



# Making it count

Research credit management in a collaborative world

Jonathan Adams, Ludo Waltman and  
Ross Potter

April 2022



# Background

ISI and scientometric indicators

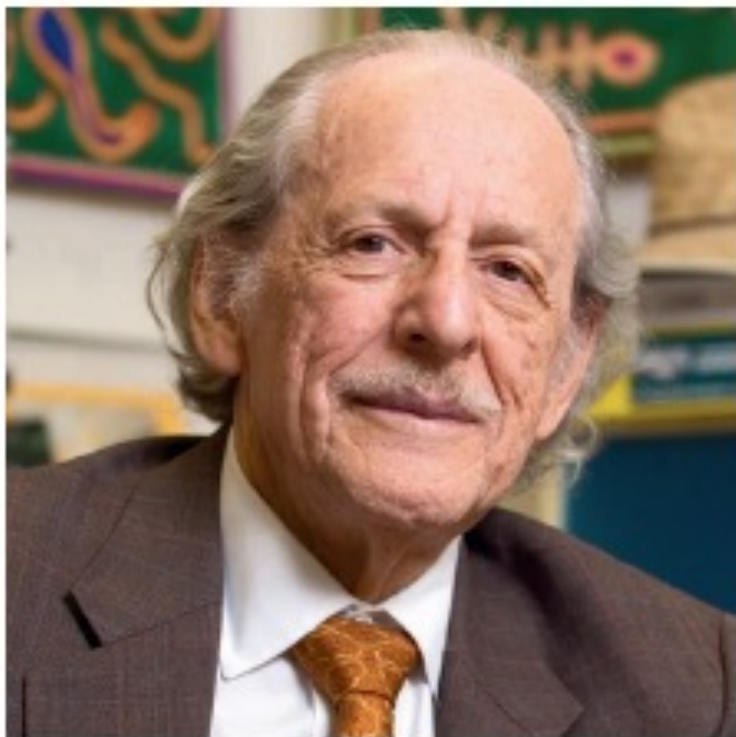
- Profiles, not metrics (2019)

- We don't want more indicators
- We want better versions of the indicators we already have
- We want to enable better presentation of indicators
- We want more informed interpretation of indicators
- We want feedback on ideas from users – not from bibliometric experts
- We want to know what works, what is useful

# The genesis of citation indexing and analysis

## Eugene Garfield

1925 – 2017



### Information retrieval:

Eugene Garfield, "Citation indexes for science: a new dimension in documentation through association of ideas," *Science*, 122 (3159): 108-11, 15 July **1955**

### History and sociology of science:

Eugene Garfield, "Citation indexes in sociological and historical research," *American Documentation*, 14 (4): 289-291, **1963**

Eugene Garfield, "Citation indexing for studying science," *Nature*, 227 (5259): 669-671, **1970**

### Structure and dynamics of science:

Eugene Garfield, M.V. Malin, and H. Small, "A system for automatic classification of scientific literature," *Journal of the Indian Institute of Science*, 57 (2): 61-74, **1975**

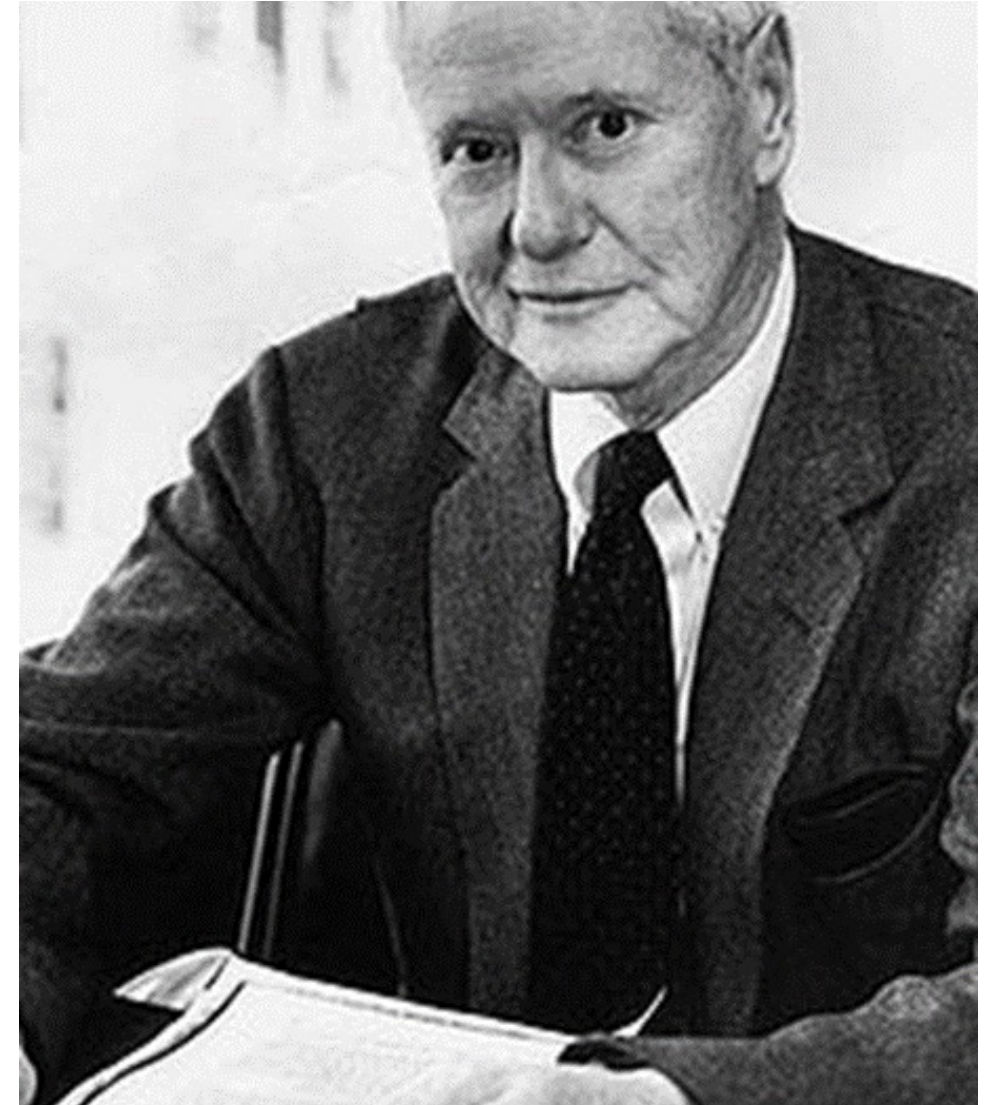
### Evaluation of research:

Eugene Garfield, "Is citation analysis a legitimate evaluation tool?," *Scientometrics*, 1 (4): 359-375, **1979**

## Theories of citation and the normative school

- Merton was a sociologist of science at Columbia University. Normative theory.
- ***Citations as currency used to repay intellectual debts. Those with many citations have gained “credits” from their peers.***
- The formal nature of publication and the moral imperative to cite.
- Other theories, including citations as rhetorical devices, constructivist theories.
- Known for coining the concepts and phrases: “self-fulfilling prophecy,” “role model,” “focus group,” “unanticipated consequences,” and “Matthew effect.”

