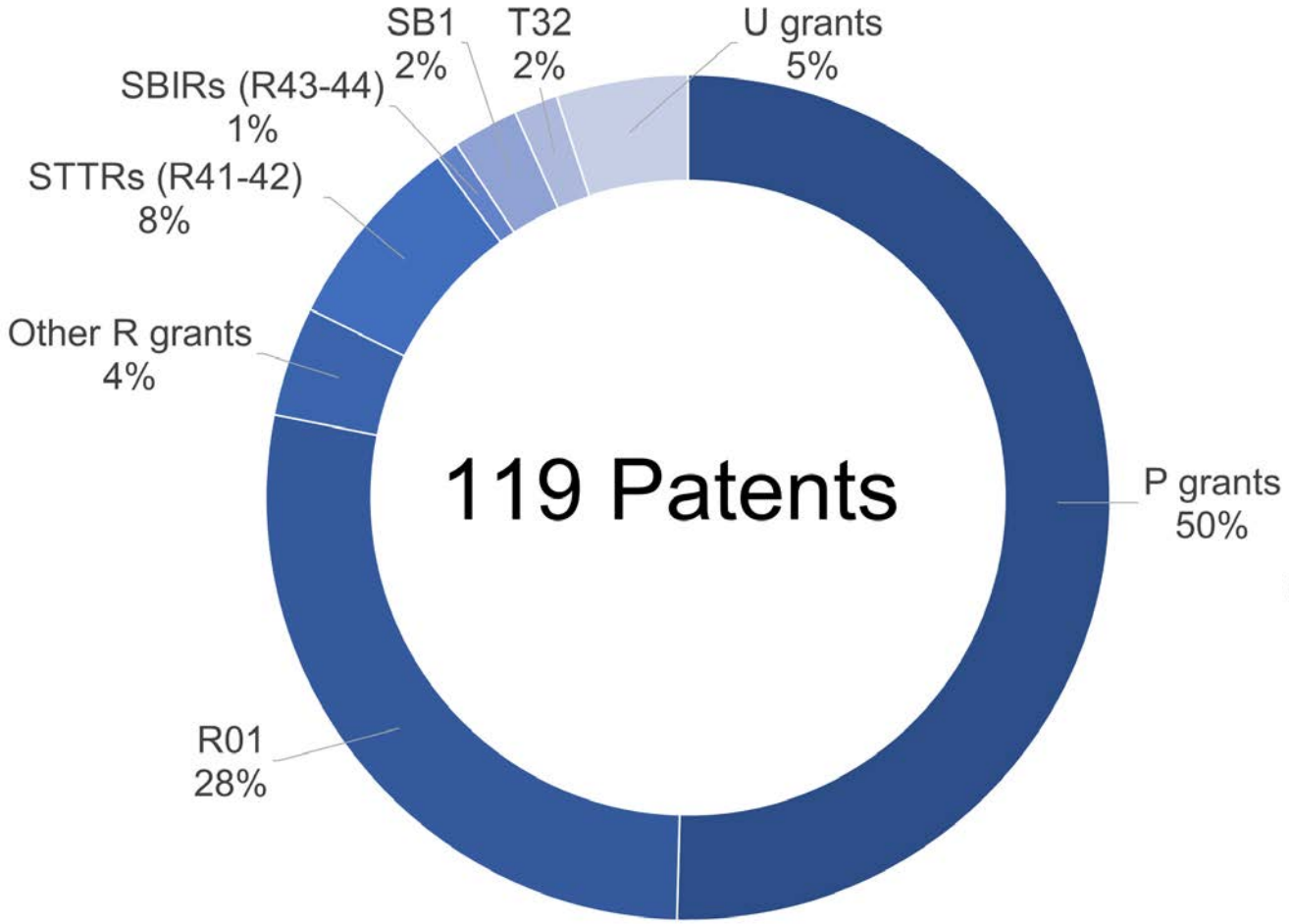
| | FY 2015 | | FY 2016 | | FY 2017 | | FY 2018 | |
|---|---|---|---|---|---|---|---|---|
| | # | % | # | % | # | % | # | % |
| Primary | 30 | 2.2% | 1 | 0.2% | 6 | 1.3% | 48 | 9.3% |
| Secondary | 74 | 5.4% | 13 | 2.6% | 16 | 3.6% | 126 | 24.4% |
| Tier II | 48 | 3.5% | 0 | 0.0% | 9 | 2.0% | 76 | 14.7% |
| Total Inaccurate | 152 | 3.7% | 14 | 1.0% | 31 | 2.3% | 250 | 16.1% |
| Total | 1360 | | 491 | | 446 | | 517 | |

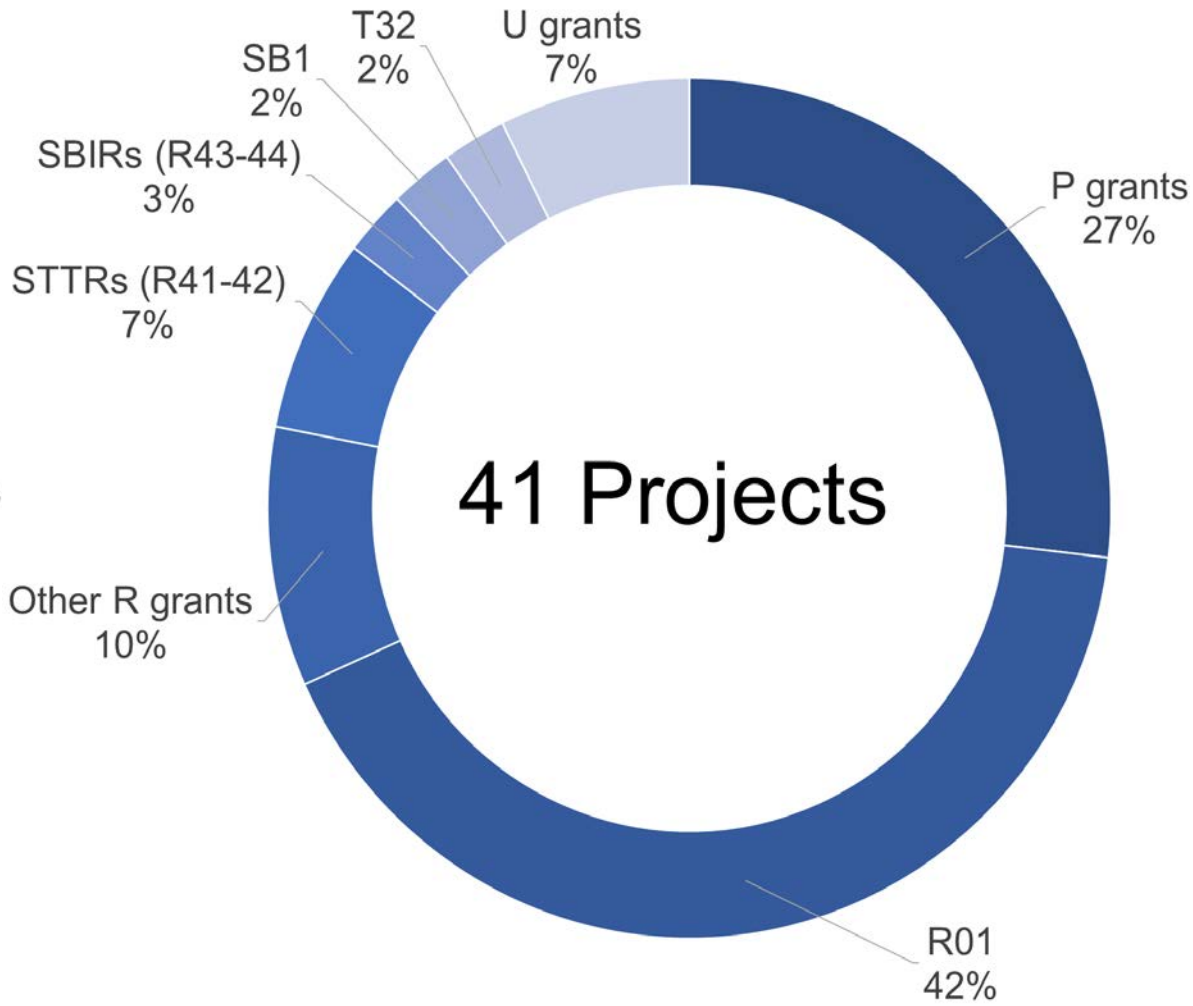
# 2018 Agreement per IC (%)

