# OFFICE OF PORTFOLIO ANALYSIS

## September 5, 2012

# The Office of Portfolio Analysis (OPA)

**OPA was established in 2011 to provide multiple services:**

- Scientific portfolio analysis…
  - of all Common Fund initiatives
  - in response to requests by senior leadership
- Coordinate portfolio analysis activities across NIH
- Train NIH staff to promote the effective use of analytical tools
  - Regularly scheduled courses
  - Ad hoc consultations
- Improve portfolio analysis at NIH (see next slide)

# Improve Portfolio Analysis at NIH

- Solve NIH-wide limitations in cross-referencing databases
- Build new computational tools that address specific NIH needs
- Model NIH output and health impact accurately
  - Advanced bibliometrics
- Track progress in the new field of “science of science” and find useful methods and synergies among the parallel efforts in…
  - Academia
  - Other government agencies
  - Private sector
- Adopt new methods that meet NIH needs and provide training opportunities for portfolio analysis stakeholders at NIH

# Scientific Portfolio Analysis at NIH

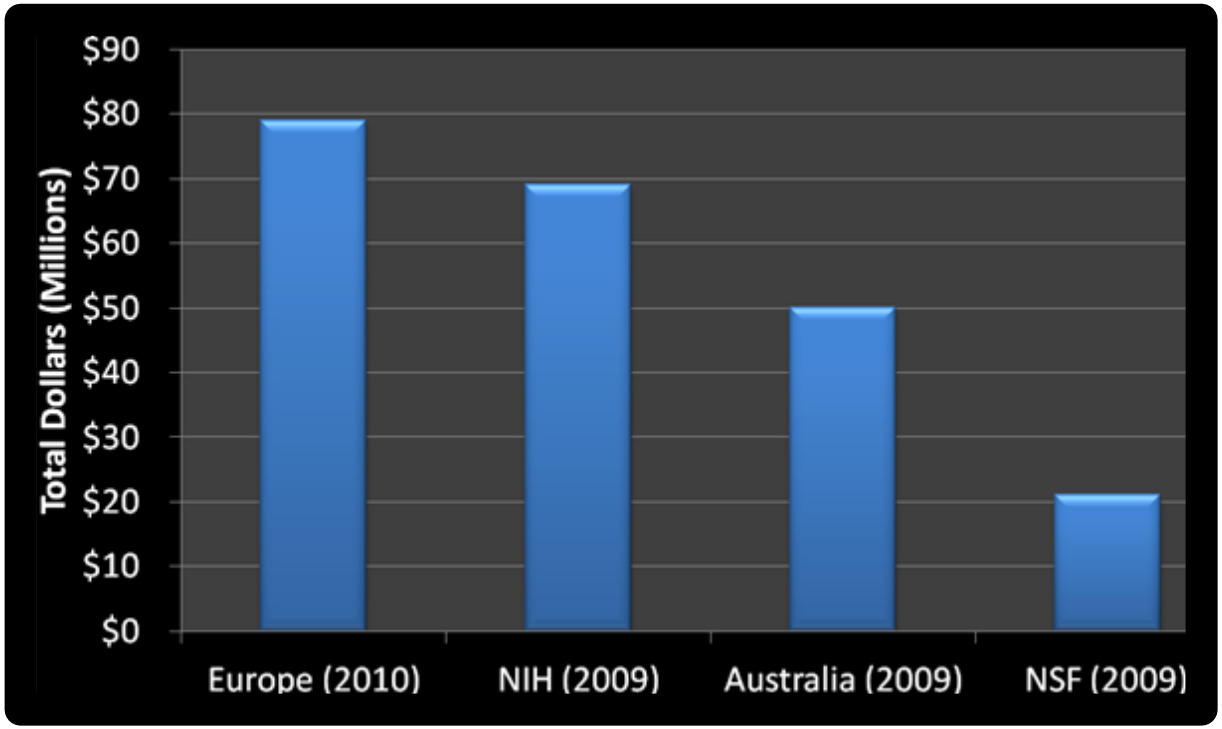
Data-driven approaches to program development

