

Experiences using Bibliometric Assessment of NIH Common Fund Programs

Amanda Greene
OD/DPCPSI/OSC
June 6, 2016



National Institutes of Health
Office of Strategic Coordination - The Common Fund



Hagen

Overview

The NIH Common Fund

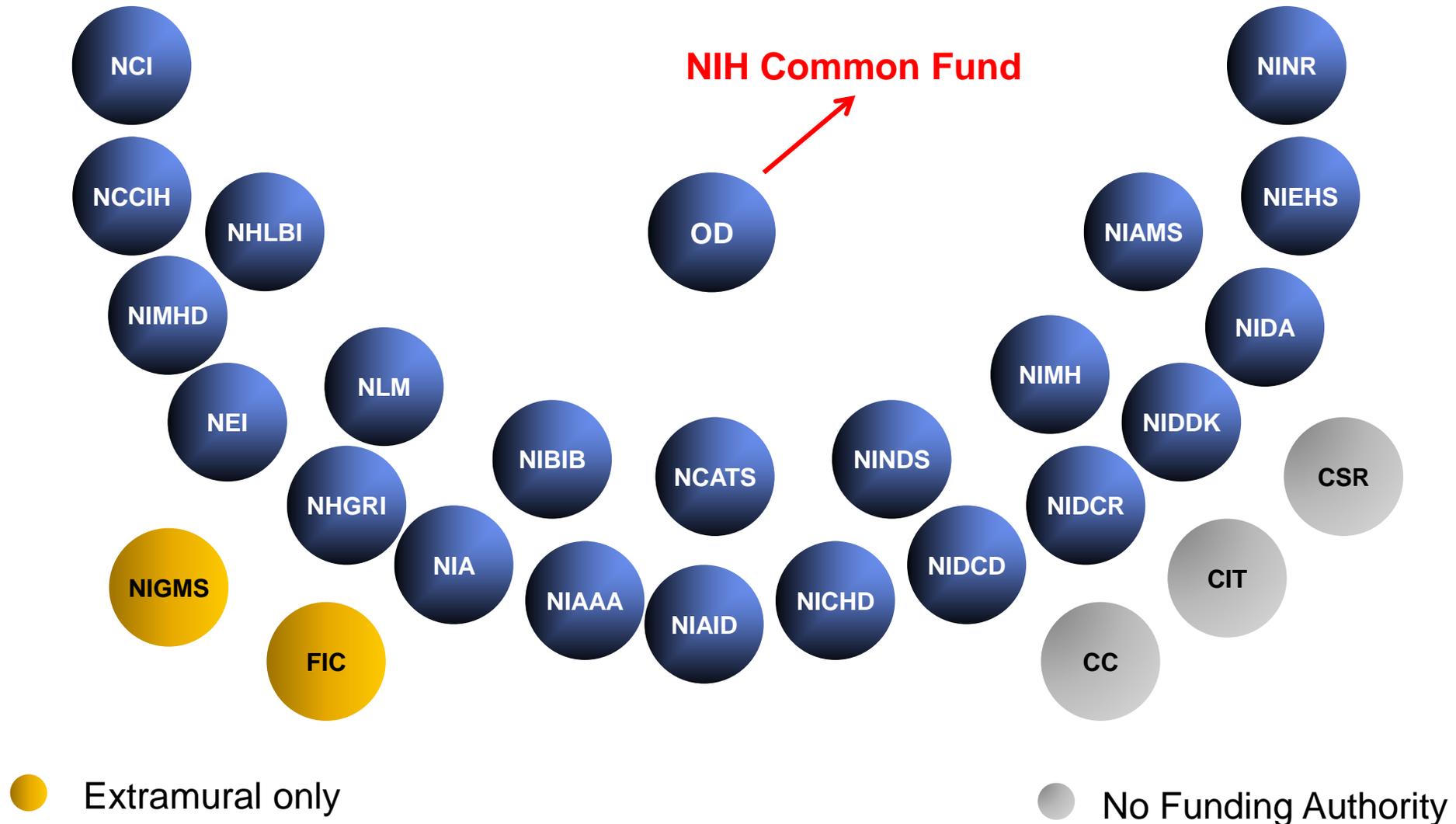
Bibliometrics: one tool in your toolbox for evaluating science

The Common Fund bibliometrics project

Challenges and Limitations

Recommendations

NIH Institutes and Centers



What is the NIH Common Fund?

Supports cross-cutting, trans-NIH programs

Provides a strategic and nimble approach to addressing key roadblocks in biomedical research

Approximately 30 programs spanning the NIH mission

Managed by Office of Strategic Coordination (DPCPSI/OD) in partnership with Institutes and Centers

Criteria for Common Fund Programs

Transformative: Programs are expected to have **exceptionally high and broadly applicable impact.**

Catalytic, Short Term and Goal-driven: Programs must achieve a goal and produce deliverables within **5-10 years.**

Synergistic /Enabling: Programs should be **value-added to the NIH ICs.**

Cross-Cutting: CF programs should address complex issues that require **trans-NIH teams** to design and manage.