| Code | FIC | NCATS | NCCIH | NCI | NEI | NHLBI | NIA | NIAAA | NIAMS | NIBIB | NICHD | NIDA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Primary | 100.0 | 0.0 | 91.7 | 90.9 | 100.0 | 88.9 | 90.1 | 87.5 | 100.0 | 100.0 | 86.5 | 92.3 |
| Secondary | 0.0 | 100.0 | 91.7 | 58.2 | 91.7 | 66.7 | 84.5 | 75.0 | 88.2 | 50.0 | 66.2 | 84.6 |
| Tier II | 100.0 | 100.0 | 100.0 | 83.6 | 83.3 | 88.9 | 94.4 | 75.0 | 100.0 | 75.0 | 90.5 | 92.3 |
| Total Projects | 1 | 1 | 12 | 55 | 12 | 18 | 71 | 8 | 17 | 8 | 74 | 13 |

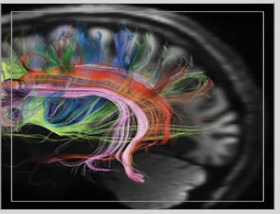
| Code | NIDCD | NIDCR | NIDDK | NIEHS | NIGMS | NIMH | NIMHD | NINDS | NINR | NLM | OD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Primary | 97.3 | 66.7 | 62.5 | 100.0 | 50.0 | 52.6 | 80.0 | 97.9 | 100.0 | 100.0 | 100.0 |
| Secondary | 90.7 | 33.3 | 62.5 | 100.0 | 50.0 | 21.1 | 60.0 | 85.6 | 83.3 | 100.0 | 100.0 |
| Tier II | 92.0 | 50.0 | 75.0 | 100.0 | 100.0 | 78.9 | 40.0 | 71.1 | 100.0 | 0.0 | 100.0 |
| Total Projects | 75 | 6 | 8 | 1 | 2 | 19 | 5 | 97 | 12 | 1 | 1 |

# 2018 Patent Data




National Institutes of Health



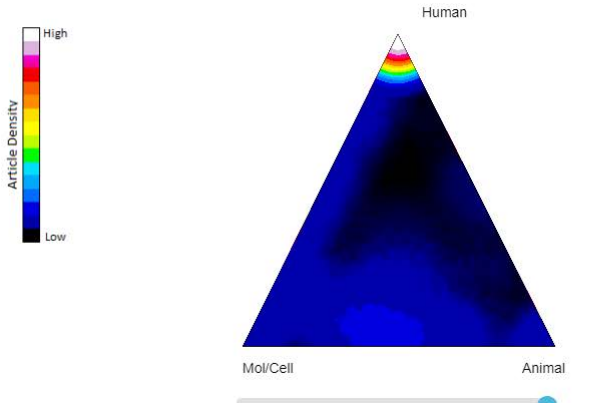

# Clinical Impact and Interdisciplinary Collaboration

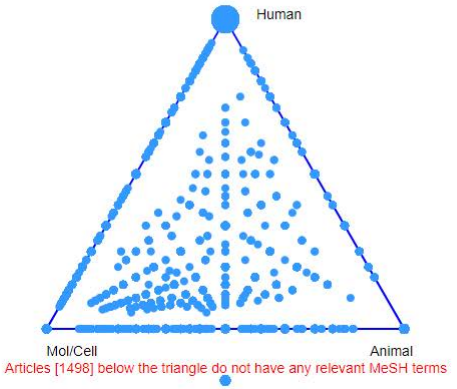
National Institutes of Health

# Draft Clinical Impact: 2018 Rehabilitation Portfolio

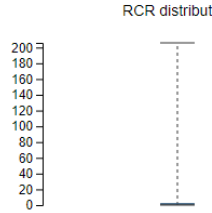
- **4 (0.01%)** are not found by *iCite* (published in 1979).
- **1,498 (5%)** have no Human, Animal or Molecular/Cellular MeSH terms so sit outside the triangle.
- **26,147 (95%)** have H, A or M/C MeSH terms and are shown in the triangle of biomedicine.
  - Of these, 12,556 (48% of those with H, A or M/C MeSH terms) have only Human MeSH terms (probably as you would expect).
- **7,202 (26%)** of the 27,645 publications in *iCite* have been cited by a clinical trial or guideline.
- **Mean RCR = 2.12, median RCR = 1.24**

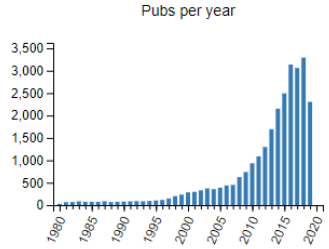
| Total Pubs | Pubs/Year | Avg. Human | Avg. Animal | Avg. Mol/Cell | Median RCR | Avg. APT | Cited by Clin. |
|---|---|---|---|---|---|---|---|
| 27,645 | 674.27 | 0.55 | 0.17 | 0.22 | 1.24 | 0.29 | 7,202 |

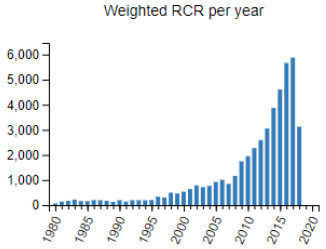



| Total Pubs | Pubs/Year | Cites/Year | | | | Relative Citation Ratio (RCR) | | | | Weighted RCR |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MAX | MEAN | SEM | MED | MAX | MEAN | SEM | MED | |
| 27,645 | 674.27 | 582.00 | 4.47 | 0.06 | 2.50 | 205.88 | 2.12 | 0.03 | 1.24 | 46437.27 |