**Robert K. Merton (1910-2003),**



# Citation distributions are skewed

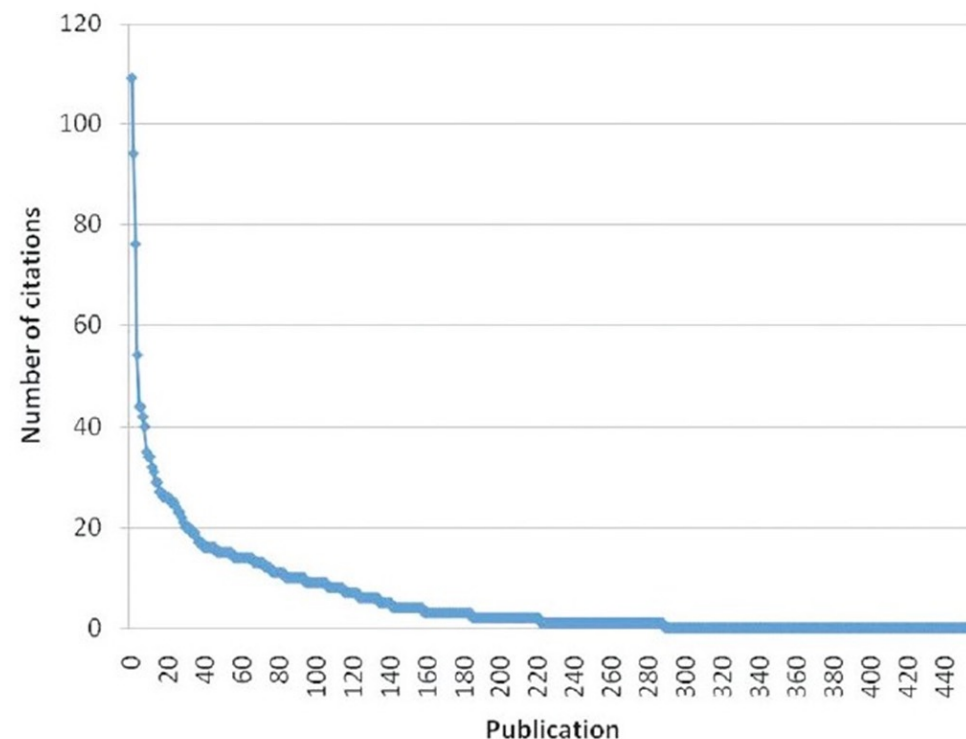
- Skewed by papers and by authors
- Created in a socio-cognitive context, in which contributions of value are recognized and those recognized then tend to attract more citations
- Cumulative advantage, preferential attachment, success breeds success, and Merton's Matthew effect

## Highly-cited chemists

First Name	Last Name	Citations
George M.	Sheldrick	52551
Michael	Graetzel	37409
Rodney S.	Ruoff	35159
Omar M.	Yaghi	24221
SonBinh T.	Nguyen	23119
Donald G.	Truhlar	22600
Stefan	Grimme	20082
Mohammad Khaja	Nazeeruddin	18578
Jeffrey R.	Long	18327
Richard D.	Piner	16480
Avelino	Corma	16240
Yan	Zhao	16080
Hong-Cai	Zhou	15730
Sasha	Stankovich	15580
Michael	O'Keeffe	15570
Hui-Ming	Cheng	15450
Hongjie	Dai	15380
Dmitriy A.	Dikin	15140
Joseph T.	Hupp	14800
Jin-Quan	Yu	14080
Prashant V.	Kamat	13920
Younan	Xia	13650
Jiaguo	Yu	13220
Jean-Marie	Tarascon	12680
Gao Qing (Max)	Lu	12570

The distribution of citations across people and papers follows a power law model (various models exist)

This is like Pareto's description of income: 20% of papers attract 80% of citations



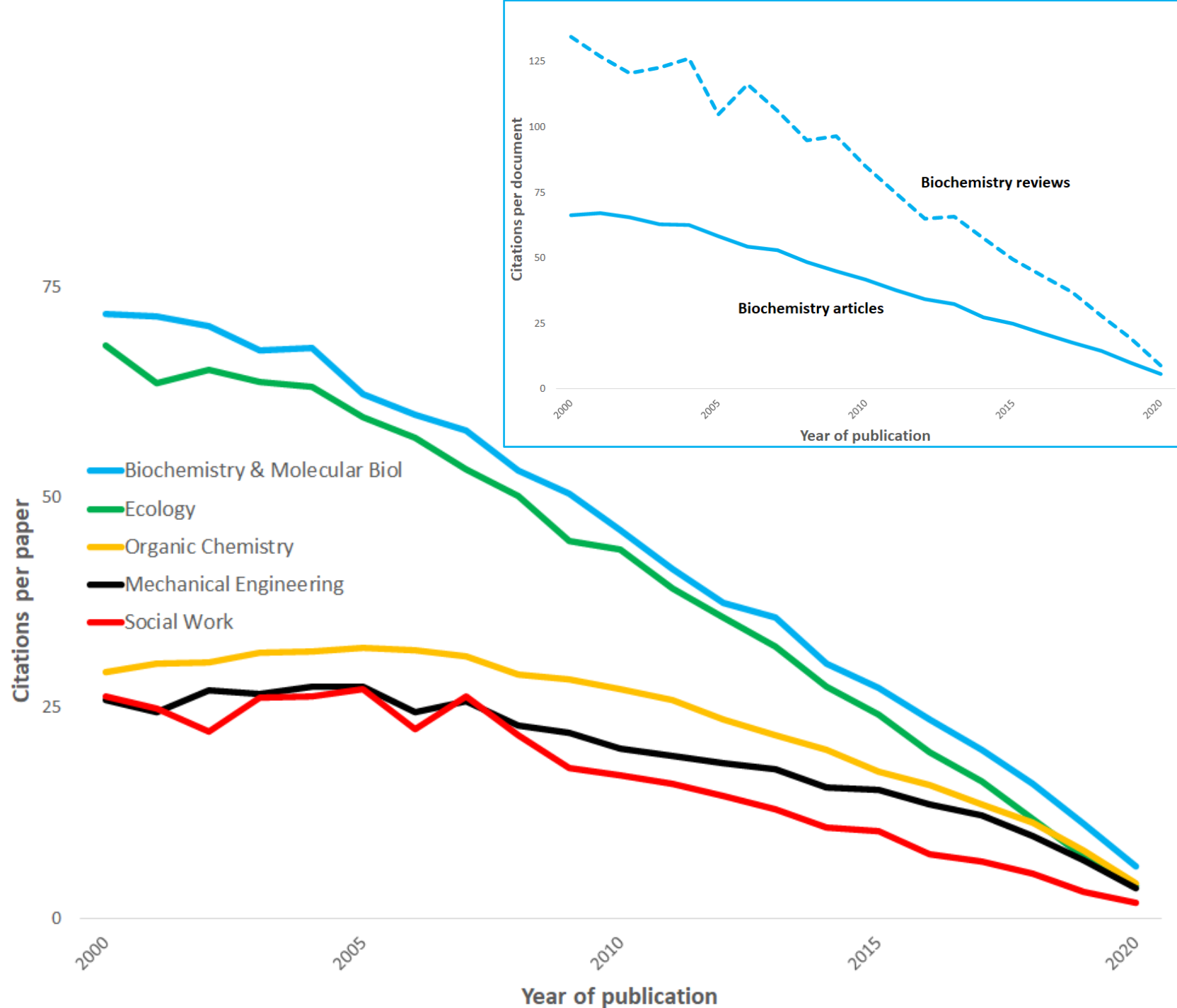
# Citation impact

- So, we have a database of publications in dated (when?) journal volumes
- Every publication has an author (who?) at an address (where?)
- The titles (what?) publications are in a citation network

- Research is supported by public money
- Governments want research funders to demonstrate value and effective resource distribution
- Research investment buys research production = papers
- Are some publications more significant than others?
- What information do the variable numbers of citations provide?
- Can the data be used in standardized analyses?
- What other metadata are needed to make these analyses valid comparisons (comparing like-with-like)?
- Can these analyses be used to inform public policy, guide public funding and encourage better research management?