Novel: Programs should provide **new solutions to specific challenges.**

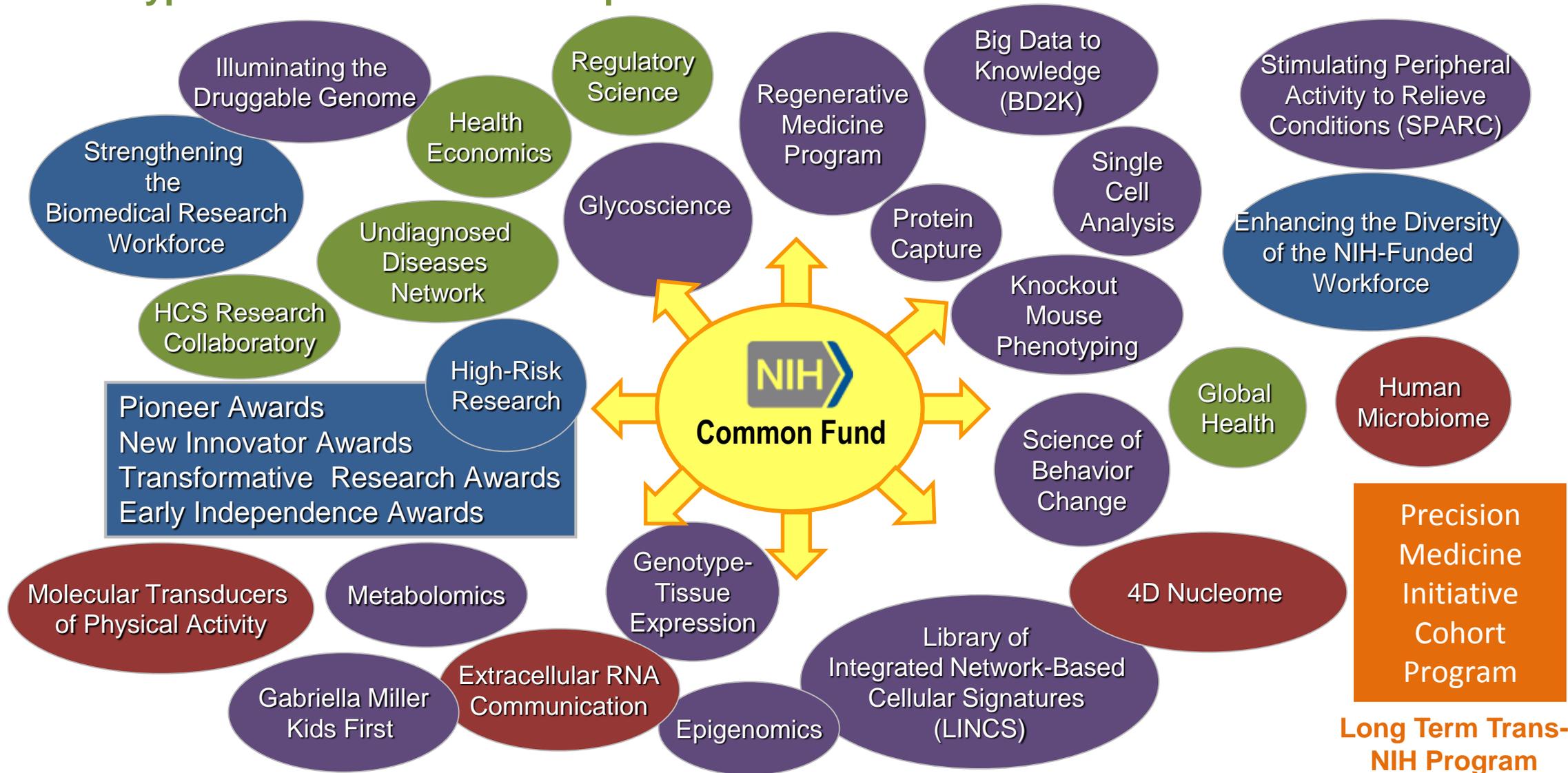


Evaluation of program outputs/outcomes is essential

Current Common Fund Programs (FY16)

New Types of Clinical Partnerships

Data/Tools/Methods



Transformative Workforce Support

New Paradigms

Long Term Trans-NIH Program

After a research paper is published...

How much influence does it have?

How can you measure the article's impact?



BIBLIOMETRICS are quantitative methods for studying science using publications.

REMEMBER:
Bibliometrics are **ONE** tool in your toolbox for evaluating science programs.

Why is the Common Fund conducting bibliometric analyses of its programs?



Determine influence and spread of research through citations

Assess the impact of research publications funded by the Common Fund in the context of other scientific work

Follow trends in science

Common Fund Bibliometric Project

Involved 24 programs

- Each program had to have at least 30 publications

Identified 18,566 publications

Focused on the following metrics

- Number publications
- Number citations
- Relative Citation Ratio (iCite)
- Citation Percentile Rank (Web of Science)

Identifying Publications

Searched QVR using RFA numbers and/or base grant numbers

- Used *History* profile when program started more than 7 years ago

Limited search to:

- Start of RFA funding year through present
- Funded awards

Excluded supplements

Downloaded standard report *Bibliography listing (BIBLIO)*

- Included PubMed Identification number (PMID)