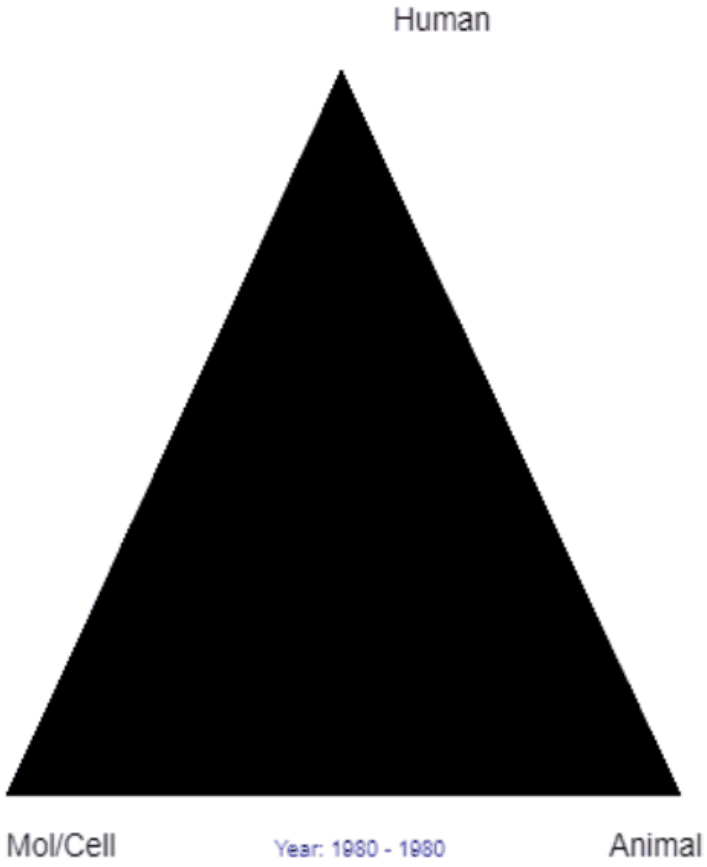



National Institutes of Health

# Draft Clinical Impact 2018: The Animated Version

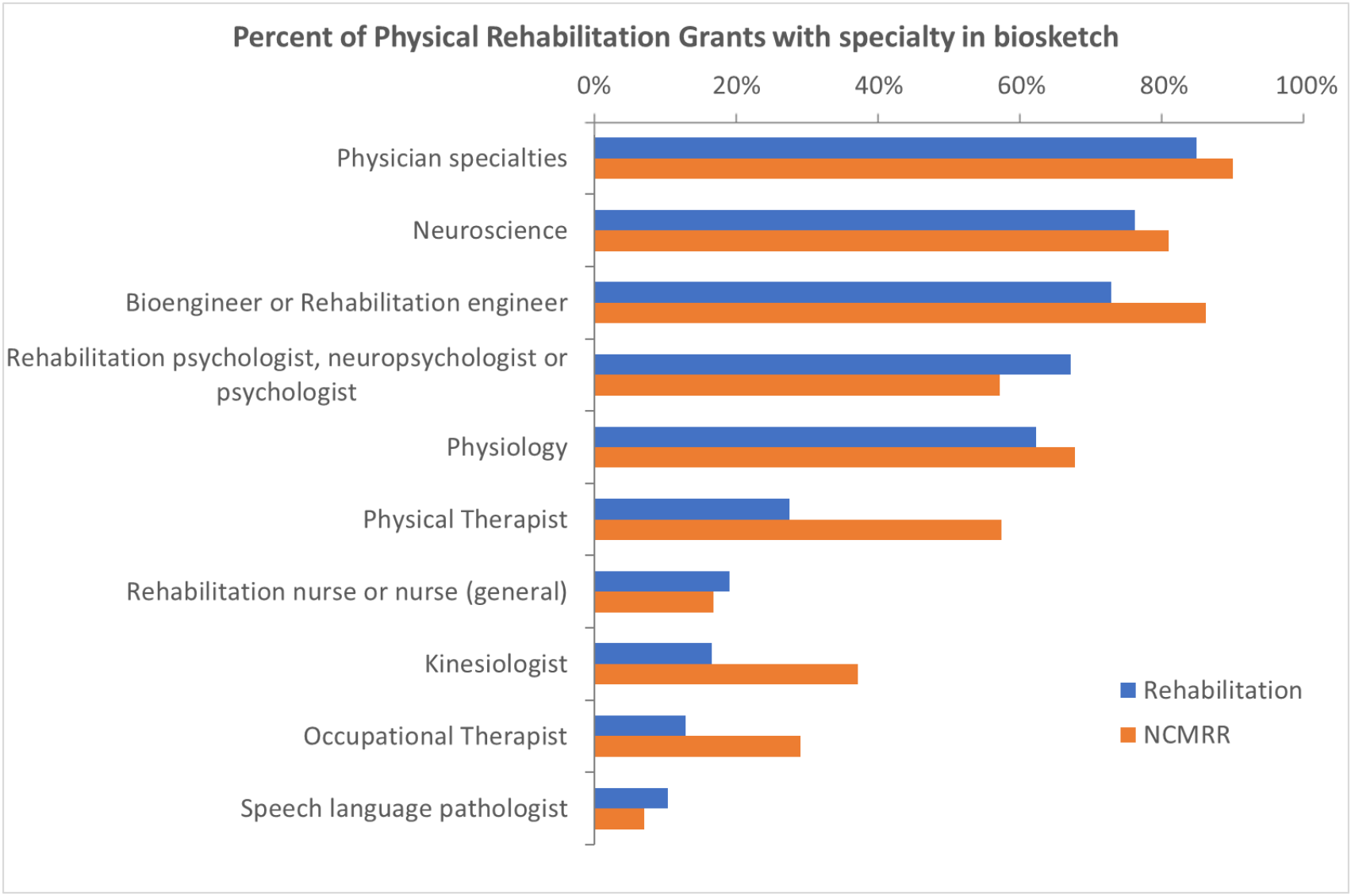


# Draft Interdisciplinary Collaboration: Rehabilitation 2018

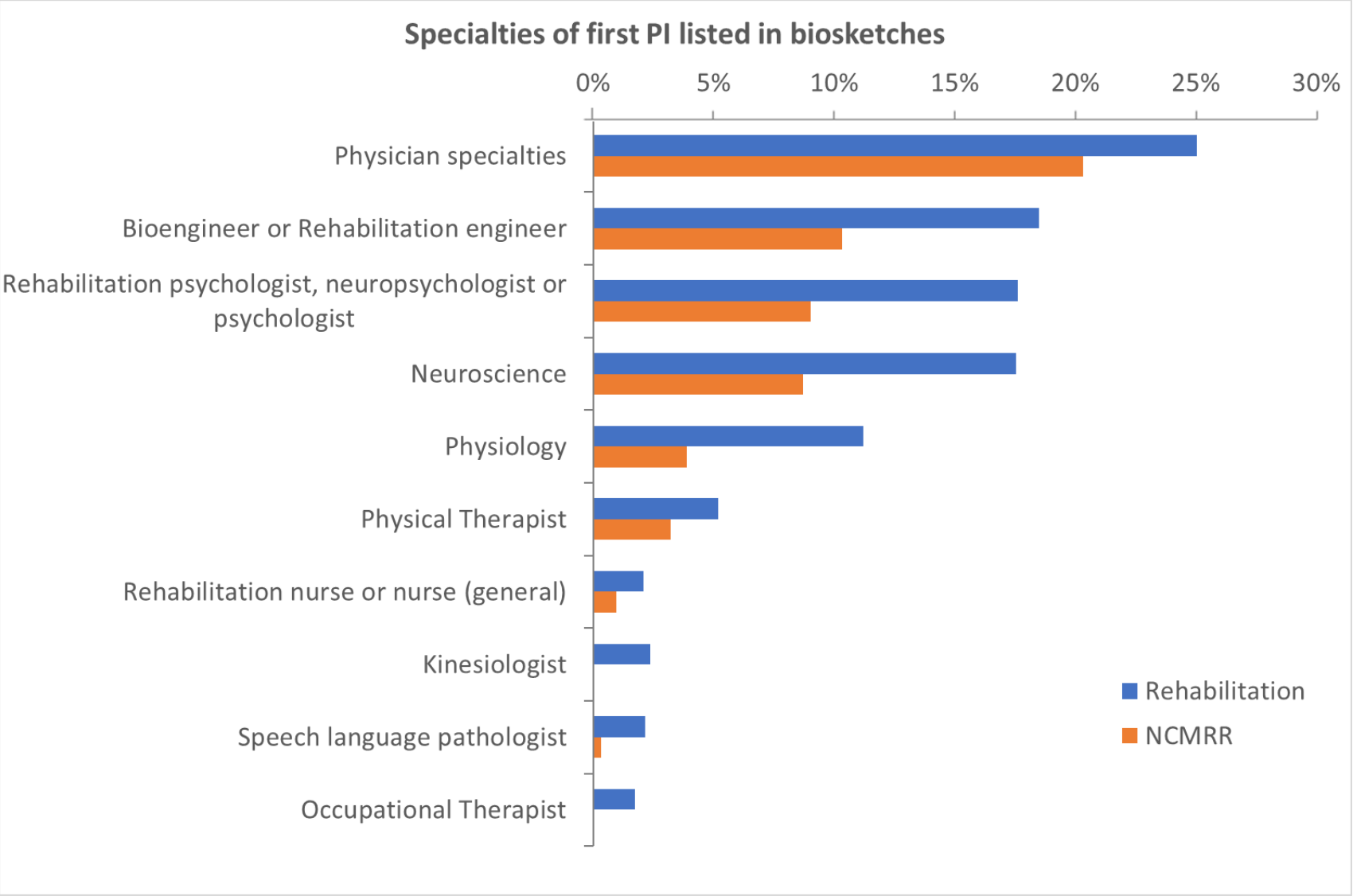
- Biosketch Analysis of the 2018 portfolio in collaboration with the Office of Portfolio Analysis
  - Used 1,591 Rehabilitation and 366 NCMRR ApplIDs, from 1,417 and 316 Grants respectively, provided by NICHD.
  - Identified awarded type 1 and 2 grant applications (as these contain full bio-sketches). Biosketches identified for 1,397 Rehabilitation Grants (99%) and 311 NCMRR Grants (98%)
  - Used text mining to identify rehabilitation specialties.
  - Identified one ApplID for each grant number. Within grant numbers, sorted by year (oldest first), Application type (Type 1s first), ApplID (lowest [oldest] first). Selected first ApplID for each project number for analysis.
  - Identify main specialties by looking at broad category with the most matches for each PI.

| Number of ApplIDs | Number of Rehabilitation Grants | Number of NCMRR Grants |
|---|---|---|
| 1 | 1,214 | 278 |
| 2 | 110 | 19 |
| 3 | 55 | 10 |
| 4 | 18 | 4 |
| **Total grants used for analysis** | **1,397** | **311** |

# Draft Percent of physical rehabilitation applications with specialty in biosketch



# Draft Specialties listed by first PI* in medical rehabilitation biosketches




## Next Steps:

1. Refine specialties as needed
2. Deal with duplicate searches due to key word strategy in algorithm
3. Review biosketch extraction process

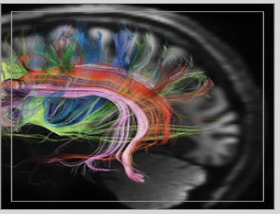
* First PI is the first identified person in the biosketch section of the application