# Citation indicators

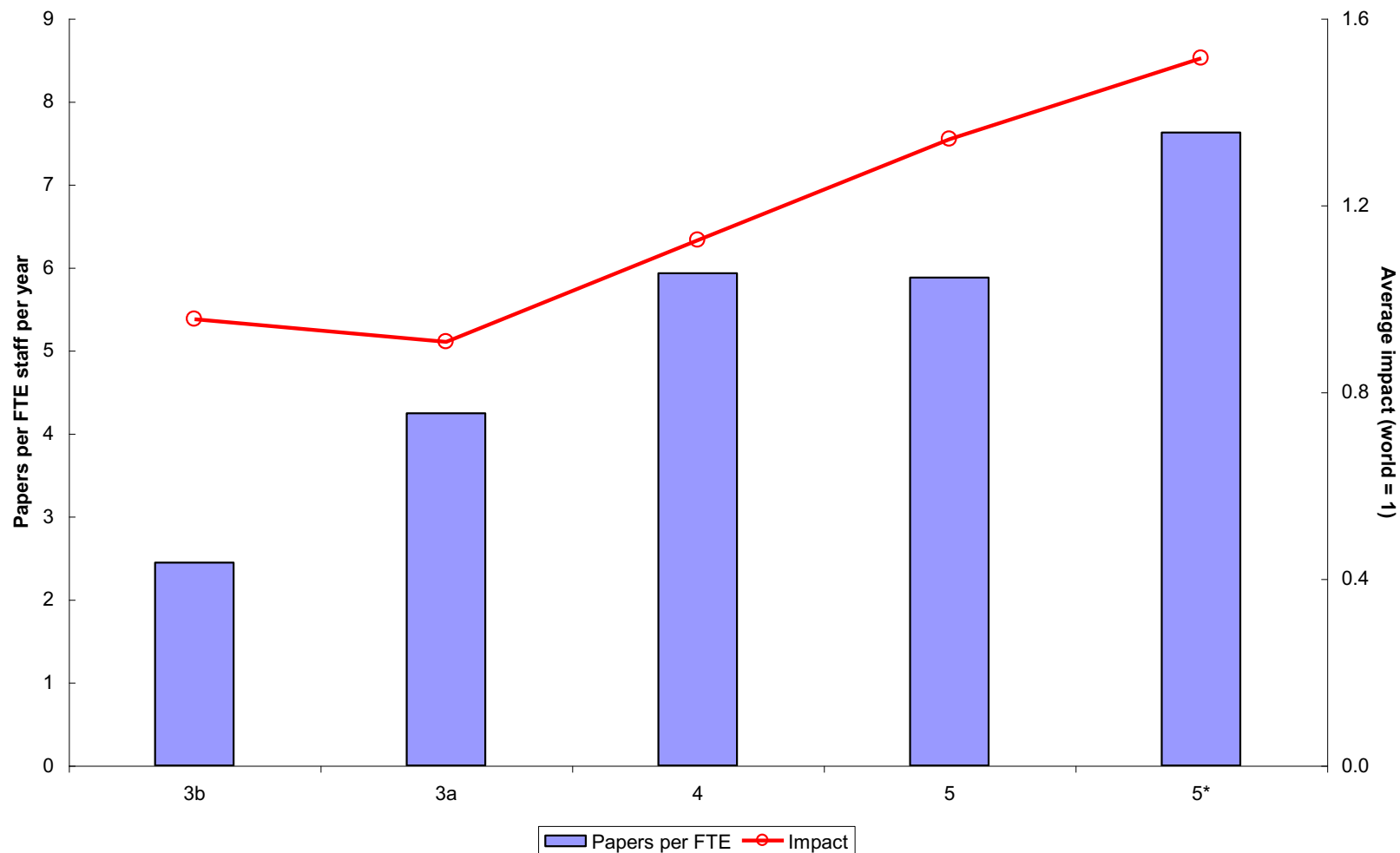
- Garfield recognised journals grouped into categories/fields
- Citation counts to any paper (article or review) rise over time at a rate that depends on field and document-type
- We index for
  - Document type
  - Year of publication
  - Journal category
- This is Category Normalized Citation Impact (CNCI)



## Citation impact correlates with peer review

- These data for Chemistry are:
  - RAE2001 peer panel grades
  - CNCI for university output 1996-2000
- There is considerable residual variance but the trend is found across all subjects

The positive relationship between peer grade and average citation impact was first found in RAE1996 (Adams, J. 1998. Nature, 396, 615-618.)



These data are for university Chemistry departments grouped by grade (1-5\*)



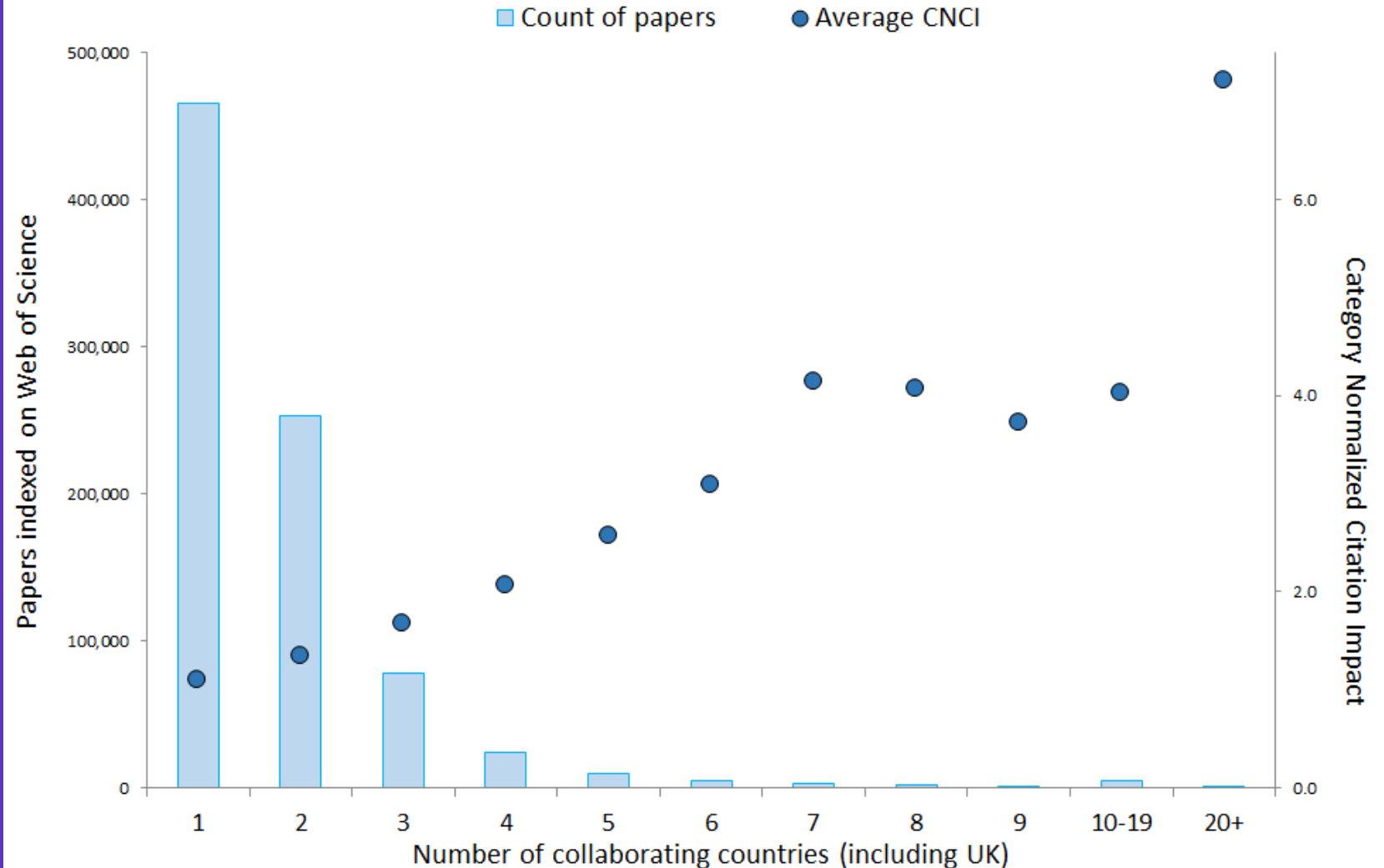
# What's the problem?