Example: QVR bibliography Listing

PUBMED Link	SCOPUS Link	SPIRES Link	Hist Link	Base Grant Num	PI Name (Contact)	Institution	State	Country	Project Title	Pub Dt	Pub Yr	Pub Month	Pub Title	Pub Author(s)	Pub Journal
PubMed	Scopus	PubSpire	Hist	PN1EY016586	SHEETZ, MICHAEL PATRICK	COLUMBIA UNIV NEW YORK MORNINGSIDE	NY	UNITED STATES	CENTER FOR MECHANOTRANSDUCTION AND NANOMEDI	2007 Feb 9	2007	2	Lamellipodial actin mechanically links myosin activity with adhesion-site formation.	Giannone, Gregory; Dubin-Thaler, Benjamin J; Rossier, Olivier; Cai, Yunfei; Chaga	Cell
PubMed	Scopus	PubSpire	Hist	PN2EY018230	GUO, PEIXUAN	PURDUE UNIVERSITY	IN	UNITED STATES	PHI29 DNA-PACKAGING MOTOR FOR NANOMEDI	2008 Nov	2008	11	Novel mechanism of hexamer ring assembly in protein/RNA interactions revealed by single	Xiao, Feng; Zhang, Hui; Guo, Peixuan	Nucleic acids research
PubMed	Scopus	PubSpire	Hist	PN2EY018230	GUO, PEIXUAN	PURDUE UNIVERSITY	IN	UNITED STATES	PHI29 DNA-PACKAGING MOTOR FOR NANOMEDI	2007 May	2007	5	Viral nanomotors for packaging of dsDNA and dsRNA.	Guo, Peixuan; Lee, Tae Jin	Molecular microbiology
PubMed	Scopus	PubSpire	Hist	PN2EY018230	GUO, PEIXUAN	PURDUE UNIVERSITY	IN	UNITED STATES	PHI29 DNA-PACKAGING MOTOR FOR NANOMEDI	2010 Oct	2010	10	Robust properties of membrane-embedded connector channel of bacterial virus phi29 DNA	Jing, Peng; Haque, Farzin; Vonderheide, Anne P; Montemagno, Carlo; Guo, Xiaob; Feng; Cai, Ying; Wang, Joseph Che-Yen; Green, Dominik; Cheng, R Holland	Molecular bioSystems
PubMed	Scopus	PubSpire	Hist	PN2EY018230	GUO, PEIXUAN	PURDUE UNIVERSITY	IN	UNITED STATES	PHI29 DNA-PACKAGING MOTOR FOR NANOMEDI	2009 Aug 25	2009	8	Adjustable ellipsoid nanoparticles assembled from re-engineered connectors of the	Guo, Peixuan; Lee, Tae Jin; Wang, Joseph Che-Yen; Green, Dominik; Cheng, R Holland	ACS nano
PubMed		PubSpire	Hist	PN2EY018244	ISACOFF, FLUID	UNIVERSITY OF CALIFORNIA	CA	UNITED STATES	NDC FOR THE OPTICAL CONTROL	2013 Oct	2013	10	The Brain Prize 2013: the anticipation	Reiner, Andreas; Isacoff, Fluid Y	Trends in neuroscience

Cleaning Identified Publication Lists

Machine curation

- Removed any publications that occurred prior to the start of the program
- Removed duplicates based on PMID
- Limited timeframe from start of program through 2014