**What is the best way to accelerate scientific progress?**

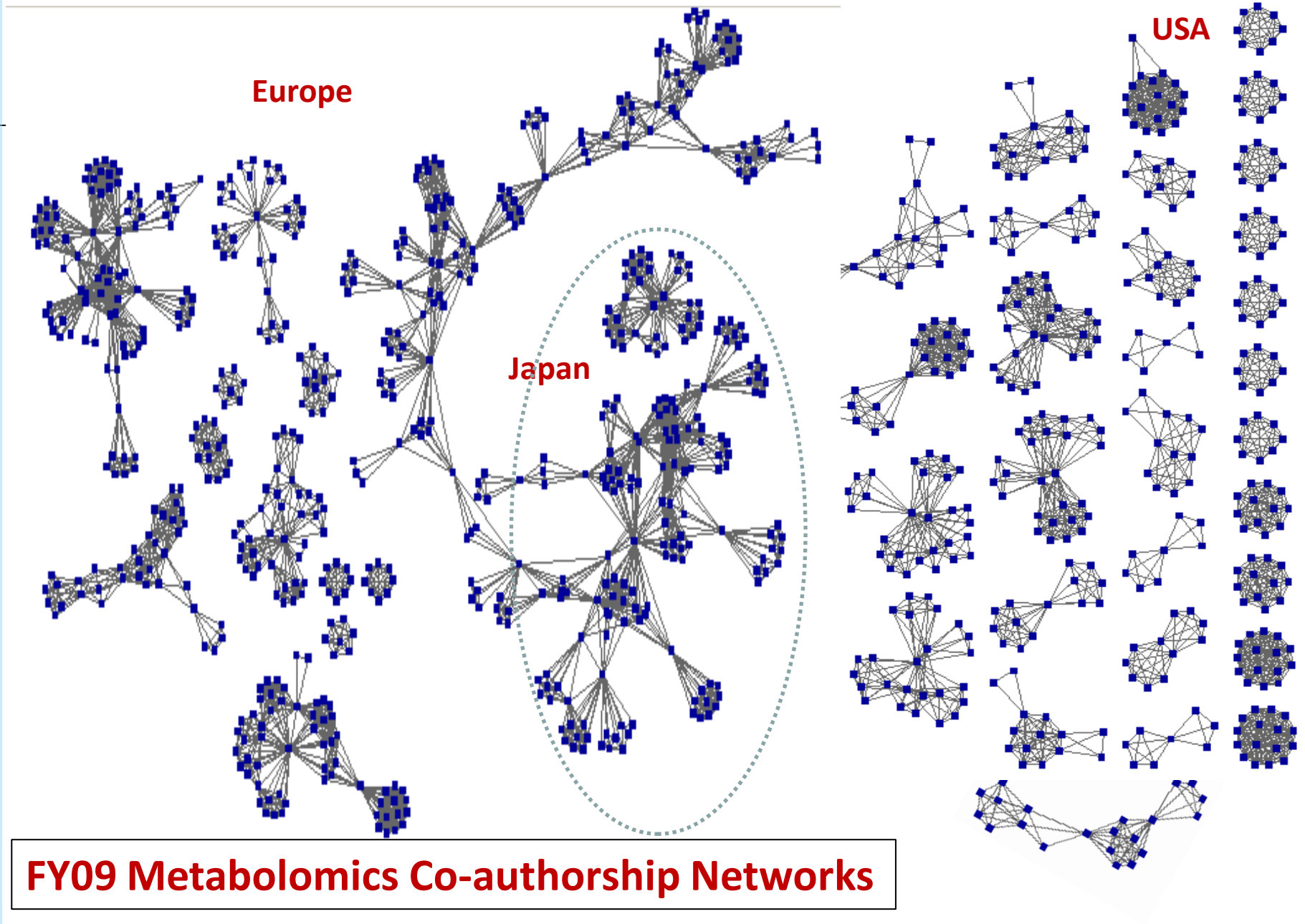
**An example:**

- OPA analysis of 38 Metabolomics Centers in the US, most of which receive funding from NIH (~$70M in 2010)

# $225M Global Investment in Metabolomics in FY10




- Netherlands Metabolomics Center: $67M*
- BBSRC, UK  Plant and Microbial Metabolomics: $10.4M
- Canadian Human Metabolome Database: $8.1M




# FY09 Metabolomics Co-authorship Networks

# Scientific Portfolio Analysis at NIH

**What is the best way to accelerate scientific progress?**

- OPA analysis of 38 Metabolomics Centers in the US, most of which receive funding from NIH (~$70M in 2010)
- These Centers have overlapping goals, yet operated in isolation with little coordination or collaboration
- Portfolio analysis shaped a new Metabolomics RFA aimed at improving coordination and leveraging existing resources

# Coordination of Portfolio Analysis Efforts at NIH

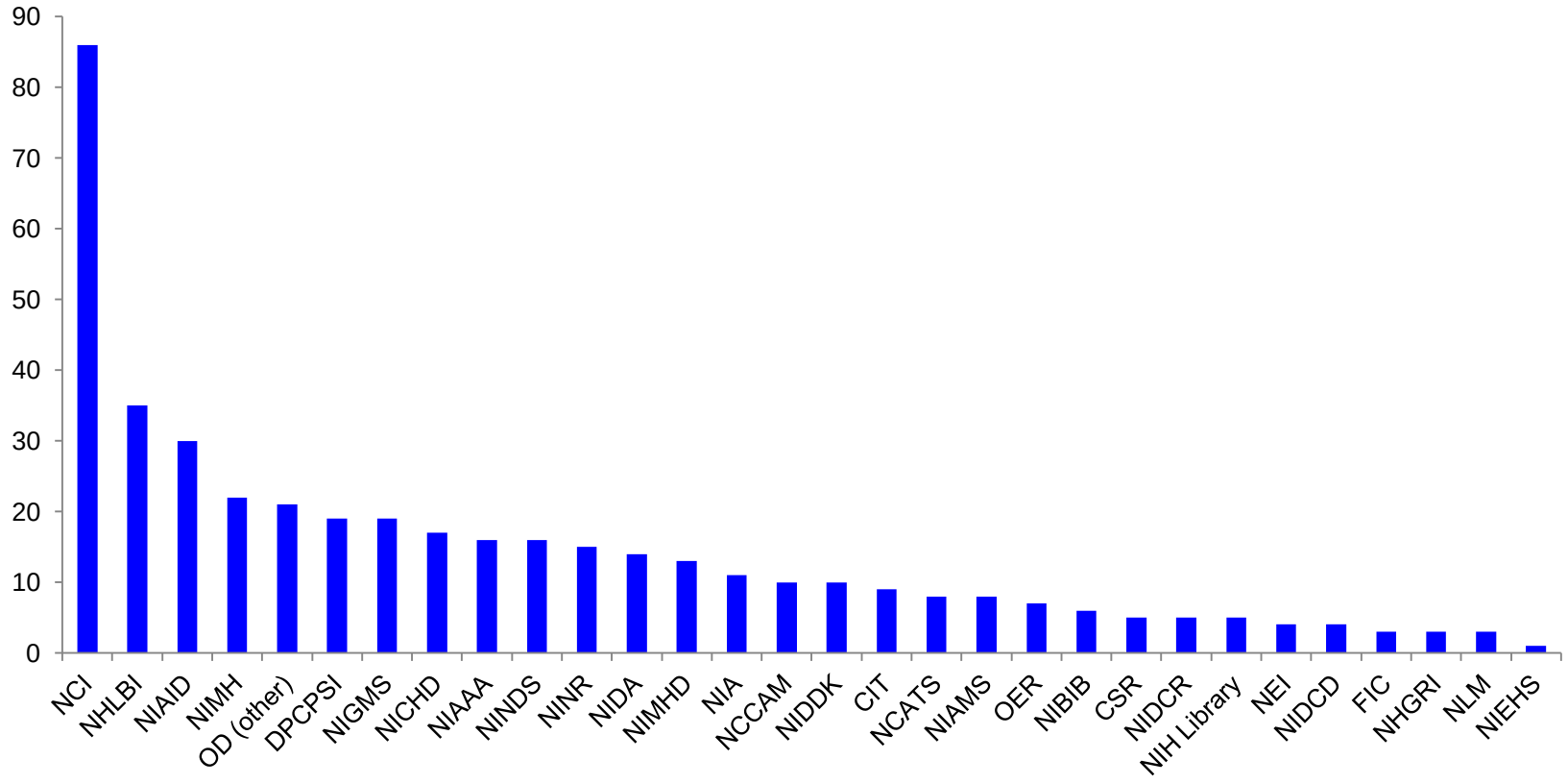
- ✓ Portfolio Analysis Workshop (February 6, Natcher Conf. Ctr.)
- ✓ Portfolio Analysis Symposium (July 23-24, Natcher Conf. Ctr.)
- ✓ Build a computer lab to tailor existing and new computational tools to NIH needs, and to train NIH staff in their use
- ➢ Centralized web-based repository to disseminate computational tools
  - http://dpcpsi.nih.gov/portfolio_analysis/
- ➢ Standing trans-NIH Working Group

# Goals of the Portfolio Analysis Workshop Feb 5-6, 2012

- Discuss perceived needs in portfolio analysis
- Use needs assessment to plan the Portfolio Analysis Symposium (July 23-24, Natcher Conf. Ctr.)
- Cover topics of broad interest to NIH decision-makers, including:
  - Strategic planning
  - Uses of portfolio analysis
  - Overlap in NIH portfolios
  - Measuring impact
  - New portfolio analysis tools
  - Identification of emerging areas

