Web of Science Group presents the *Research. Smarter.* webinar series. Essential resources, tips, and guidance to help you power through each stage of your research journey.
Intelligent metrics: The need for a Responsible Use of Research Metrics

Guillaume Rivalle, Manager, Solution consultants, Europe

December 2019
Research evaluation is increasingly being conducted using bibliometric methodology and citation analysis.

Find out how *InCites* supports a comprehensive class of advanced bibliometric indicators and a responsible research evaluation.

<table>
<thead>
<tr>
<th></th>
<th>Reliable metrics require reliable data</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>The need for normalization</td>
</tr>
<tr>
<td>3</td>
<td>Combining metrics</td>
</tr>
<tr>
<td>4</td>
<td>Working within the appropriate data perimeter</td>
</tr>
</tbody>
</table>
Reliable metrics require reliable data
What data?

The reliability of all analysis is directly linked to the quality and curation of the data you will use.

The Web of Science Core Collection is a trusted, high quality collection of journals, books and conference proceedings.

Disciplinary and geographical coverage has to be balanced and should be a true reflection of the research community publication activities.

Sources are carefully selected by our editorial team, and decisions are made by our expert in-house editors, with no affiliations to publishing houses or research institutes.

For the last 60 years, bibliographic metadata (all authors, affiliations, cited references, etc.) has been meticulously captured with the same policy, across the entire archive. This eliminate gaps issues and ensures the stability of the data throughout the years.

Web of Science Core Collection (1980-present)

40 years of publication activities, with full cited references and consistent bibliographic metadata

Bibliometric analysis in InCites

1. Initial Triage
   - ISSN
   - Journal Title
   - Journal Publisher
   - URL (online journals)
   - Content Access
   - Presence of Peer Review Policy
   - Contact Details

2. Editorial Triage
   - Scholarly Content
   - Article Titles and Article Abstracts in English
   - Bibliographic Information in Roman Script
   - Clarity of Language
   - Timeliness and/or Publication Volume
   - Website Functionality/Journal Format
   - Presence of Ethics Statements
   - Editorial Affiliation Details
   - Author Affiliation Details

3. Editorial Evaluation
   - Editorial Board Composition
   - Validity of Statements
   - Peer Review
   - Content Relevance
   - Grant Support Details
   - Adherence to Community Standards
   - Author Distribution
   - Appropriate Citations to the Literature

Quality Criteria

Impact Criteria

Successful outcomes

Starts editorial triage
Starts editorial evaluation
Enters ESCI and is evaluated for impact
Enters SCIE/SSCI/AHCI

Unsuccessful outcomes

Submission cannot be completed
Re-submission welcome as soon as issues have been resolved

Failed editorial triage
Re-submission welcome as soon as issues have been resolved

Failed editorial quality evaluation
Re-submission subject to embargo of at least two years

Failed editorial impact evaluation
Entry/continued coverage in ESCI
Re-evaluation subject to embargo of at least two years
The need for normalization
Young papers are usually less cited than older ones. The age of publications should therefore be taken into account when measuring their impact.
Citation rates variations

Citation averages of the world’s publications by document types

The types of publications should therefore be taken into account when measuring their impact.
Citation rates variations

Citation counts of the world’s publications in various broad disciplines

The amount of citations received greatly varies between large disciplines.

The subject categories of publications should therefore be taken into account when measuring their impact.
Citation rates variations
Citation averages of the world’s publications in two disciplines

Even when selecting the same document type, the difference in “citation behavior” is very strong between disciplines.

In mathematics, researchers cite older papers than in other areas, e.g. Social Sciences.
Normalization in action
Normalization has to be calculated at article level

At the heart of InCites:
The following is calculated every month for each of the 65M+ publications of the InCites dataset

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>00034696950001</td>
<td>10.1186/s12284-014-0029-y</td>
<td>26224558</td>
<td>The roots of future rice harvests</td>
<td>RICE</td>
<td>AGRONOMY</td>
<td>Review</td>
<td>2014</td>
<td>20</td>
<td>29.88</td>
<td>0.66</td>
<td>32.33</td>
<td>0.62</td>
<td>47.27</td>
</tr>
</tbody>
</table>

How strong/weak is this impact?
### Normalization in action

Normalization has to be calculated at article level

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
</table>

Category expected citations is calculated in the context of the category, document type and publication year.
## Normalization in action

Normalization has to be calculated at article level.