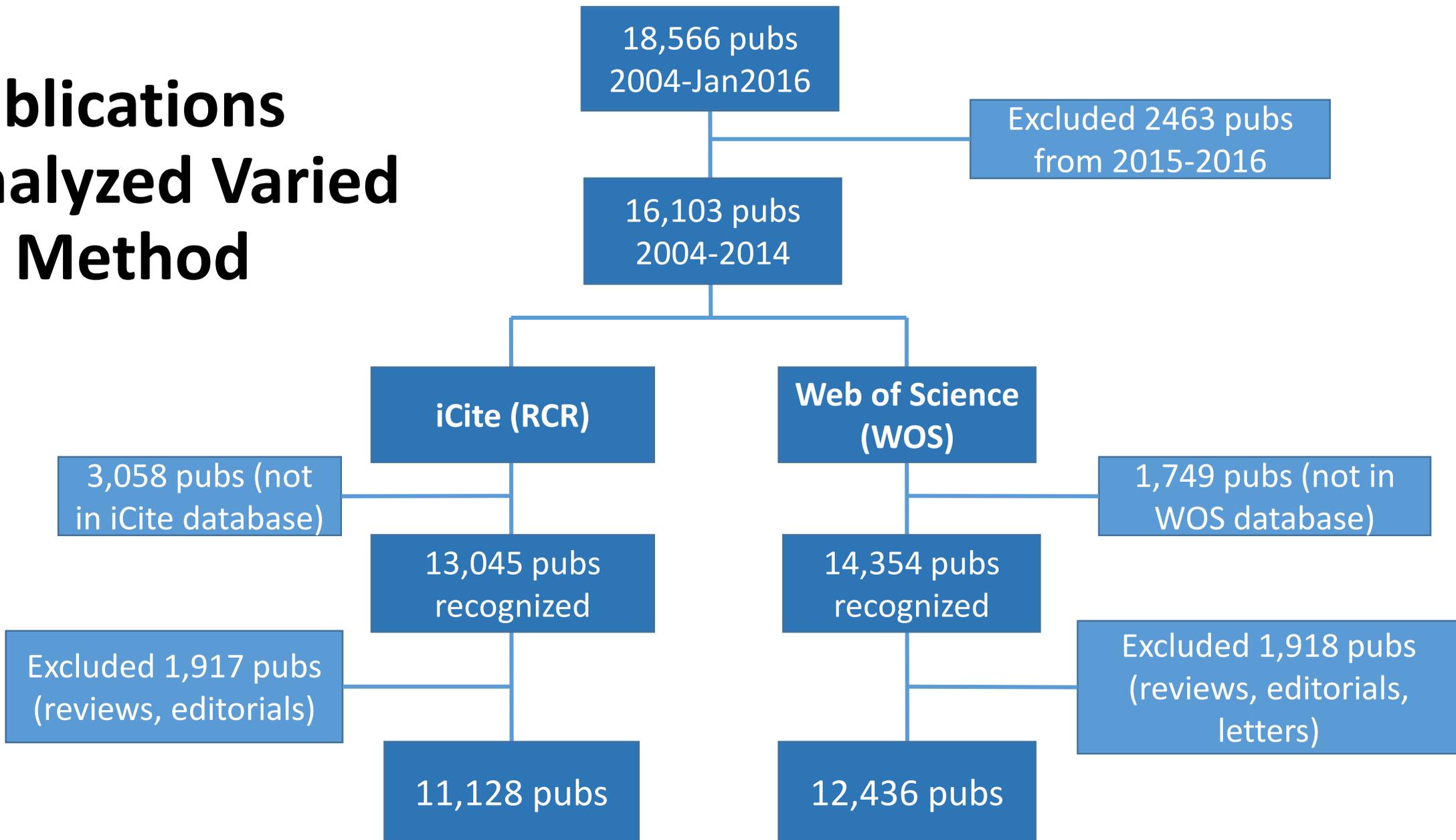
Tips

- Keep raw data file with date stamp
- Create working files
- Determine if list needs to be manually curated
- Keep log of all actions/changes to working file
- In Excel, use Pivot Tables

Identifying Publications is an Iterative Process

- Topic and/or Question
- Refine topic – identify search terms OR RFA/PA and/or grant numbers
- Determine where you will search? (e.g., NIH IMPAC II, PubMed, Web of Science, Scopus)
- Consider other resources such as Annual Reviews in your topic area or Cochrane Library
- Refine you list of publications (e.g., delete duplicates, remove out of range publications)
- Determine if manual curation is needed and how this will be done
- Keep a log of all steps and decisions

Publications Analyzed Varied by Method

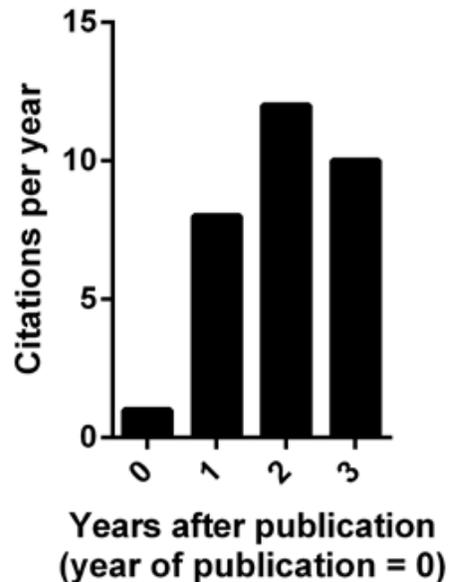


Calculating Relative Citation Ratio (RCR) using iCite

$$RCR = \frac{\textit{Article Citation Rate}}{\textit{Expected Citation Rate}}$$

Article Citation Rate

(denominator excludes year of publication)



Source: Santangelo, NIH OPA (2015)

- Article-level metric
- Number of cites per year
- RCR changes over time with the accrual of new citations
- Scalable to large portfolios containing tens of thousands of articles

Hutchins, Yuan, Anderson, & Santangelo (2015) Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. BioRxiv.