National Institutes of Health

# Draft Collaborations (Rehabilitation)

| First PI specialty | Number of Grants | Percent of Grant applications featuring other investigators referencing: Physician specialties | Neuroscience | Bioengineer or Rehabilitation engineer | Rehabilitation psychologist, neuropsychologist or psychologist | Physiology | Physical Therapist | Rehabilitation nurse or nurse (general) | Kinesiologist | Occupational Therapist | Speech language pathologist |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Physician specialties | 213 | 99% | 87% | 73% | 77% | 64% | 26% | 22% | 11% | 17% | 5% |
| Bioengineer or Rehabilitation engineer | 118 | 66% | 75% | 97% | 37% | 69% | 29% | 3% | 27% | 11% | 8% |
| Rehabilitation psychologist, neuropsychologist or psychologist | 73 | 75% | 77% | 58% | 93% | 53% | 18% | 8% | 10% | 8% | 21% |
| Neuroscience | 46 | 76% | 100% | 76% | 61% | 76% | 17% | 7% | 7% | 7% | 2% |
| Physical Therapist | 27 | 96% | 81% | 93% | 41% | 85% | 100% | 15% | 56% | 44% | 7% |
| Physiology | 24 | 79% | 75% | 75% | 54% | 92% | 38% | 4% | 33% | 13% | 4% |
| Occupational Therapist | 3 | - | - | - | - | - | - | - | - | - | - |
| Kinesiologist | 3 | - | - | - | - | - | - | - | - | - | - |
| Speech language pathologist | 0 | - | - | - | - | - | - | - | - | - | - |
| Rehabilitation nurse or nurse (general) | 0 | - | - | - | - | - | - | - | - | - | - |
| Multiple Specialties | 65 | 86% | 83% | 72% | 63% | 68% | 31% | 11% | 15% | 11% | 12% |
| Other Specialties only | 818 | 84% | 70% | 68% | 67% | 56% | 26% | 23% | 15% | 11% | 11% |
| All grants with multiple biosketches | 1390 | 84% | 76% | 72% | 66% | 61% | 28% | 19% | 16% | 13% | 10% |