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>00034696950001</td>
<td>10.1186/12284-014-0029-y</td>
<td>26224558</td>
<td>The roots of future rice harvests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>29.88</td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Category Normalized Citation Impact calculates how above or below expected the times cited count is.
### Normalization in action

Normalization has to be calculated at article level

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
</table>

Journal expected citations is calculated in the context of the Journal, document type and publication year.
### Normalization in action

Normalization has to be calculated at article level

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>00034696950001</td>
<td>10.1186/12284-014-0029-y</td>
<td>26224558</td>
<td>The roots of future rice harvests</td>
<td>EAGLE AGRO</td>
<td>Review</td>
<td>2014</td>
<td>29.88</td>
<td>0.66</td>
<td>32.33</td>
<td>0.62</td>
<td>47.27</td>
<td>47.27</td>
<td></td>
</tr>
</tbody>
</table>

Journal Normalized citation Impact calculates how much below or above expected the times cited count is
Normalization in action

Normalization has to be calculated at article level

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>00034696950001</td>
<td>10.1186/s12284-014-0029-y</td>
<td>26224558</td>
<td>The roots of future rice harvests</td>
<td>AGRONOMY</td>
<td>Review</td>
<td>2014</td>
<td>20</td>
<td>47.27</td>
<td>32.33</td>
<td>0.62</td>
<td>0.66</td>
<td>47.27</td>
<td>0.66</td>
</tr>
</tbody>
</table>

The percentile in which the paper ranks in the category
Normalization in action

Normalization has to be calculated at article level

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>DOI</th>
<th>Pubmed ID</th>
<th>Article Title</th>
<th>Journal</th>
<th>Research Area</th>
<th>Document Type</th>
<th>Publication Date</th>
<th>Times Cited</th>
<th>Category Expected Citations</th>
<th>Category Normalized Citation Impact</th>
<th>Journal Expected Citations</th>
<th>Journal Normalized Citation Impact</th>
<th>Percentile in Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>00034696950001</td>
<td>10.1186/s1284-014-0029-y</td>
<td>26224558</td>
<td>The roots of future rice harvests</td>
<td>RICE</td>
<td>AGRONOMY</td>
<td>Review</td>
<td>2014</td>
<td>20</td>
<td>29.88</td>
<td>0.66</td>
<td>32.33</td>
<td>0.62</td>
<td>47.27</td>
</tr>
</tbody>
</table>

ANVUR: 16 categories
CAPES: 179 categories
CSADC 89 categories
ESI: 22 categories
FAPESP: 81 categories
GIPP: 6 categories
KAKEN: 76 categories
OECD: 48 categories
UK RAE: 67 categories
UK REF: 36 categories
WoS: 250 categories
TOTAL: 870 categories

870 x 43 x 40 = 1.5M Category Expected Citations are calculated every month
Normalization in action
Normalized metrics applied to groups of publications

Using powerful filters, publications can be grouped and normalized metrics can be aggregated, allowing for fair benchmarking.

Publications can be grouped by any metadata:
- Country
- Institution
- Author
- Journal
- Publisher
- Research funder
- Category
- Open Access
- Etc.
The JIF, as a value, should always be looked at within the context of the categories the journal is indexed in.

Journal Impact Factor values, but also quartiles, are available in InCites Benchmarking analytics.
Combining several metrics
Use a combination of metrics
One single metric will never be able to answer all questions

Combining different metrics provides a much clearer picture and helps make balanced conclusions

Exporting reports and list of documents from InCites will allow further analysis of all group studied.
E.g. Citation distribution

The citation distributions of the two group are very different
Use a combination of metrics
One single metric will never be able to answer all questions

Radar charts are a good way to plot several metrics simultaneously

Group A
Group B
Group C
Use a combination of metrics

One single metric will never be able to answer all questions

Scatter Plots can plot three metrics simultaneously
Are you working with the right data perimeter?
Limiting the analysis to meaningful clusters for meaningful findings

Example: I am looking for Scottish organizations publishing influential papers in Environmental sciences

Are these groups of papers clusters that I can base my conclusions on?