<http://biorxiv.org/content/early/2015/10/22/029629>

RCR/iCite Results for Common Fund Programs

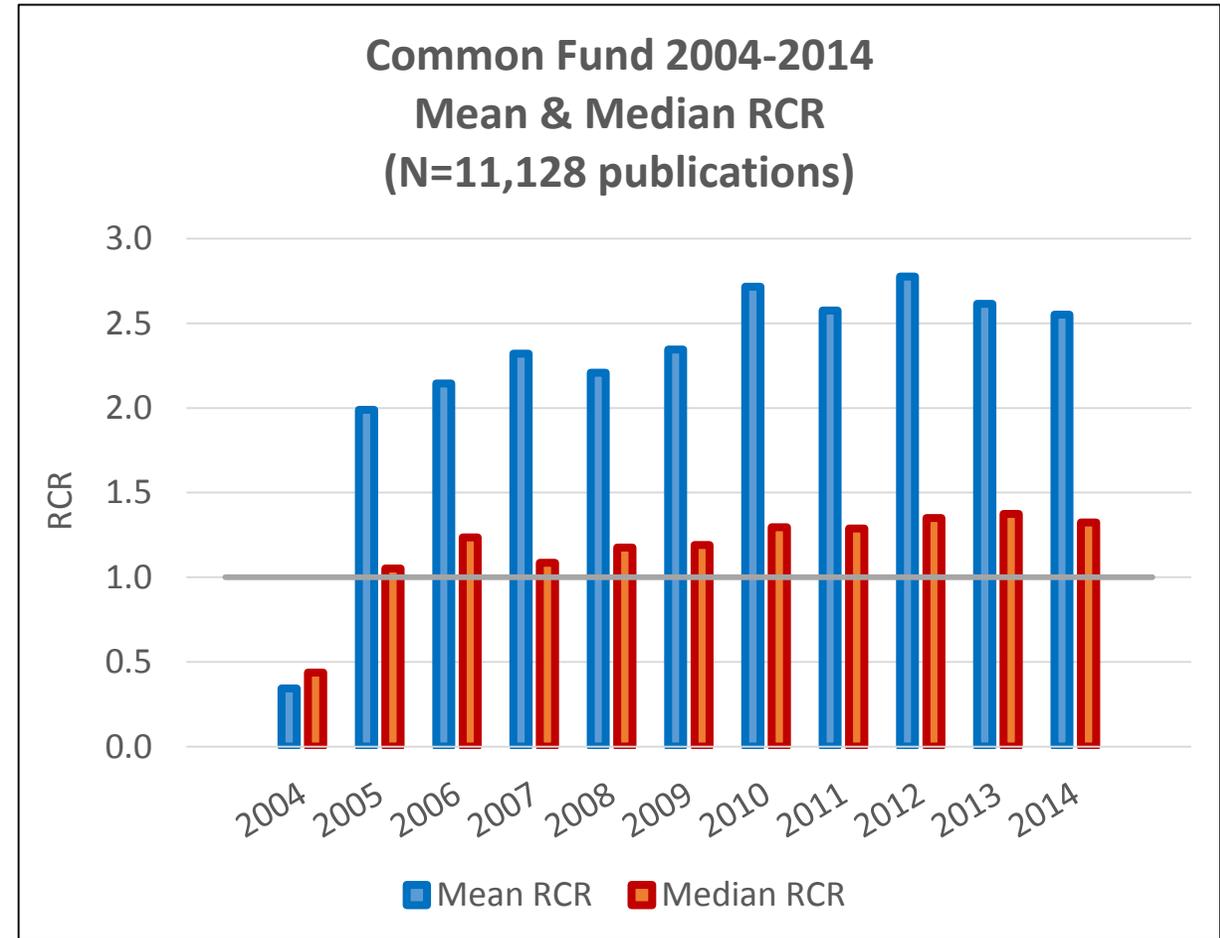
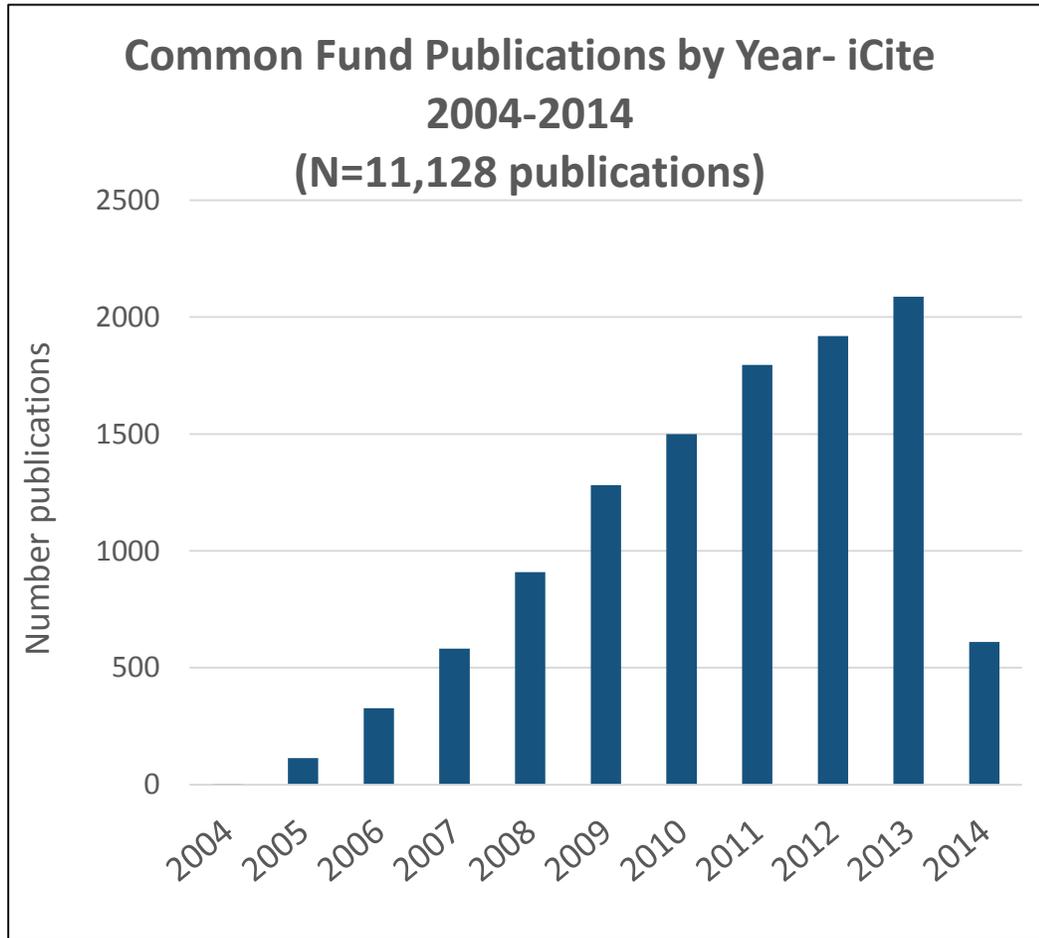
All Publications