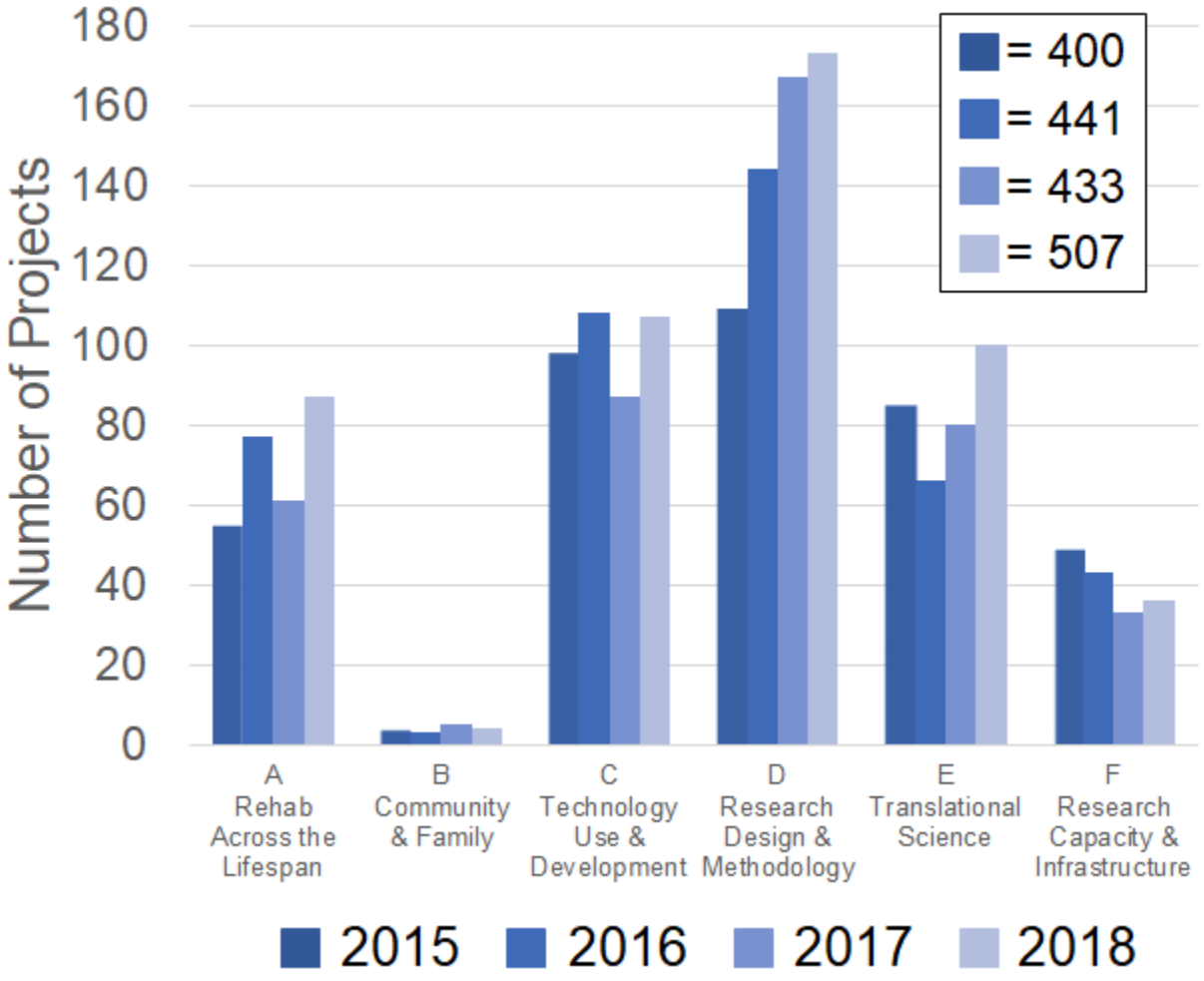
- Table shows first PI (IDed first in biosketch) main specialty in the first column. Cells contain the percentage of applications where other investigators in the bio-sketch report a specialty.
  - Applications where a PI has multiple specialties are included in the ‘Multiple’ category.
  - The ‘Other Specialty’ category is PIs whose bio-sketch returns no matches for the Rehabilitation search terms used.
  - Percentages not calculated where base is less than 10.