# Outcomes of the Portfolio Analysis Workshop

- Full registration within 48 hrs of the announcement
- >500 participants – approved for ESA training credit



# Outcomes of the Portfolio Analysis Workshop

**Survey results**

- Topics included:

| | **Highest priority:** |
|---|---|
| ○ Measuring impact | **47%** |
| ○ Gaps and overlap in NIH portfolios | **30%** |
| ○ Identification of emerging areas | **18%** |
| ○ Categorizing portfolios | **5%** |

- What needs should OPA try to address?

| | |
|---|---|
| ○ Build better tools / easier to use tools | **65%** |
| ○ Provide training and support | **50%** |
| ○ Develop targeted case studies | **15%** |

# Goals of the Portfolio Analysis Symposium July 23-24, 2012

- Bring outside experts from academia, government, and the private sector to discuss and demo state-of-the-art approaches in scientific portfolio analysis
- Choose those with expertise in areas identified as critical and of greatest interest to portfolio analysis stakeholders at NIH
  - Measuring Impact
  - Identifying Gaps and Impact
  - Identifying Emerging Areas
- Facilitate the development of collaborations that address NIH needs in scientific portfolio analysis

# Outcomes of the Portfolio Analysis Symposium

- Links to the Symposium agenda and archived videocast are available on the OPA web site:
  http://dpcpsi.nih.gov/opa/index.aspx

- Outside experts presented new tools and approaches in each topic; one of these tools ($Sci^2$) was particularly well received by Symposium attendees and is being added to QVR

- The Symposium resulted in several collaborations. Examples include OPA initiatives in building tools for co-author network analysis, analysis of patent development, and tracking of patent licenses

# The Office of Portfolio Analysis (OPA)

**OPA was established in 2011 to provide multiple services:**

- Scientific portfolio analysis…
  - of all Common Fund initiatives
  - in response to requests by senior leadership
- Coordinate portfolio analysis activities across NIH
- Train NIH staff to promote the effective use of analytical tools
  - Regularly scheduled courses
  - Ad hoc consultations
- Improve portfolio analysis at NIH

# Build Computational Tools and Train NIH Staff

**12 laptops**
**Two workstations**
**Smartboard**
**Linux server**
**Videoconferencing capabilities**

Building 1, Room B301

**OPA TOOLS LAB**

# OPA Consultations with NIH Staff

| IC | Contact | Activity |
|---|---|---|
| NCCAM | John Williamson | Consult |
| NCI | John Hewes | Consult & Collaboration |
| NCI | Tanya Agurs-Collins | Consult |
| NHLBI | Marc Charette | Consult & Collaboration |
| NIAMS | Faye Chen | Consult |
| NIAMS | Anita Linde | Consult |
| NIDDK | Lisa Spain | Consult |
| NIGMS | Michelle Hamlet | Consult |
| NIGMS | Ward Smith | Consult |
| NIGMS | Kelley Smith | Consult & Training |
| NIMH | Yancy Bodenstein | Consult & Collaboration |
| NIMHD | Ligia Artiles | Consult & Training |
| NLM | Alan Vanbierlet | Consult &Training |

# Improve Portfolio Analysis at NIH

Database management


Tool development

Modeling NIH output

# Improve Portfolio Analysis at NIH

## DATABASE MANAGEMENT

**Our current ability to track awards, output, and health impact is limited**

- Funding of clinical trials
  - Inadequate linkage between ClinicalTrials.gov and IMPAC II data
  - NCI and CIT are collaborating to address this problem
- Patents and licensing of intellectual property
  - No database exists for NIH-funded patents and university licensing
  - OPA is addressing this in collaboration w/ NIH tech transfer officers
- Output of awards
  - Inadequate linkage between NIH awards and literature/citation data
  - OPA is developing a next-gen disambiguation tool

# Improve Portfolio Analysis at NIH

**TOOL DEVELOPMENT**

**New tools are needed to provide NIH staff, grant applicants, et al., with a current and accurate picture of NIH investments**

- In collaboration with Calvin Johnson's group at CIT, OPA has developed several new tools
- Examples include the Hierarchical Clustering & Classifier tools
  - Comparative Effectiveness Research (CER)
  - NIGMS Technology Development
  - Disruptive Proteomics Common Fund Analysis
  - Similarity Matrix of Standing Study Sections (SRGs)
    - Characterize the assignment of applications to SRGs
    - Evaluate how SRGs are presented to PIs

# Cancer Genetics Study Section vs. All 165 Study Sections

2011, 3 cycles

| Study Section | Code | Match Score |
|---|---|---|
| **Cancer Genetics** | **CG** | **1.000** |
| **Cancer Molecular Pathobiology** | **CAMP** | **0.509** |
| Tumor Cell Biology | TCB | 0.452 |
| **Molecular Oncogenesis** | **MONC** | **0.445** |
| **Tumor Progression and Metastasis** | **TPM** | **0.405** |
| Basic Mechanisms of Cancer Therapeutics | BMCT | 0.397 |
| Developmental Therapeutics | DT | 0.394 |
| Genetics of Health and Disease | GHD | 0.394 |
| Tumor Microenvironment | TME | 0.393 |
| Genomics, Computational Biology and Technology | GCAT | 0.365 |
| **Cancer Biomarkers** | **CBSS** | **0.364** |
| **Molecular Genetics B** | **MGB** | **0.362** |
| Chemo/Dietary Prevention | CDP | 0.360 |
| **Epidemiology of Cancer** | **EPIC** | **0.353** |
| Radiation Therapeutics and Biology | RTB | 0.346 |
| Drug Discovery and Molecular Pharmacology | DMP | 0.339 |
| Molecular Neurogenetics | MNG | 0.331 |
| **Cancer Etiology** | **CE** | **0.326** |
| **Molecular Genetics C** | **MGC** | **0.321** |
| **Molecular Genetics A** | **MGA** | **0.288** |

# Neural Basis of Psychopathology, Addictions and Sleep Disorders Study Section vs. All 165 Study Sections

2011, 3 cycles

| Study Section | Code | Match Score |
|---|---|---|
| **Neural Basis of Psychopathology, Addictions, and Sleep Disorders** | **NPAS** | **1.000** |
| **Adult Psychopathology and Disorders of Aging** | **APDA** | **0.528** |
| **Pathophysiological Basis of Mental Disorders and Addictions** | **PMDA** | **0.493** |
| Cognition and Perception | CP | 0.398 |
| Biobehavioral Regulation, Learning and Ethology | BRLE | 0.378 |
| Cognitive Neuroscience | COG | 0.368 |
| Clinical Neuroscience and Neurodegeneration | CNN | 0.363 |
| Child Psychopathology and Developmental Disabilities | CPDD | 0.353 |
| Neurobiology of Motivated Behavior | NMB | 0.341 |
| Molecular Neuropharmacology and Signaling | MNPS | 0.339 |
| **Developmental Brain Disorders** | **DBD** | **0.326** |
| **Neurotoxicology and Alcohol** | **NAL** | **0.294** |
| **Genetics of Health and Disease** | **GHD** | **0.189** |