Publications	13,045
Publications with no citations	320 (2.4%)
Citations/year	13.2 mean 6.0 median
RCR	2.6 mean 1.3 median

Publications excluding Reviews

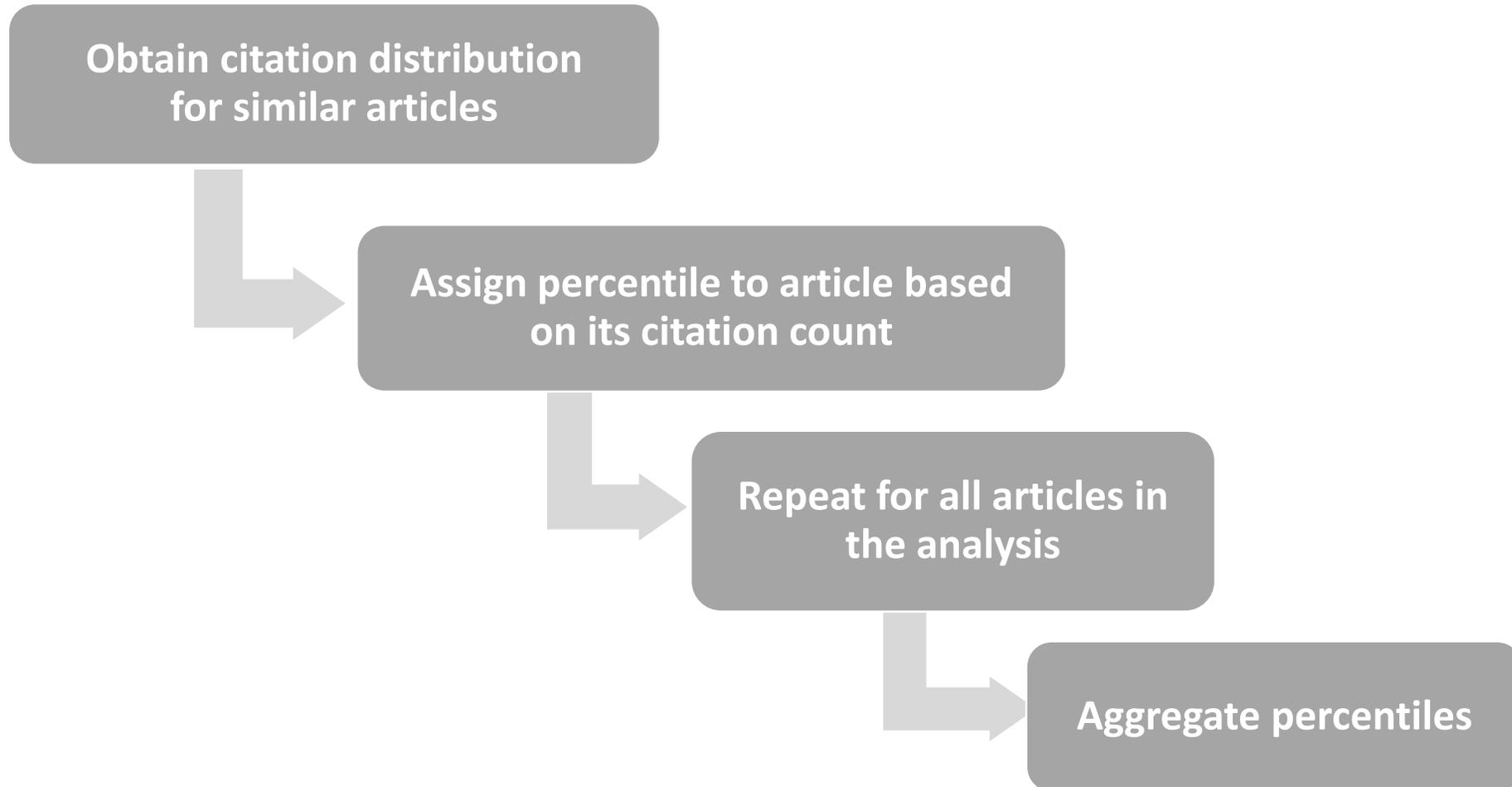
Publications	11,128
Publications with no citations	245 (2.2%)
Citations/year	13.0 mean 5.8 median
RCR	2.5 mean 1.3 median