- Research collaboration affects citation impact
- Users of our data and products are not always clear about how this works
- They are also unclear about the best way to manage 'credit' for multi-author work

- Research collaboration is increasing and more collaborative authorship is linked to higher citation counts
- Managers want to understand how this affects performance indicators
- Standard CNCI uses whole counting
- Many analysts prefer Fractional counts but this uses arbitrary weighting and offers only spurious 'correction'
- ISI has developed Collab-CNCI, which keeps standard normalization of citation impact but does so within collaborative types
- Collab-CNCI can be used to provide both
  - a single summary impact metric
  - the distribution of that impact across output types
- This provides more complete data for research management and policy-making

# The problem: collaboration and credit

- Research collaboration and co-authorship has increased
- In the 1980s it was less than 10% of UK output but now stands at 70%
- Collaboration is correlated with higher citation counts
- How should credit be managed?



## Methods for assigning credit

- Whole counting was standard but many analysts now use a variant of fractional counting
- There is NO definitive 'correct' outcome as opinions differ among researchers as to how credit accumulates
- Interpretation is difficult

- Whole counting
  - A paper counts once for each author and once for each institution and each country in their address affiliations
- Fractional counting
  - Credit is partitioned among the authors and their affiliations
  - There are many variants using equal and weighted fractions
  - Credit for institutions may be calculated separately or as a sum of the affiliated authors
- None of these methods explains the contributions of different collaborations to a final indicator value (such as the average CNCI for a university)

on challenge



Leiden University | CWTS | CWTS B.V. | Other CWTS sites

**CWTS** Leiden Ranking  
Meaningful metrics

Home | **Ranking** | Information | Downloads | Services | Contact

## CWTS Leiden Ranking 2021

[Start tour](#)

**List view** | Chart view | Map view

**Time period, field, and region/country**

Time period: 2016–2019  
Field: All sciences  
Region/country: World  
Min. publication output: 100

**Indicators**

Type of indicators: Scientific impact  
Indicators: P, P(top 10%), PP(top 10%)  
Order by: PP(top 10%)  
 Calculate impact indicators using fractional counting

	University	P	P(top 10%)	PP(top 10%)	
1	Rockefeller Univ	887	283	31.9%	
2	MIT	10507	2616	24.9%	
3	Princeton Univ	5332	1214	22.8%	
4	Caltech	5125	1125	21.9%	
5	Stanford Univ	16454	3563	21.7%	
6	Harvard Univ	34234	7246	21.2%	
7	Univ California - Berkeley	10006	2103	21.0%	
8	Weizmann Inst Sci	2536	502	19.8%	
9	Ecole Polytech Féd Lausanne	5704	1076	18.9%	
10	Univ California - San Francisco	10070	1888	18.8%	
11	Univ Cambridge	14080	2628	18.7%	
12	Univ Chicago	7679	1428	18.6%	

Dashboard offering bibliometric indicators of the performance of universities in four areas:

- Scientific impact
- Collaboration
- Open access
- Gender diversity

Scientific impact is measured by counting highly cited publications authored by a university

Two approaches for handling collaborative publications:

- Full (whole) counting
- Fractional counting

# Fractional counting

June 01 2020

## A principled methodology for comparing relatedness measures for clustering publications


Ludo Waltman  , Kevin W. Boyack  , Giovanni Colavizza  , Nees Jan van Eck 

 Check for updates


▼ Author and Article Information

Ludo Waltman 

Centre for Science and Technology Studies, Leiden University, The Netherlands

Kevin W. Boyack 

SciTech Strategies, Inc., Albuquerque, NM, USA

Giovanni Colavizza 

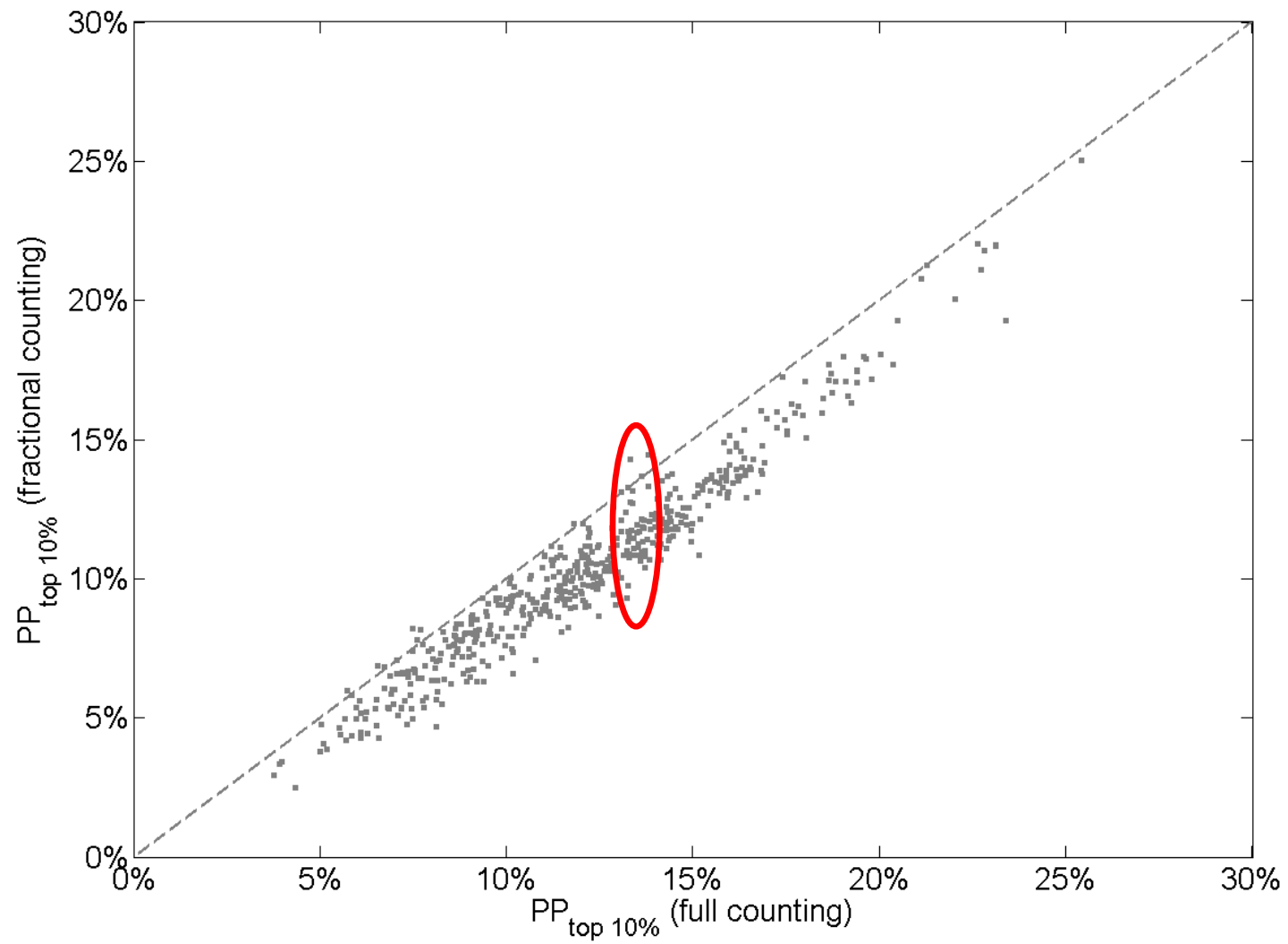
University of Amsterdam, The Netherlands