# Improve Portfolio Analysis at NIH

Database management


Tool development


Modeling NIH output

# MODELING NIH OUTPUT

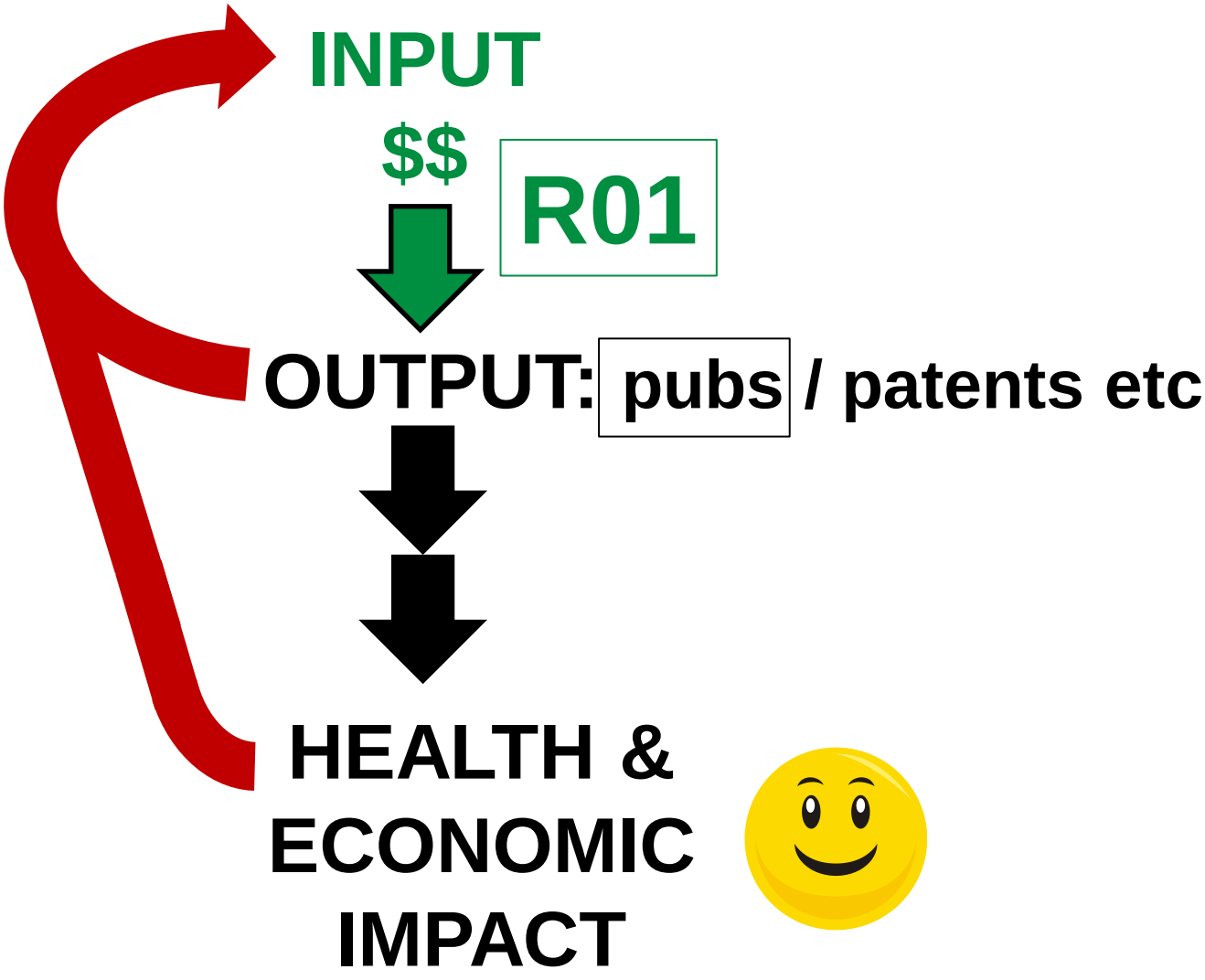
**INPUT**

**$$**

**OUTPUT: pubs / patents etc**

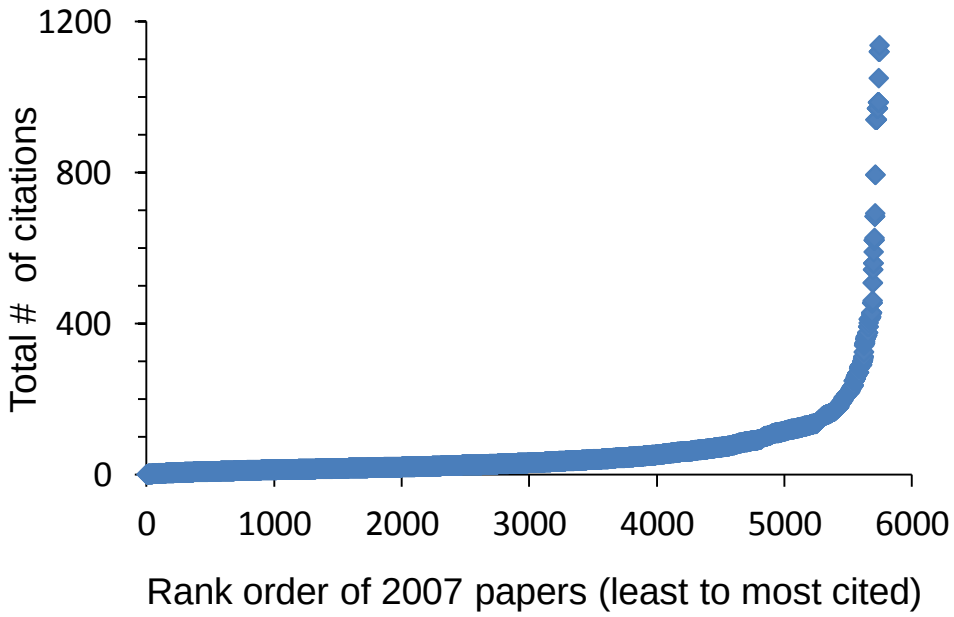
**HEALTH & ECONOMIC IMPACT**

# MODELING NIH OUTPUT: Advanced Bibliometrics

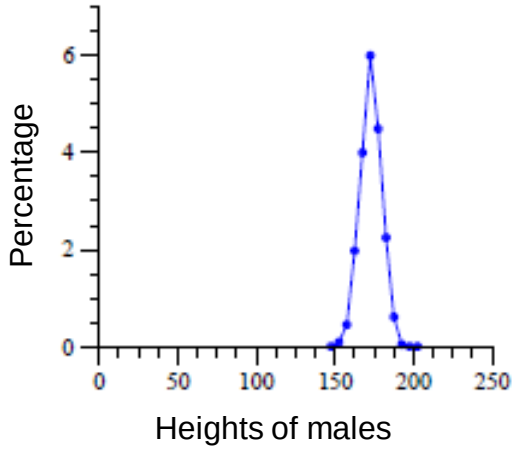


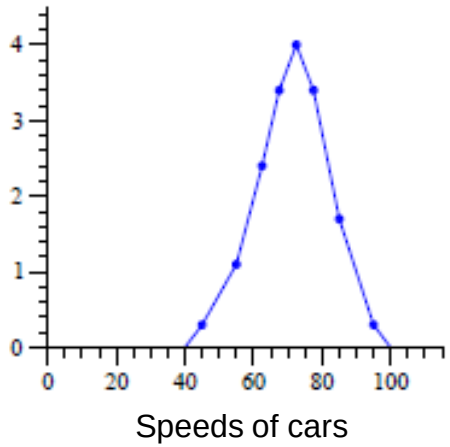
# Citation Profile for 12 Selected Biomed Journals

- Four journals from each category:
  High impact factor >25
  Medium IF 13-17
  Low IF 1-6
- All papers in those 12 journals in a single year: 2007
- All citations of those papers since 2007

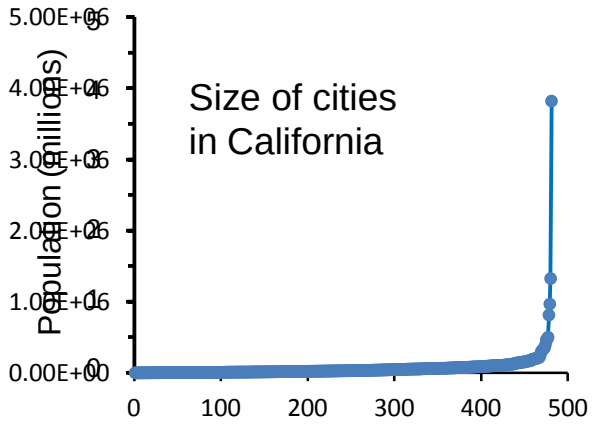


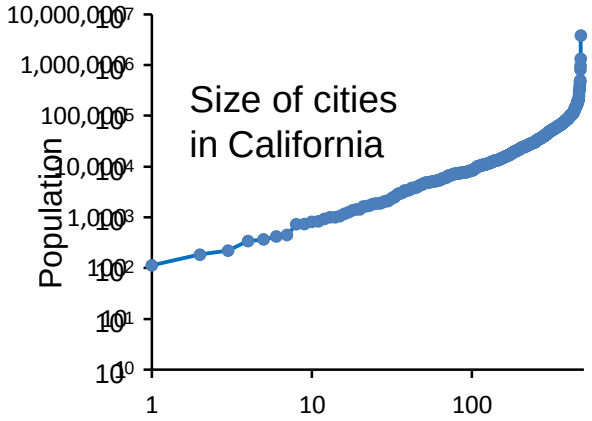
# Averages Apply Only to Gaussian Distributions




Adapted from
MEJ Newman
*Contemp. Physics*
46:322-351 (2005)



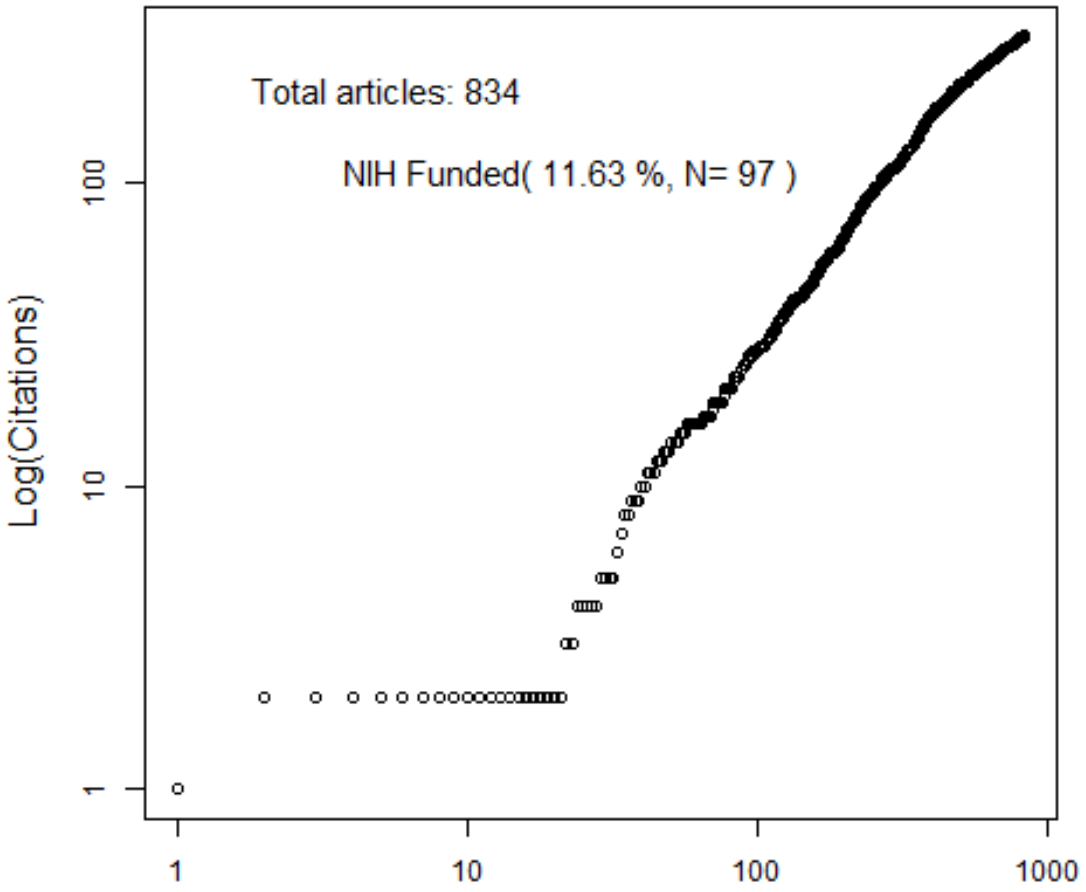




Rank order (smallest to largest)

Power law
$f(x) = Cx^{-\alpha}$
(with a tail)