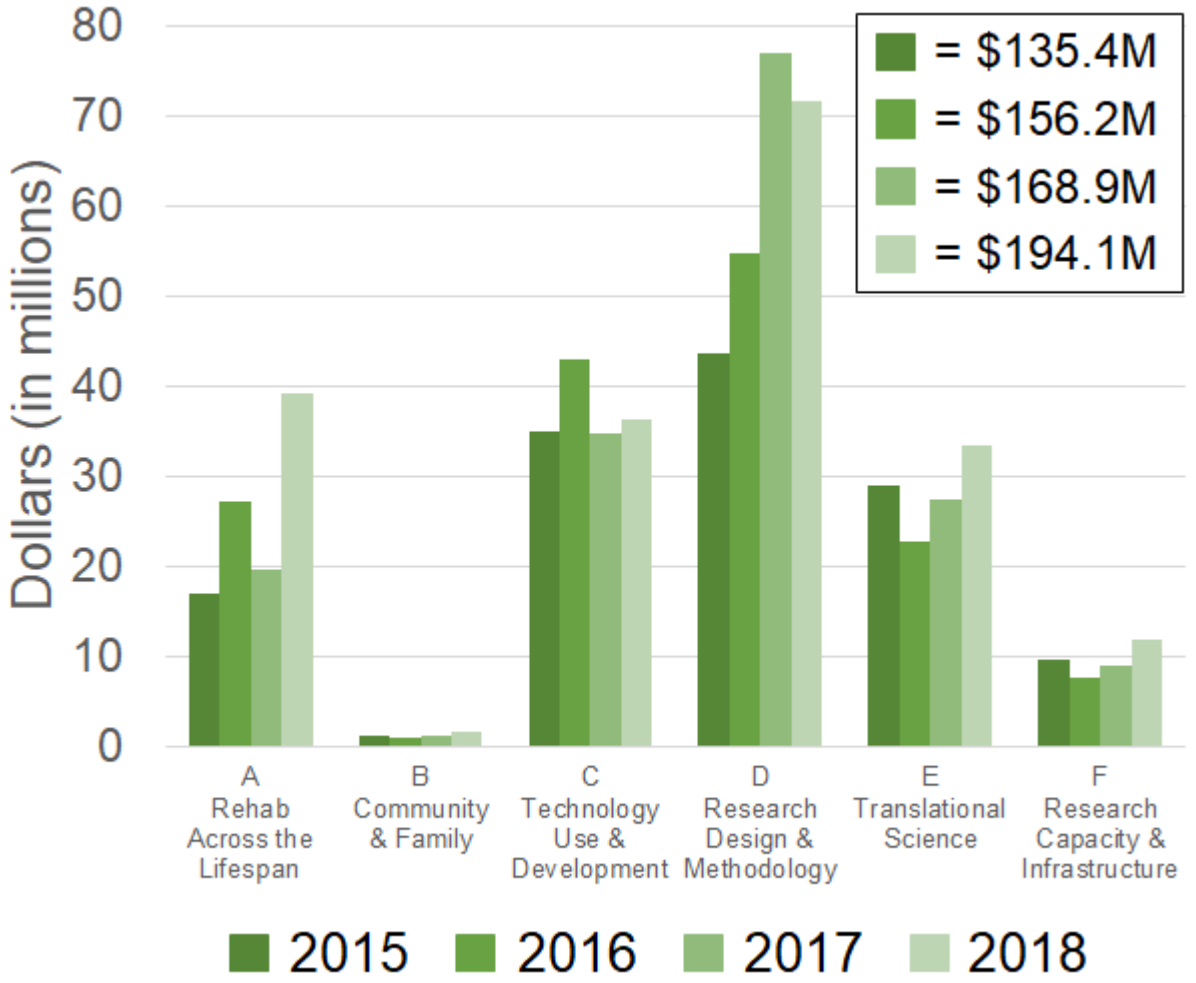



# Year-to-Year Trends

National Institutes of Health

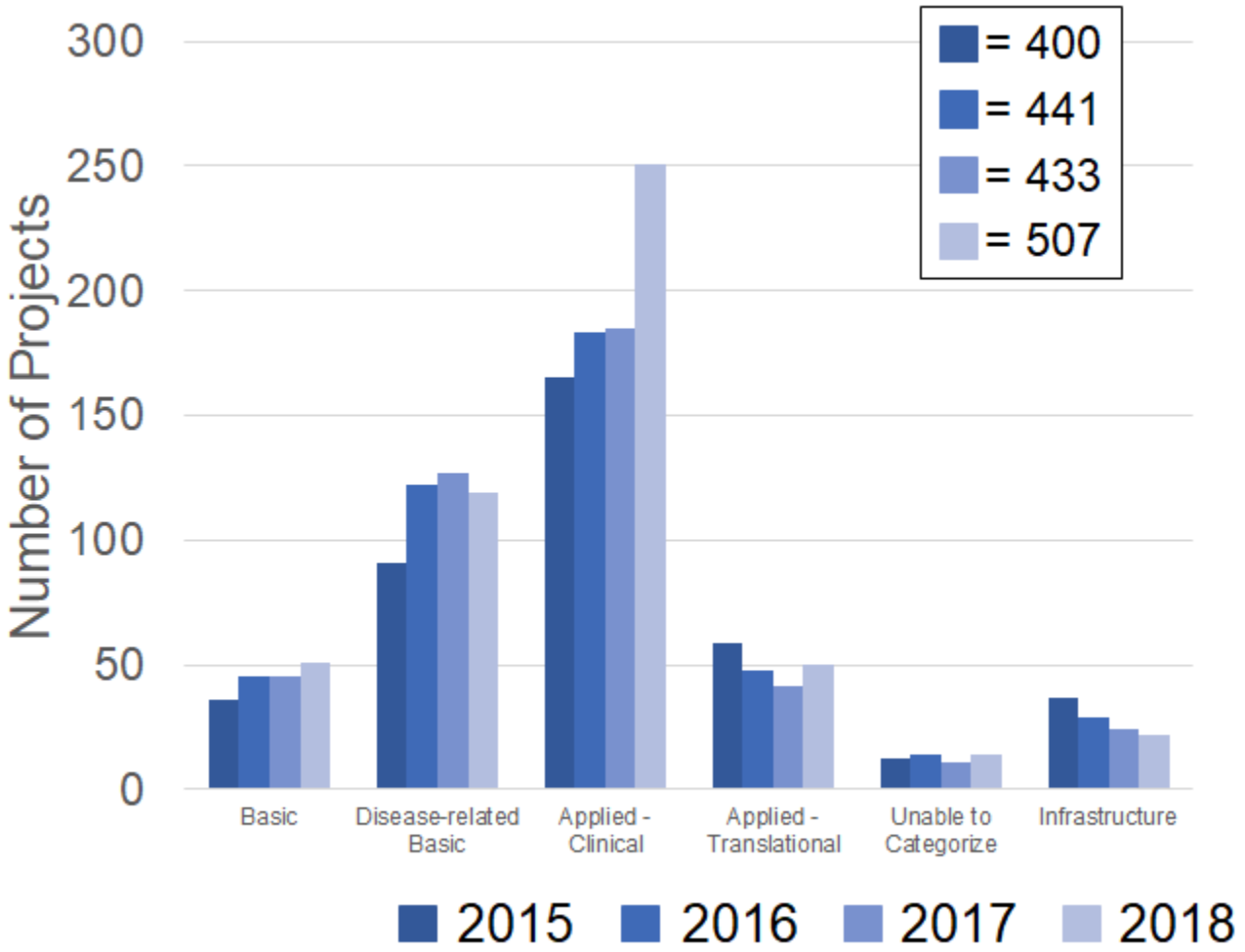
# Trends (2015-2018): Primary Category (types 1&2 only)