Nees Jan van Eck 

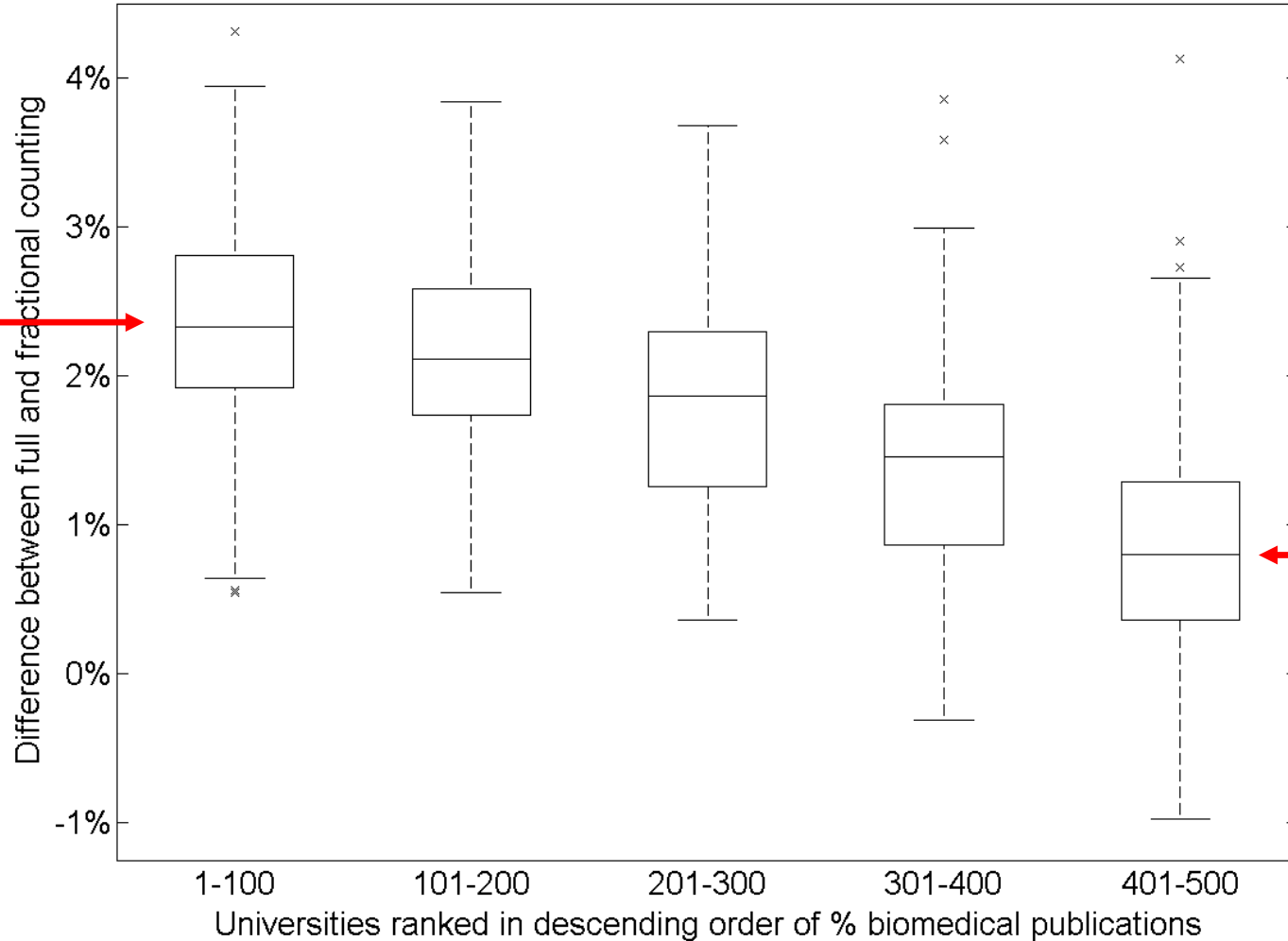
Centre for Science and Technology Studies, Leiden University, The Netherlands

In the Leiden Ranking fractional counting approach, this publication is assigned to the three collaborating organizations with the following weights:

- Leiden University: 0.50
- University of Amsterdam: 0.25
- SciTech Strategies: 0.25



Universities with strong biomedical focus benefit from 'double counting' of collaborative publications in full counting approach



Universities focusing on other disciplines, such as social sciences and humanities, are disadvantaged





# Conclusions

- Full counting approach easily leads to misleading conclusions, biased in favor of disciplines with a high level of collaboration
- Fractional counting approach offers an alternative that doesn't suffer from this problem
- New Clarivate approach offers yet another way of understanding the complex interplay between collaboration on research impact