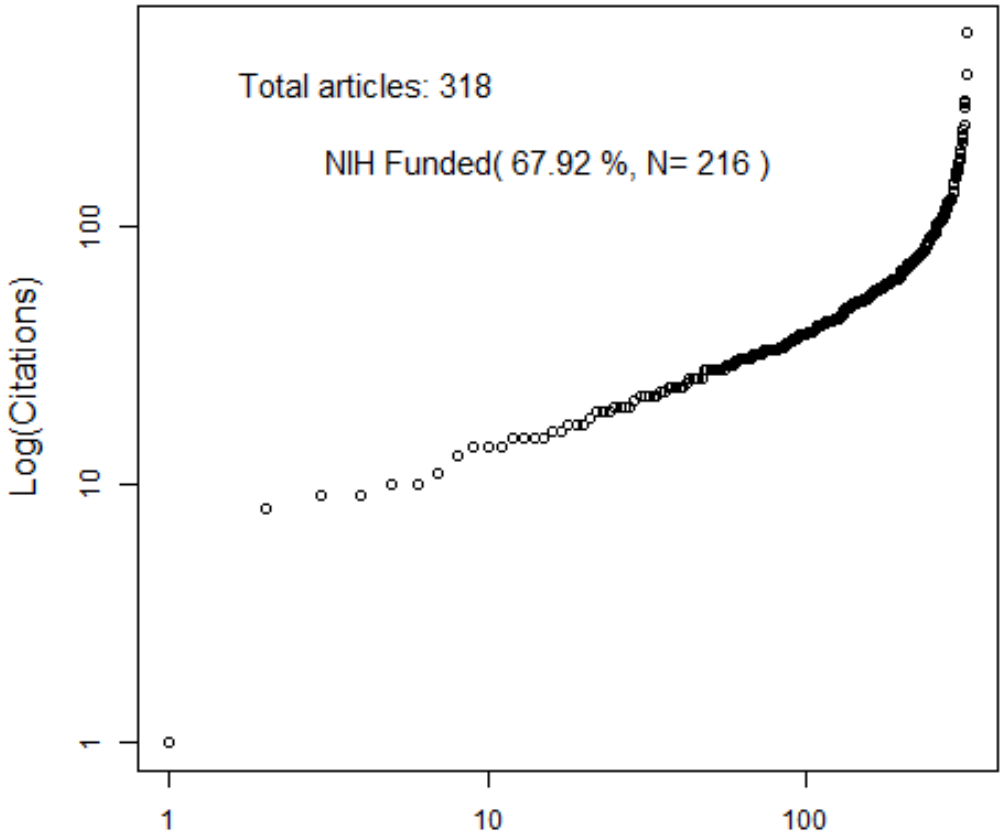
# Power Law Distribution for a High IF Journal

**HIGH IMPACT FACTOR JOURNAL: 2007**



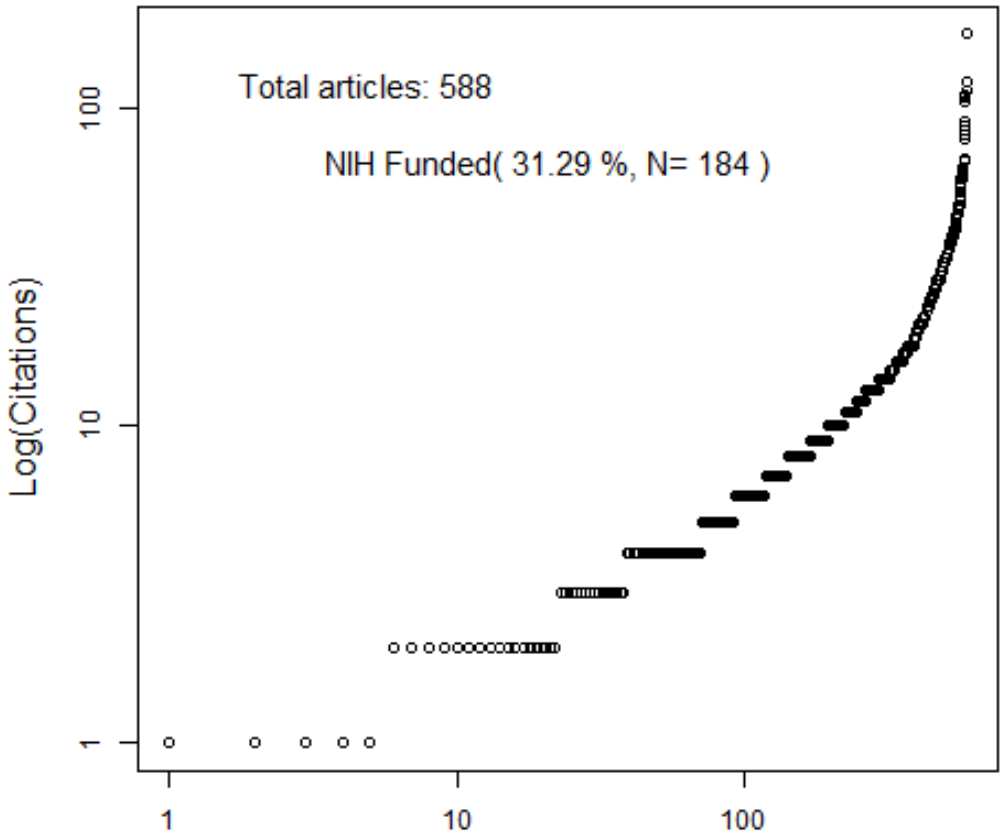
# Power Law Distribution for a Medium IF Journal