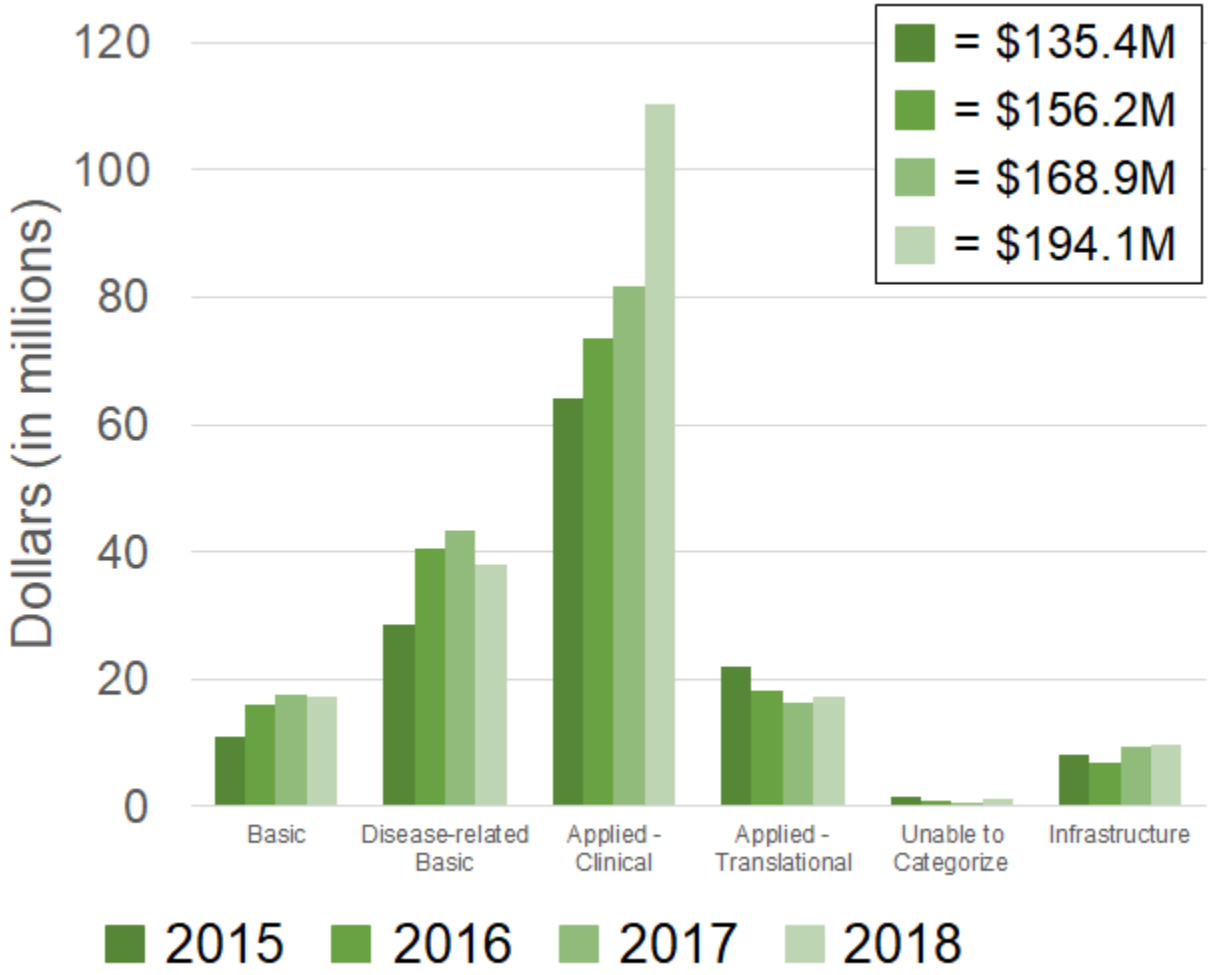



National Institutes of Health

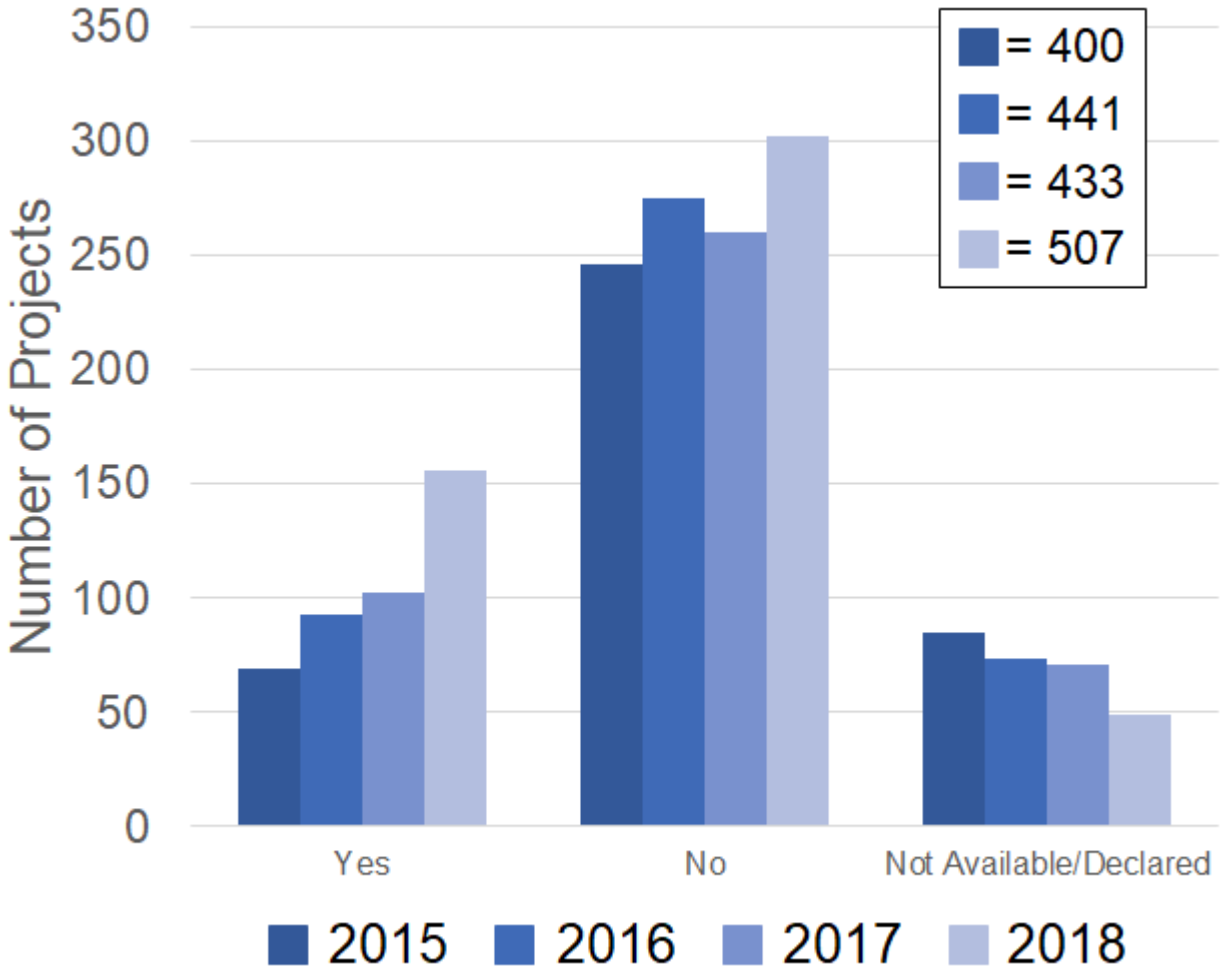
# Trends (2015-2018): Tier II Category (types 1&2 only)