RCR for Common Fund Programs*

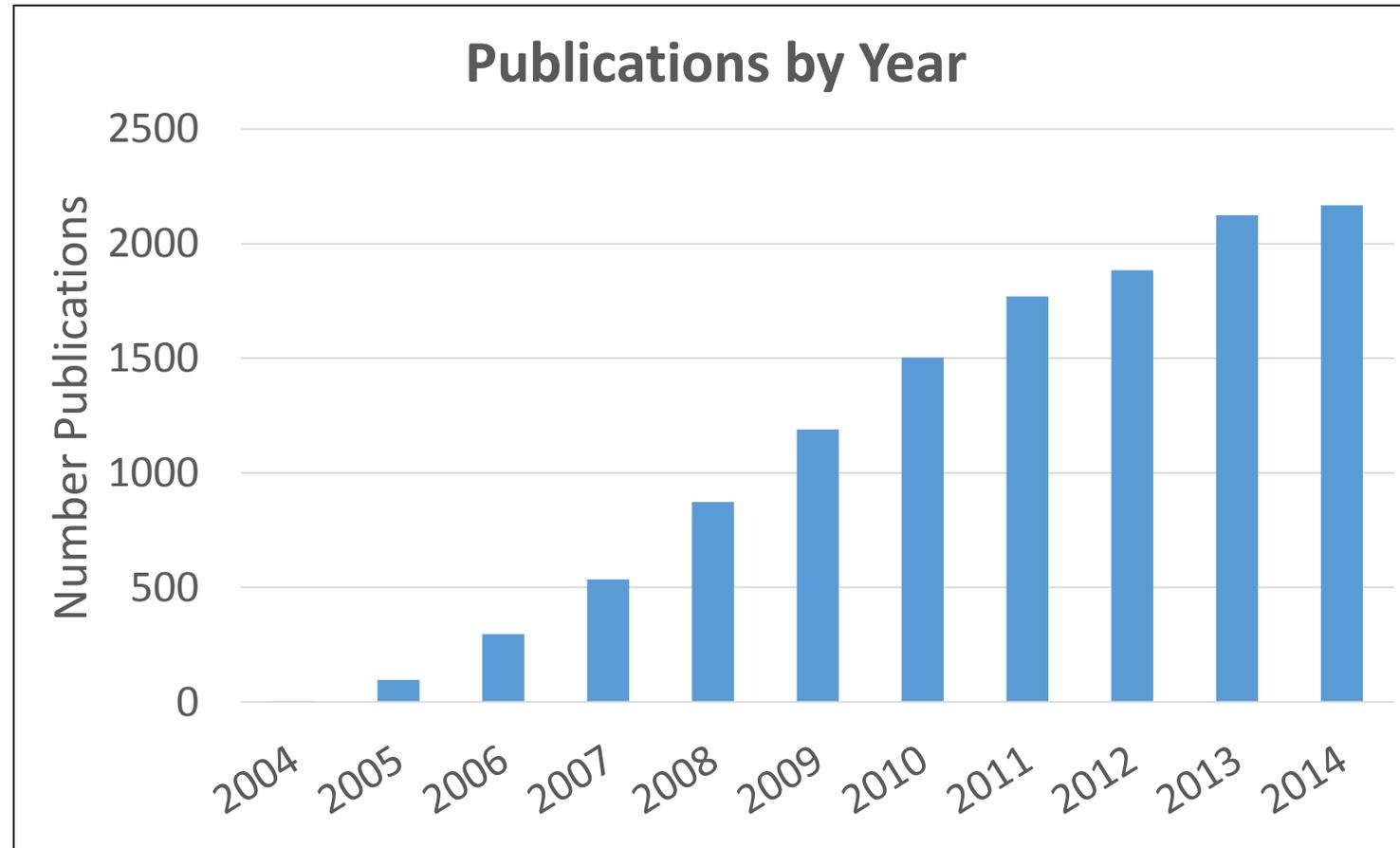


* Excludes reviews and editorials.

Citation Percentile Ranking using Web of Science Essential Science Indicators



Common Fund Publications identified by Web of Science (N=12,440*)

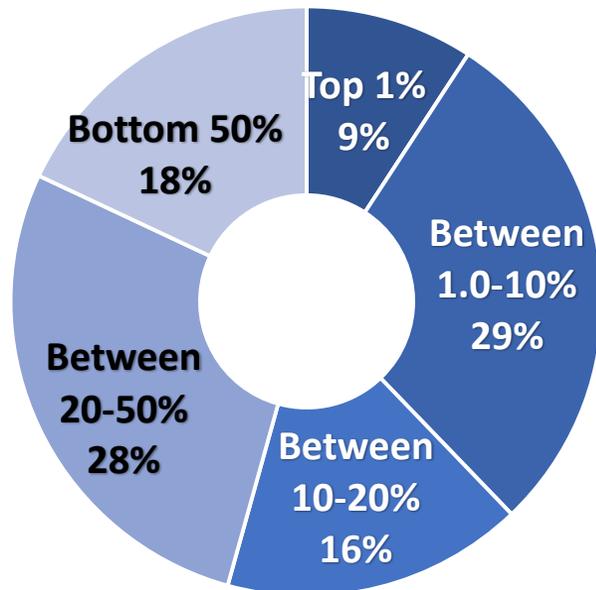


* Excludes reviews, editorials, letters

Citation Percentile Rank for Common Fund Programs (N=12,436 publications)

Percentile Ranking by Essential Science Indicators Categories

% Articles by Citation Percentile Rank



* excludes Reviews, Letters, Proceedings

Number & Percent of Publications by Citation Percentile Rank by Essential Science Indicators Categories

	Publications	
	Number	Percent
Top 1%	1143	9%
Between 1.0-10%	3561	29%
Between 10-20%	2049	16%
Between 20-50%	3445	28%
Bottom 50%	2238	18%

* excludes Reviews, Letters, Proceedings

Reporting Results

Tailor the report to the audience(s)

Statement of why conducting a bibliometric assessment

- What questions are you trying to answer?