**MEDIUM IMPACT FACTOR JOURNAL: 2007**

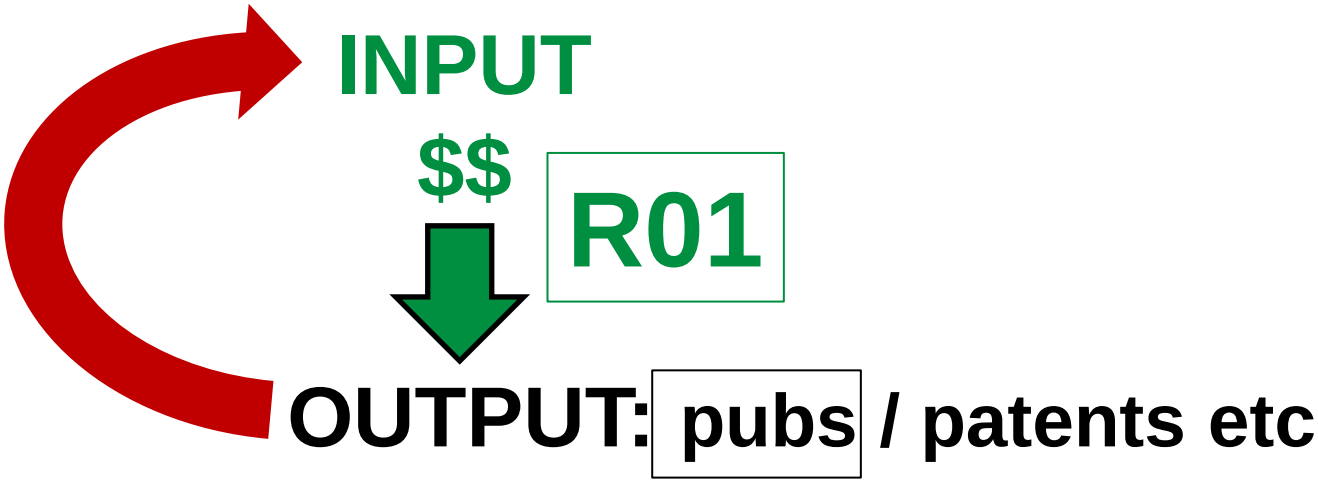


# Power Law Distribution for a Low IF Journal

**LOW IMPACT FACTOR JOURNAL: 2007**



# Modeling NIH output: Advanced Bibliometrics

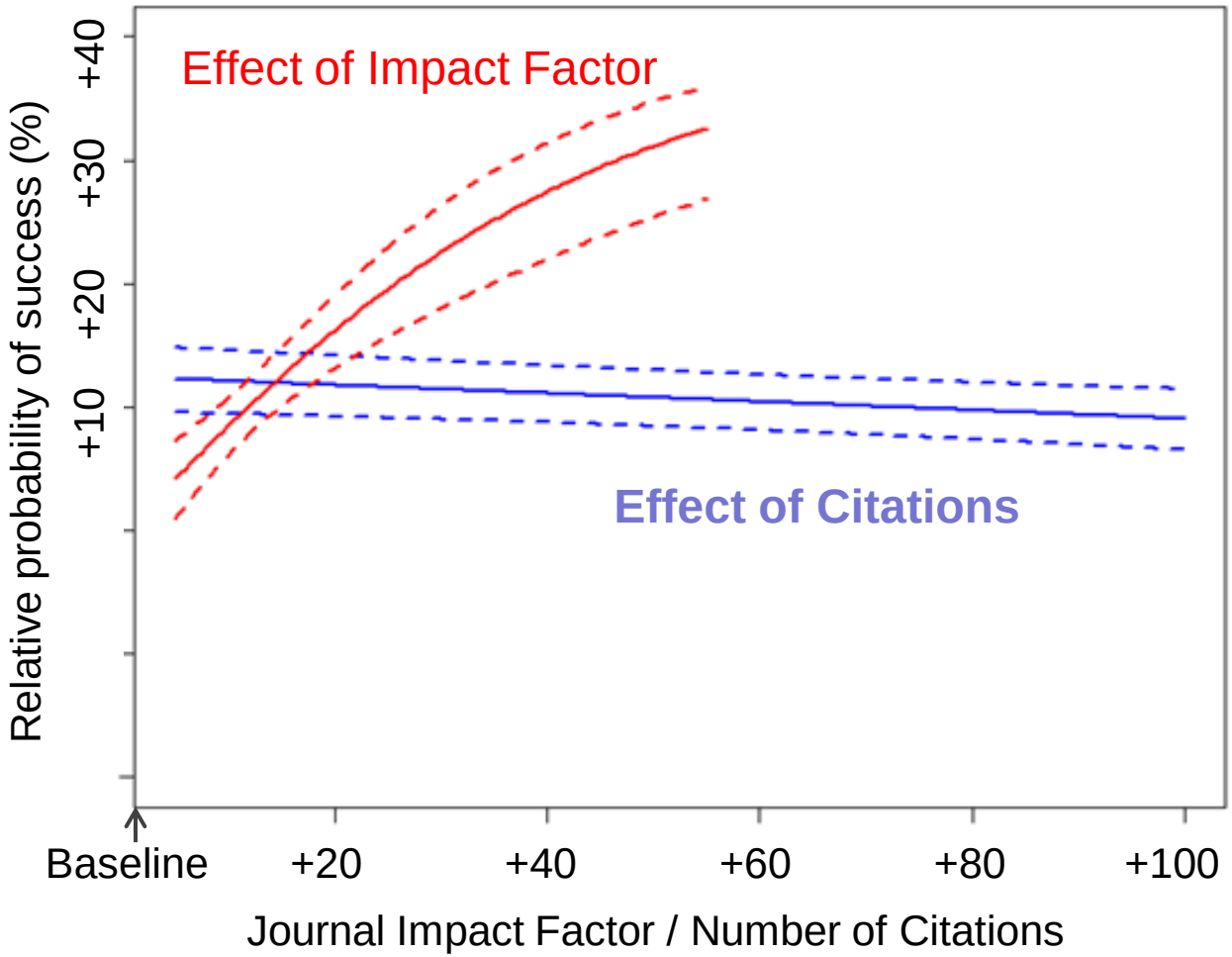



For NIH R01 awardees who published in those 12 journals in 2007:

- Most of the projects have been resubmitted (T2 applications)
- What is the relationship, if any, between prior output and review outcomes?

# Journal IF Affects R01 Fate More Than # of Citations

R01 T2 renewal applicants who published in selected 2007 journals (N = 1631)