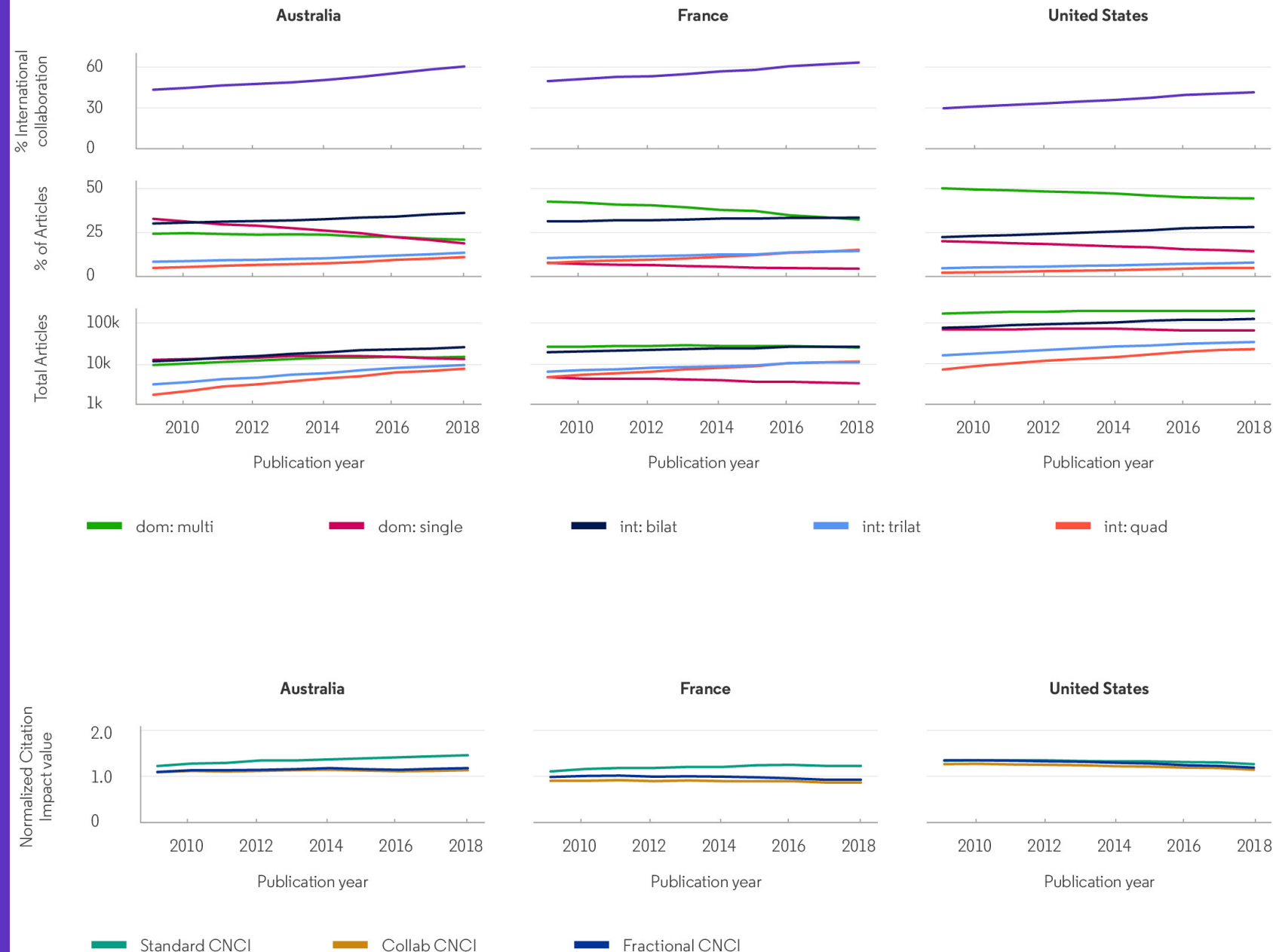
# A new approach – Collaborative CNCI

- Profiles, not metrics
- ISI's new method breaks down a typology of collaboration
- The components show where impact appears
- A net, overall value gives an indicator
- Journal of Informetrics paper:  
<https://doi.org/10.1016/j.joi.2020.101075>

- Consider collaboration *without* fractionalising credit
- Domestic papers (no international co-author)
  - Single author (dom: single)
  - Multiple authors (dom: multi)
- Internationally co-authored papers
  - Two countries (int: bilat)
  - Three countries (int: trilat)
  - Four or more countries (int: quad, less than 4%)
- Follow the same method as CNCI, but with the additional normalisation by collaboration type
  - Avoids complex fractional analysis
  - No more difficult to understand or calculate than standard CNCI
  - Can compare Collaborative CNCI values to standard and fractional methods
  - Deconstruct research portfolios by collaboration type

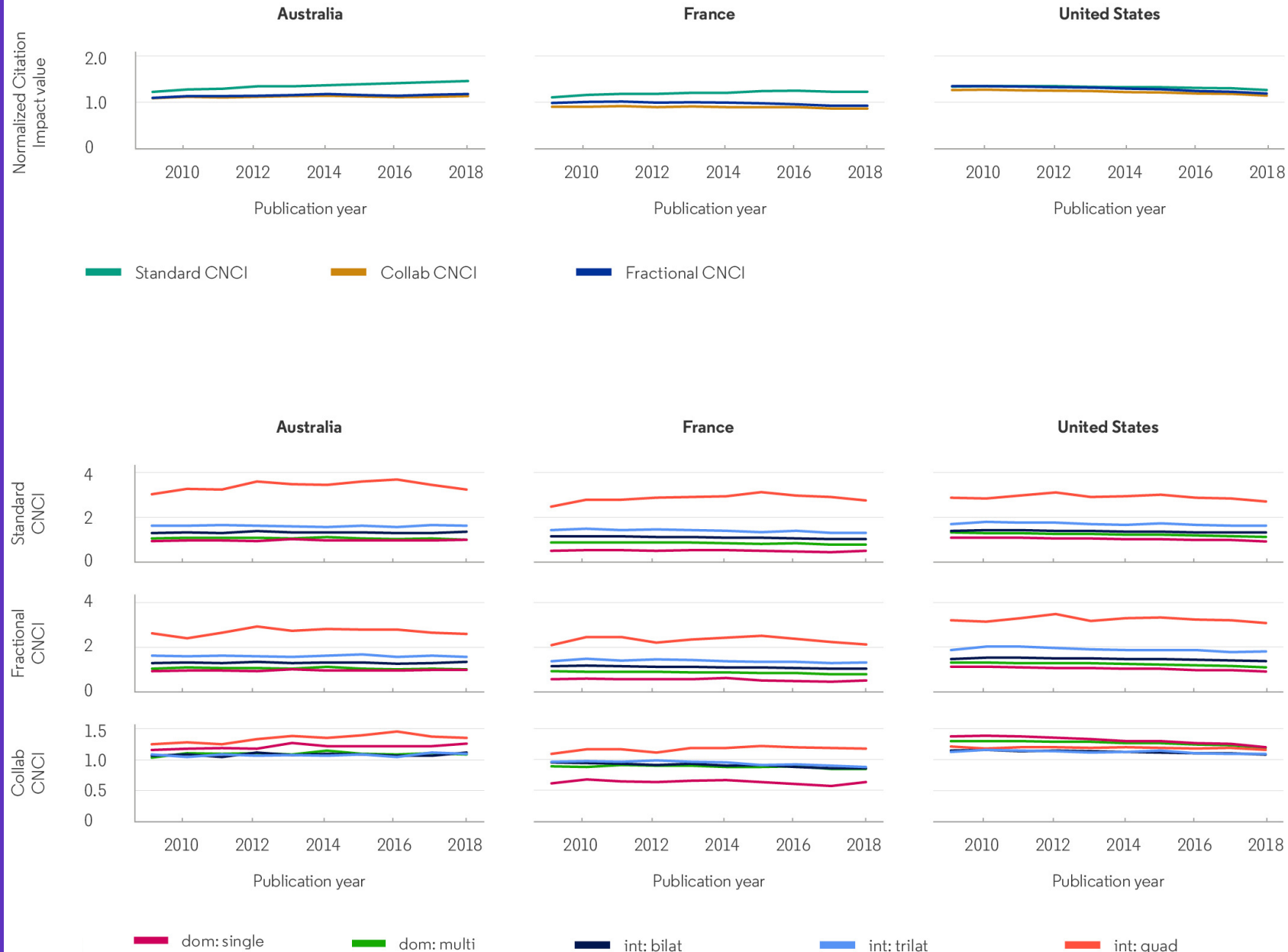
# Established research economies

- Three countries with a strong research base
- Rising collaboration, mostly bilateral
- Compare CNCI by three methods
- Collab CNCI is generally less than Standard (whole counting) but similar to a common Fractional method