By limiting the clusters to a minimum size, I can ensure to obtain more reliable metrics

Groups of papers can also be downloaded, taken apart and analyzed (e.g. for citation distribution analysis)
Limiting the analysis to meaningful clusters for meaningful answers

I am looking for organizations with strong influence in Physics

I know certain physics papers tend to contain many authors. Shouldn’t I try to split my analysis in different groups?
Using different Research Areas Schemes

From broad disciplinary definitions to more granular subject categories

Running analysis on different schema to identify strengths and weaknesses from top to granular levels

A mapping of all Research Areas can be found in InCites Help pages
Working with unified data

Example:
National Hospital Norway
206 unification rules:

Consistent capture of all author affiliations

Careful manual unification of Research organization names +

Reliable Organizational level analysis.

Example: Collaboration analysis:
Working with unified data
Research funding organizations

Example:

**Versus Arthritis**
212 unification rules:

- Arthritis Research, UK
- Arthritis Research, UK (BeBack)
- Arthritis Research-UK
- Arthritis Research-UK Fellowship
- Arthritis Research-United Kingdom Grant
- Arthritis Rheumatism Campaign UK
- Arthritis UK
- Arthritis UK (Equipment Grant)
- ARUK Rheumatoid Arthritis Pathogenesis Centre for Excellence (RACE)
- Biotechnology and Biological Sciences Research Council Industry Interchange Award programme, Arthritis Research UK
- Biotechnology and Biosciences Research Council and an Arthritis Research Campaign (ARC)
- British Arthritis Research Campaign
- Cambridge NIHR Biomedical Research Centre (Metabolism, Endocrinology, Bone and Bio-materials Theme), Arthritis Research UK
- Career Development Fellowship from Arthritis Research United Kingdom
- Centre for Osteoarthritis Pathogenesis Grant charity Arthritis Research UK

Careful manual unification of Research funding organization names + Consistent capture of author acknowledgments

Reliable Funder level analysis.
Example: Proportion of Open Access output

![Diagram showing the process of careful manual unification and consistent capture of author acknowledgments leading to reliable funder level analysis.](Image)
Working with unified data

Authors

Publons/RID:

Efficient publication claiming tools for researchers

ORCID:

Automated ingestion into the bibliographic dataset:

= Reliable author level analysis:

Unlocking bat immunology: establishment of Pteropus alecto bone marrow-derived dendritic cells and macrophages

Reliable author level analysis:
Build a more precise perimeter of analysis
Go beyond our unifications and categorization

Use Web of Science to:
Define a precise research domain by searching specific keywords in the indexed publications
Search a department name in a author addresses to recreate a departmental perimeter
Search for RID and/or ORCID’s to recreate a team

Use your own data source (CRIS, repository, database, etc.) to build a list of Unique WoS identifiers, DOI’s or PubMed ID's that represents the group or publications you want to study (e.g. department)

Reproduce your organization structure in My Organization, using:
Your own data sources (e.g. CRIS)
Our Faculty Match Service

Analyze these precise datasets in InCites.

Upload to InCites via file upload or API

Analyze your full organization in InCites:
- Control the full dataset
- Departmental analysis within a 5 levels hierarchy
- Control all labels (departments and researchers)
- Researcher level validated publications for true people analysis
- Internal collaborations analysis at researcher and department levels

Save to InCites

Upload to InCites
Useful Links

Libguide  clarivate.libguides.com/home
YouTube videos  youtube.com/WoSTraining
Blog  clarivate.com/blog/category/science-research-connect/
Google Group  groups.google.com/forum/#!forum/incites-user-group
Thank you

Guillaume Rivalle
guillaume.rivalle@clarivate.com
clarivate.com/products/incites