# Acknowledgments

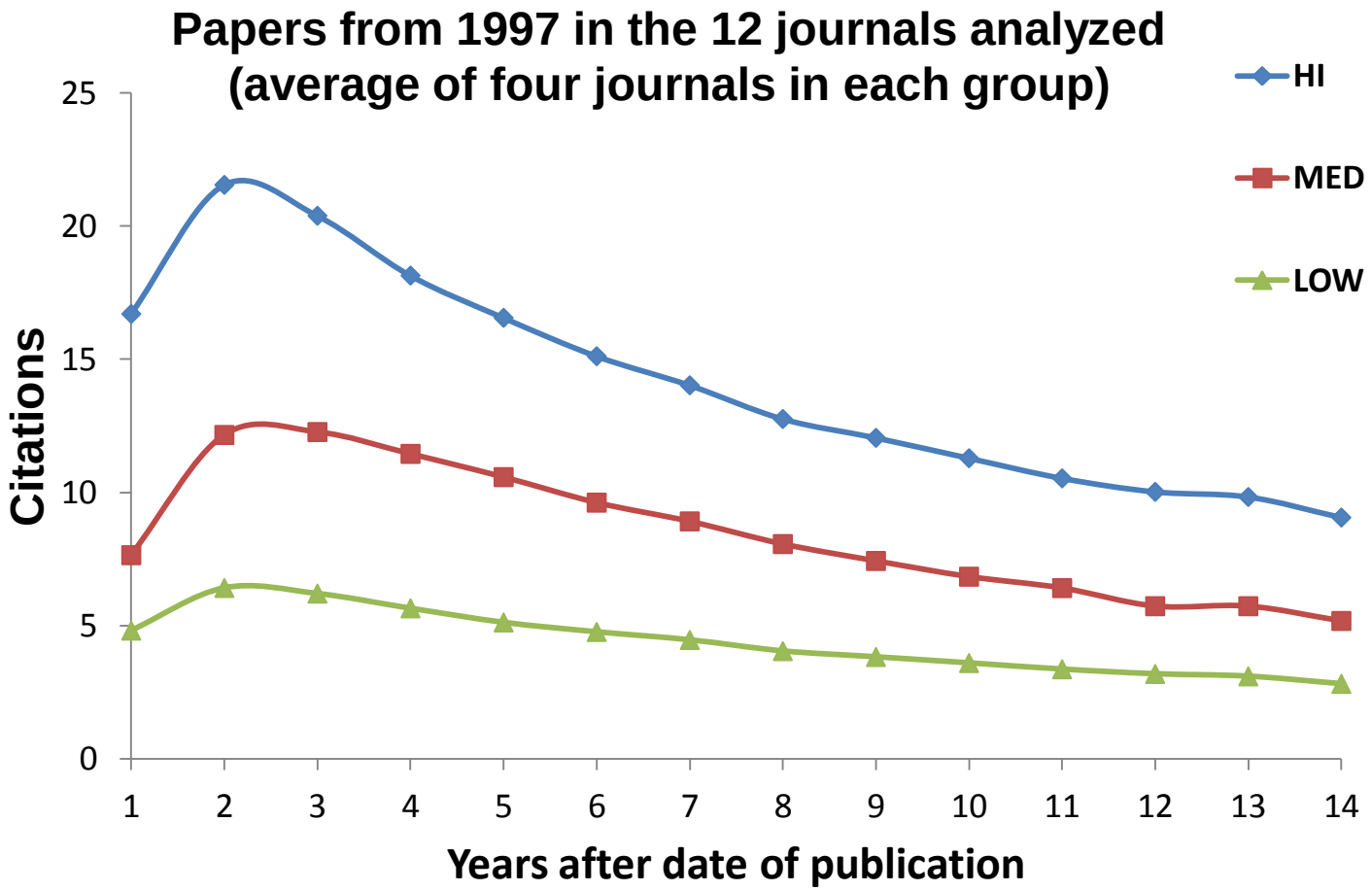

## THE OPA TEAM

| | |
|---|---|
| Geetha Senthil, Ph.D. | Science Analyst |
| Paula Fearon, Ph.D. | Science Analyst |
| Jean Yuan, Ph.D. | Science Analyst |
| Carole Christian, Ph.D. | Health Science Administrator |
| Ehsan Haque, M.S. | Data Analyst |
| Chuck Lynch, Ph.D. | IT Specialist |
| Kevin Small, Ph.D. | Software Developer |

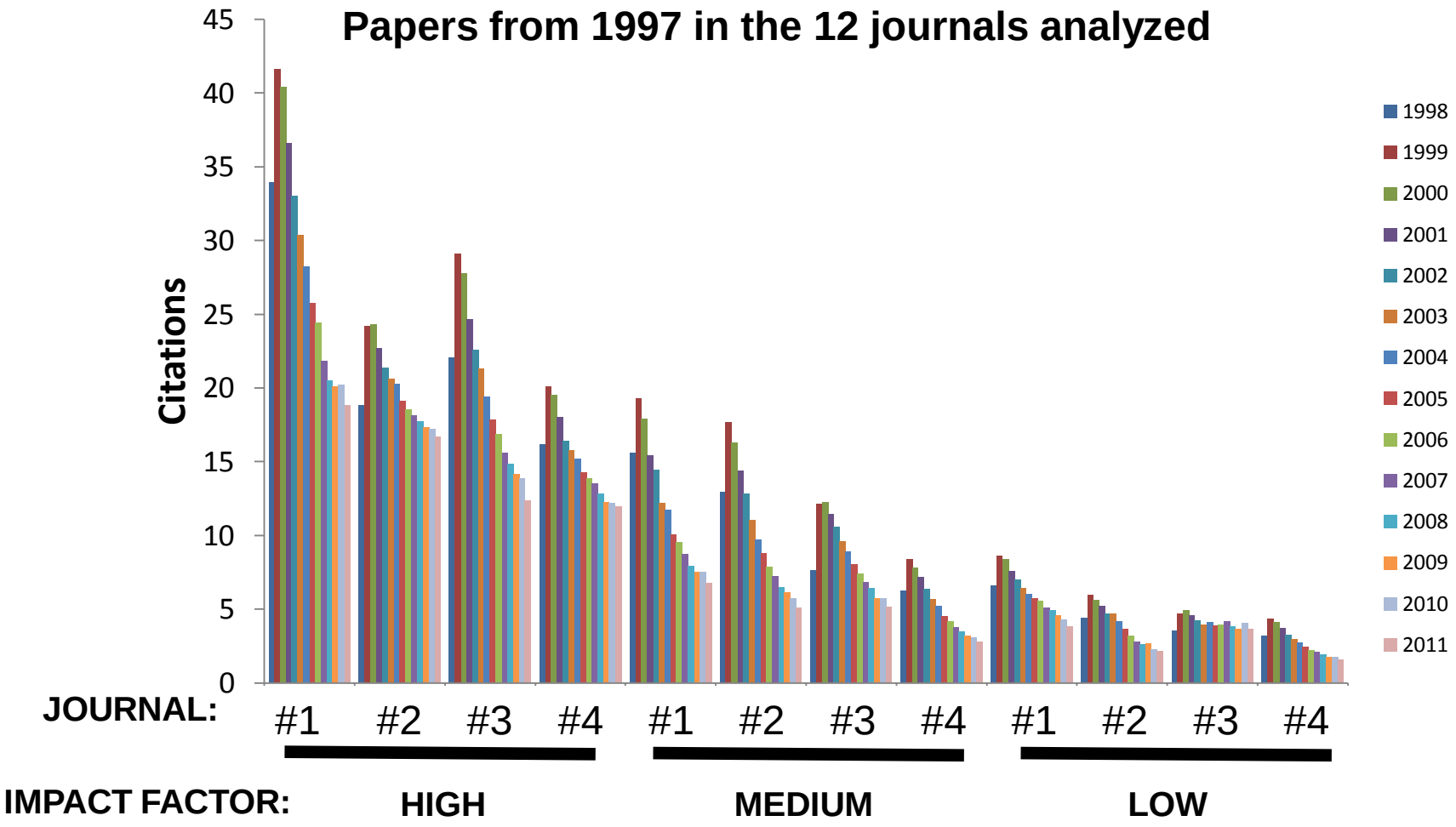
## COLLABORATORS

| | |
|---|---|
| Calvin Johnson, Ph.D. | Chief, High Performance Computing & Informatics, CIT |
| Katy Borner, Ph.D. | Victor H. Yngve Professor of Information Science, Indiana University |

# The Citation Pattern is Established Soon After Publication



# The Citation Pattern is Established Soon After Publication



# The Citation Pattern is Established Soon After Publication