1-2 page executive brief

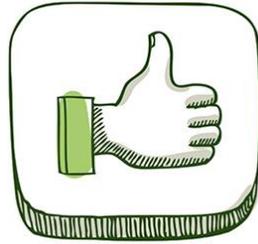
Stand-alone tables, graphs, and figures

Definitions of bibliometrics used

Methods so analysis is replicable

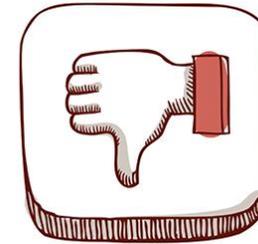
Conclusions/Limitations/Recommendations

Bibliometrics: Pros and Cons



PROS

- Handles large data sets
- Produces reproducible results
- Influence according to a large sample



CONS

- Requires expertise to generate and interpret
- Only measures publications
- Citation counts may be misleading
- Does not measure quality
- Limited measure of impact

Citing a Work \neq Agreement with Findings

Citation Patterns Differ Between Subjects

Journal Quality \neq Article Quality

Self Citations



Challenges

- New analytic approach for staff – learning curve
- Identifying publications
- Used beta-version of iCite
- Not all publications recognized by iCite or Web of Science
- Machine versus manual curation



Recommendations

Compare like with like

- Similar research areas
- Similar journals (discipline)
- Stage of academic career
- Similar size institutions



Don't use just one tool

- Coverage varies in content, depth, discipline

Data need to be looked at in context

Use a variety of metrics and other qualitative information where appropriate

The Leiden Manifesto presents ten principles for best practice in metrics-based research assessment

COMMENT

SUSTAINABILITY Data needed to drive UN development goals **#432**

CONSERVATION Economics and environmental catastrophe **#434**

GEOLOGY Questions raised over proposed Anthropocene dates **#436**

HISTORY Music inspired Newton to add more colours to the rainbow **#438**



The Leiden Manifesto for research metrics

Use these ten principles to guide research evaluation, urge **Diana Hicks, Paul Wouters** and colleagues.

Data are increasingly used to govern science. Research evaluations that were once bespoke and performed by peers are now routine and reliant on metrics¹. The problem is that evaluation is now led by the data rather than by judgement. Metrics have proliferated: usually well intentioned, not always well informed, often ill applied. We risk damaging the system with the very tools designed to improve it, as evaluation is increasingly implemented by organizations without knowledge of, or advice on, good practice and interpretation. Before 2000, there was the Science Citation Index on CD-ROM from the Institute for Scientific Information (ISI), used by experts for specialist analyses. In 2002, Thomson Reuters launched an integrated web platform, making the Web of Science database widely accessible. Competing citation indices were created: Elsevier's Scopus (released in 2004) and Google Scholar (beta version released in 2004). Web-based tools to easily compare institutional research productivity and impact were introduced, such as InCites (using the Web of Science) and SciVal (using Scopus), as well as software to analyse individual citation profiles using Google Scholar (Publish or Perish, released in 2007). In 2005, Jorge Hirsch, a physicist at the University of California, San Diego, proposed the *h*-index, popularizing citation counting for individual researchers. Interest in the journal impact factor grew steadily after 1995 (see 'Impact factor obsession'). Lately, metrics related to social usage

ILLUSTRATION BY CLAUDE BAYLON

23 APRIL 2015 | VOL 520 | NATURE | 429
© 2015 Macmillan Publishers Limited. All rights reserved

Hicks, Wouters, Waltman, de Rijcke, & Rafols (2015) [Bibliometrics: The Leiden Manifesto for research metrics](#). *Nature* 520, 429–431 (23 April 2015) doi:10.1038/520429a.

Evaluative Challenges

Can new bibliometric methods be developed to:

Measure innovation?

Estimate long term impact from short term data?

Identify paradigm shifts?

Acknowledgements

Office of Strategic Coordination/ Common Fund

Marishka Brown

Rachel Britt

Stephanie Courchesne-Schlink

Kristi Faulk

Scott Jackson

Rebecca Lenzi

Aron Marquitz

Becky Miller

Jessica Smith

Joe Tiano

OER, Office of Data Analysis Tools and Systems (DATA Systems)

Brian Haugen

NIH Office of Portfolio Analysis

Ian Hutchins

NIH Library

Chris Belter

Ya-Ling Lu

Questions

For more information, contact:

Amanda Greene

NIH Common Fund / Office of Strategic Coordination

amanda.greene@nih.gov

URL: <http://commonfund.nih.gov/>