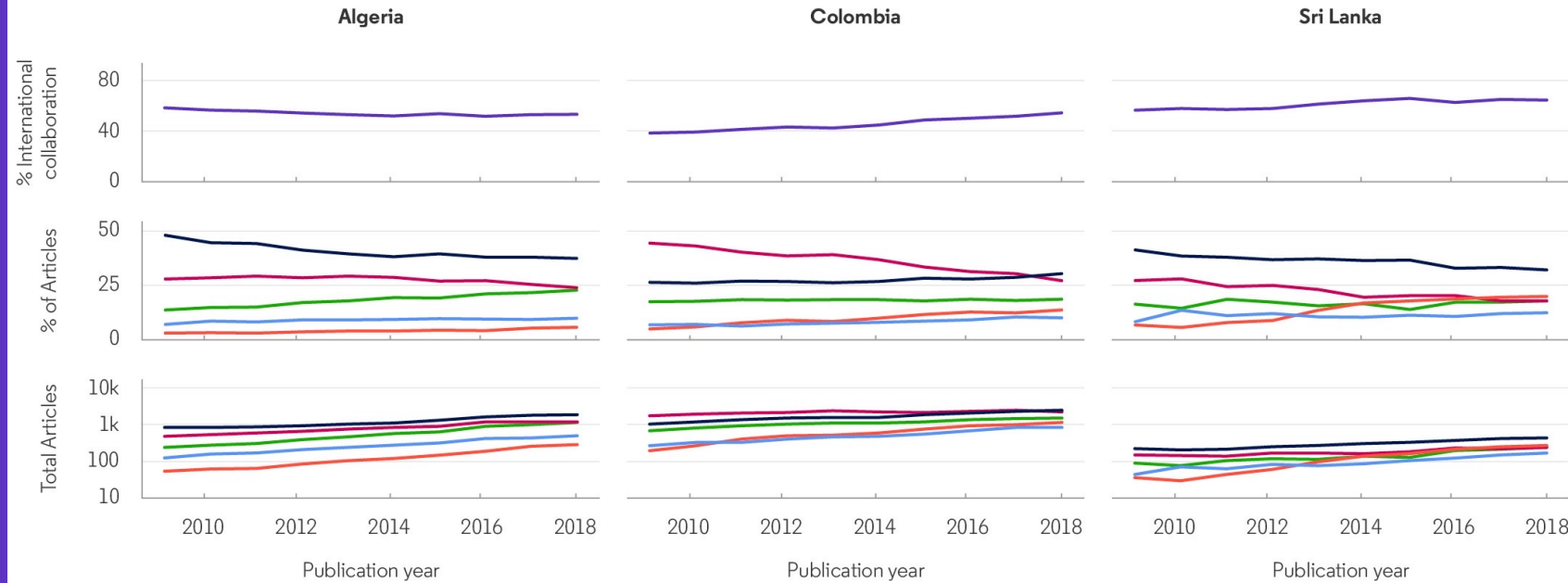
# Deconstructed impact

- Int:quad has high average CNCI in standard and fractional counting
- Normalising by collaboration type moderates this effect
- It also reveals other data, such as France's low domestic impact
- Information content is better, interpretation is easier

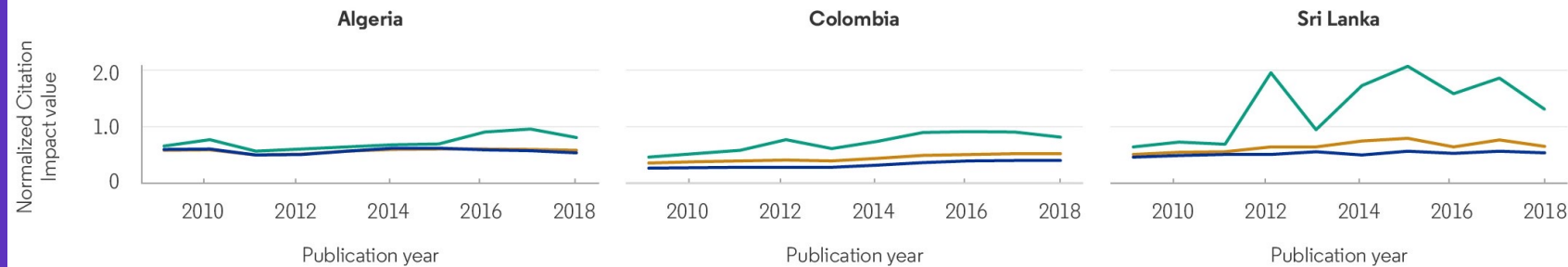


# Growing research economies

- Many more dom:single papers than larger countries
- Higher % international collaboration
- Difference between Standard and other results
- Erratic variation in Standard metrics



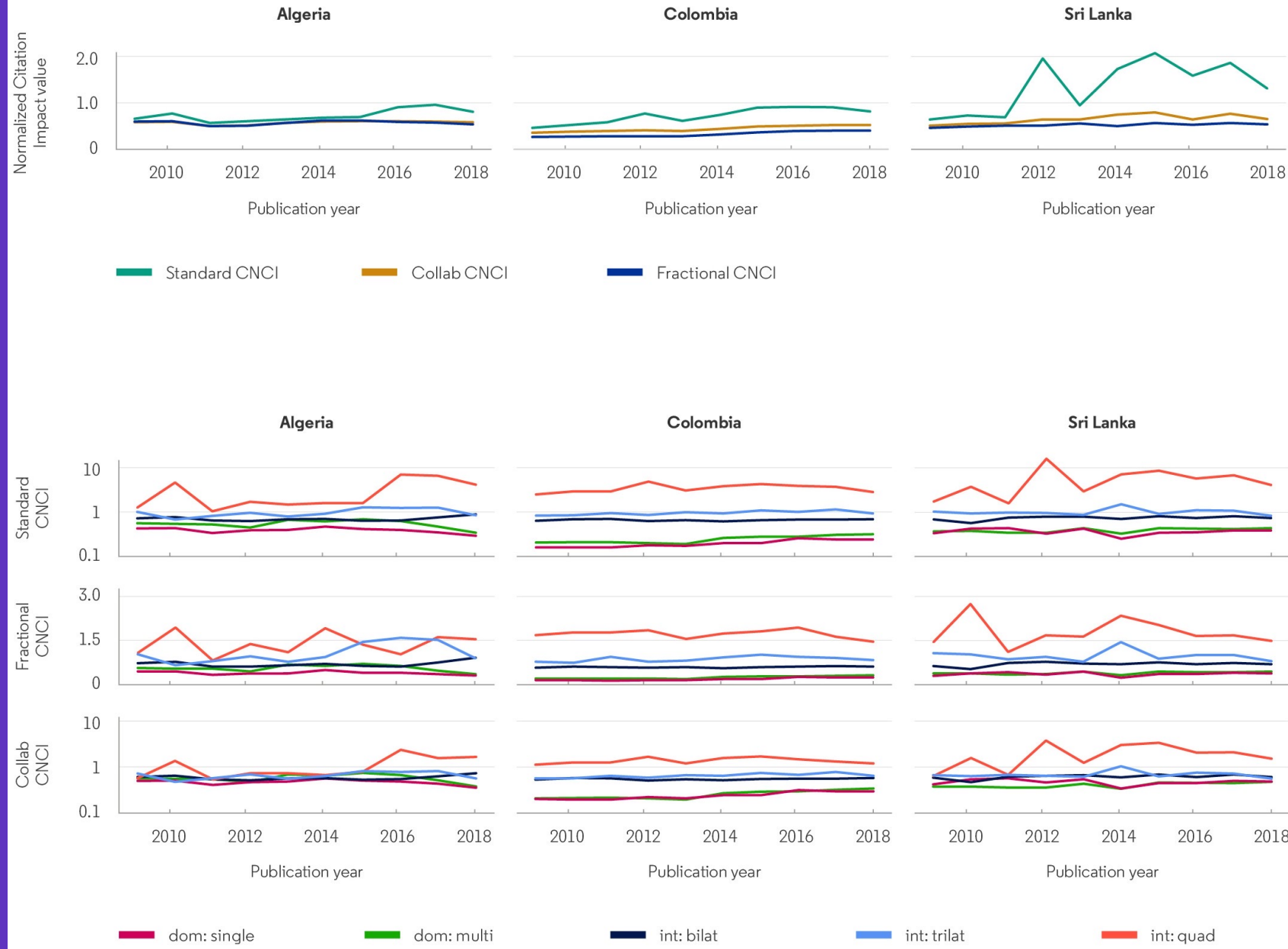
■ dom: single     ■ dom: multi     ■ int: bilat     ■ int: trilat     ■ int: quad