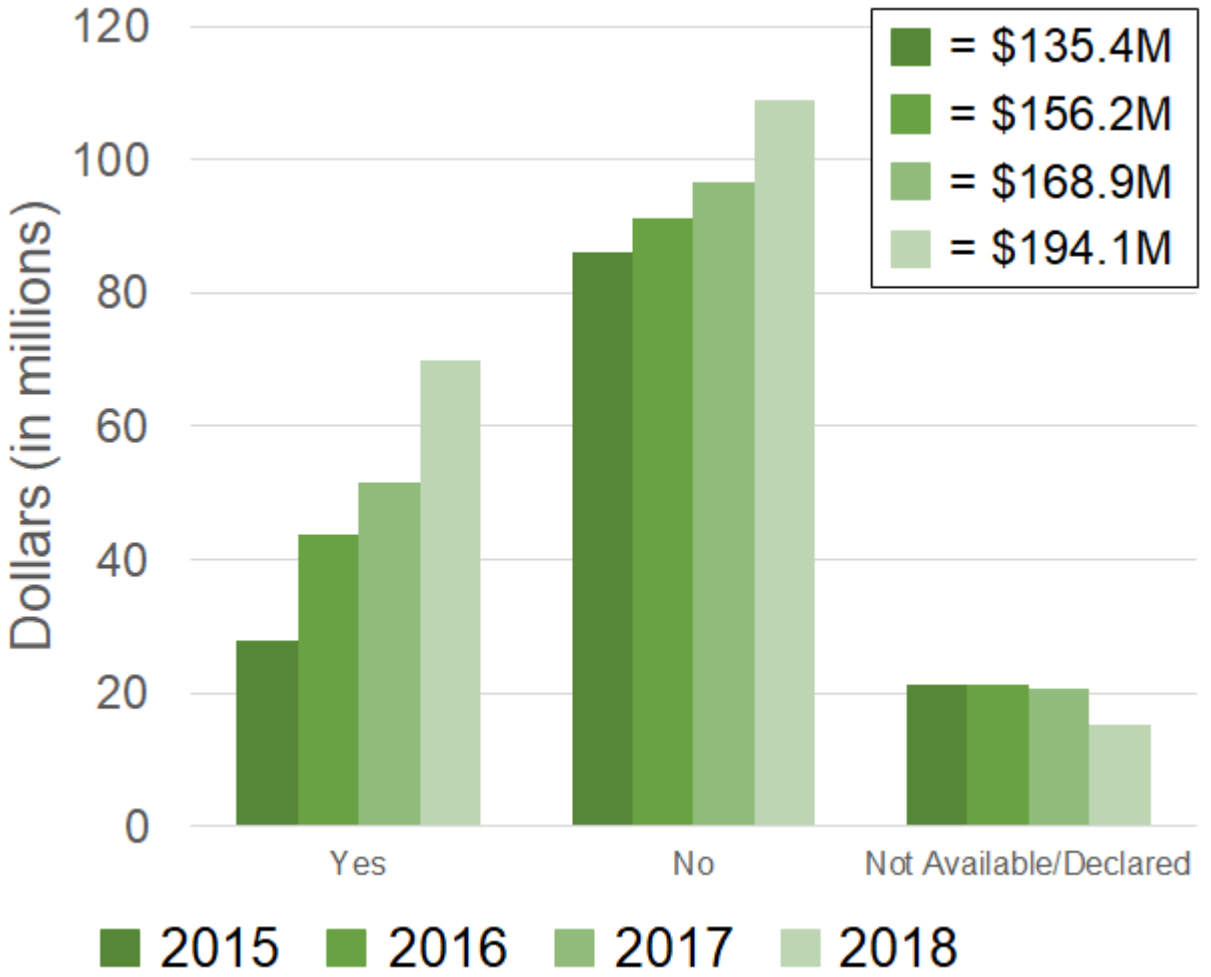



National Institutes of Health

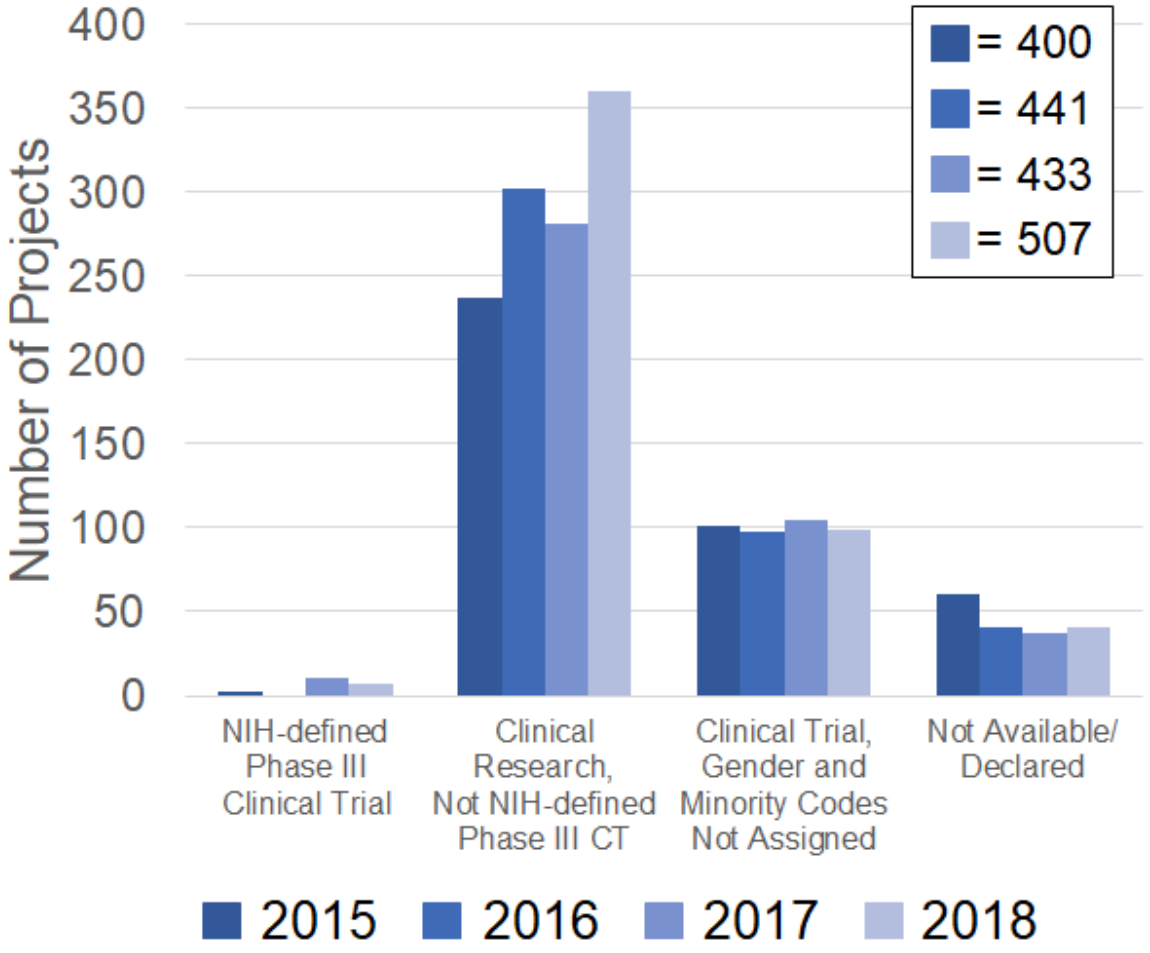
# Trends (2015-2018): Clinical Trials (types 1&2 only)

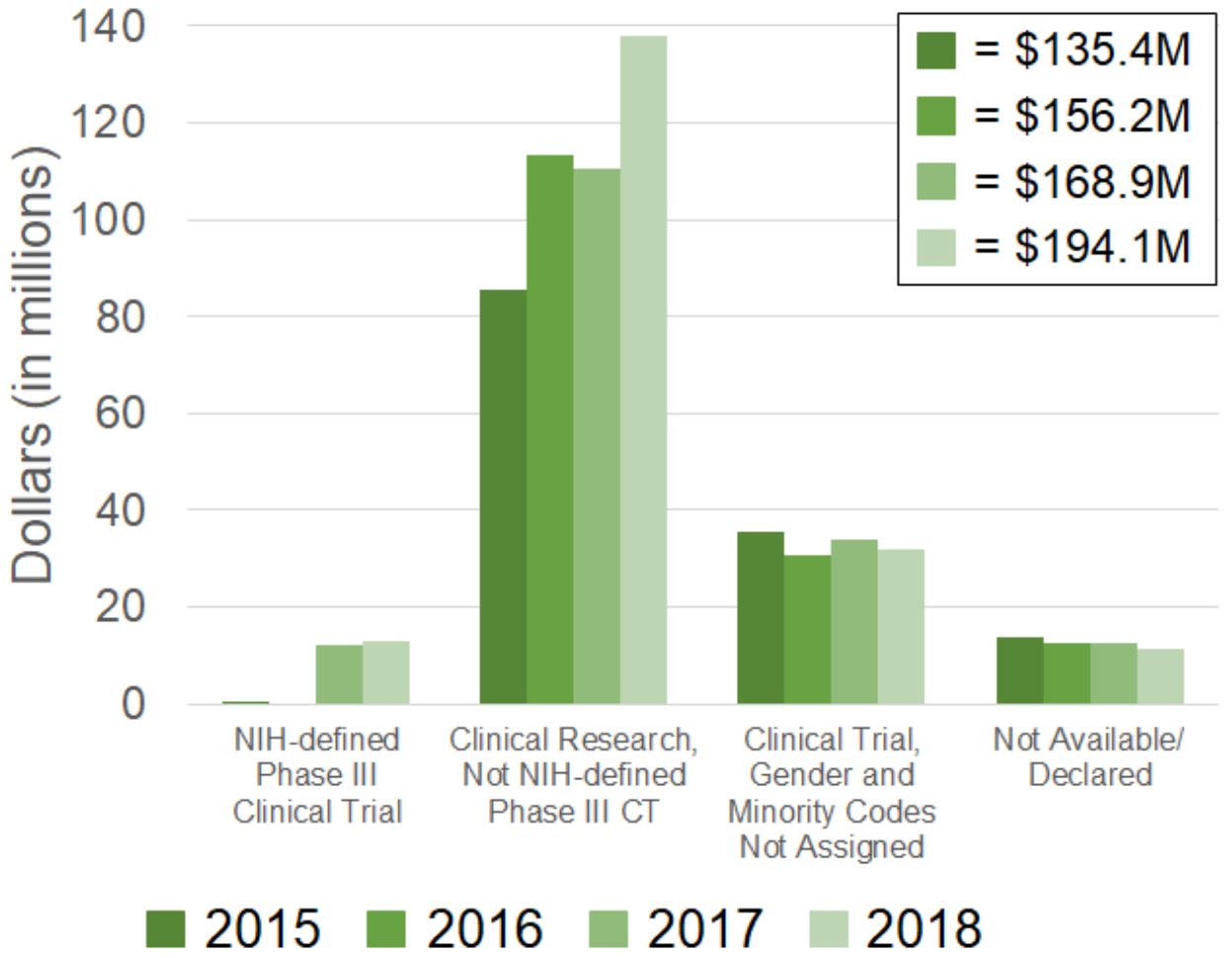



National Institutes of Health

# Trends (2015-2018): Phase III Clinical Trials (types 1&2 only)




National Institutes of Health

# Challenges

- Categorical definitions and weighting can change each year or could be revisited as a result of this analysis
- Category does not include all projects that have a rehabilitation focus and may include some that are not rehabilitation-related
- Changes in overarching NIH budget could have impacts on all portfolios
- Changes in other federal funding agencies could impact NIH portfolios

# Questions

jennifer.Jackson@nih.gov