■ Standard CNCI     ■ Collab CNCI     ■ Fractional CNCI

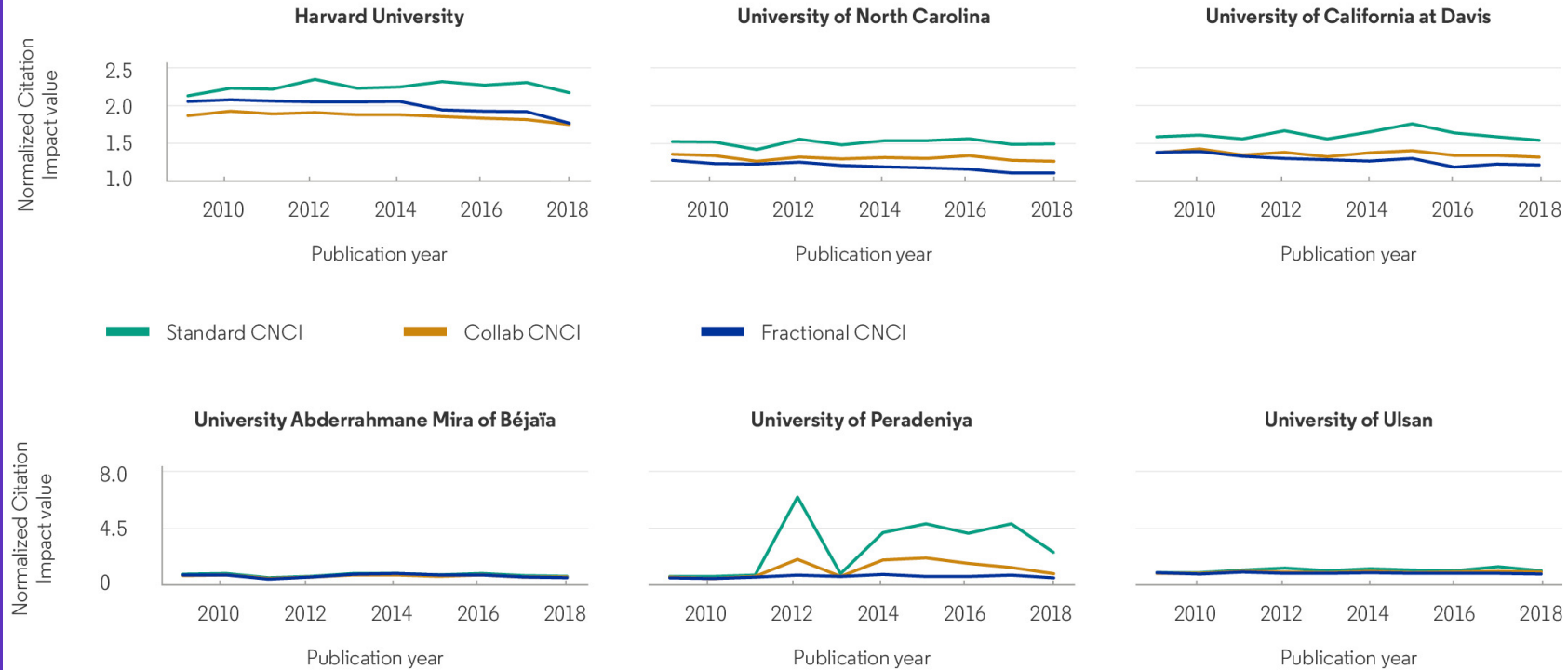
# Deconstructed impact

- Int-quad stands out for all methods and reveals erratic annual change
- Colombia and Sri Lanka – weak domestic research impact
- Single indicator values cannot be properly interpreted without the deconstruction



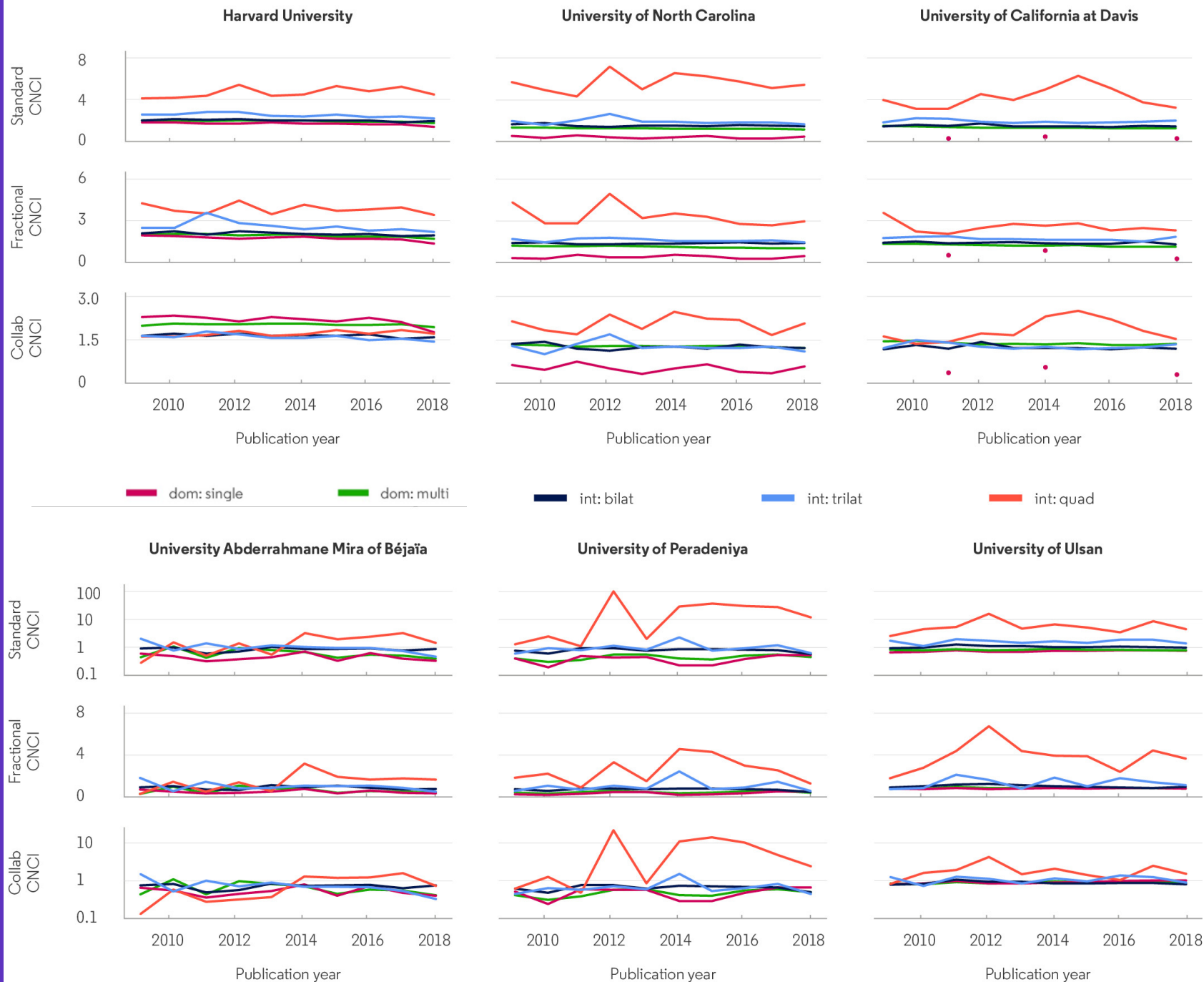
# Individual institutions

- Three US universities
- One from Algeria, Sri Lanka and South Korea
- Innovative comparisons
  - Harvard/collab CNCI
  - Peradeniya spikes



# Institutions deconstructed

- NOTE linear scale (US) & log scale (OTHERS)
- Relative impact value of different modes of publication now clear for managers
- Int:quad always high and variable, but more stable for US







# Making it count

Thank you for your attention

Jonathan Adams, Ludo Waltman and  
Ross Potter

April 2022

