

INFLATED JOURNAL VALUE RANKINGS: PITFALLS YOU SHOULD KNOW ABOUT HTML AND PDF USAGE

Chan Li & Jacqueline Wilson

California Digital Library, University of California

ALA 2015, San Francisco



UNIVERSITY
OF
CALIFORNIA

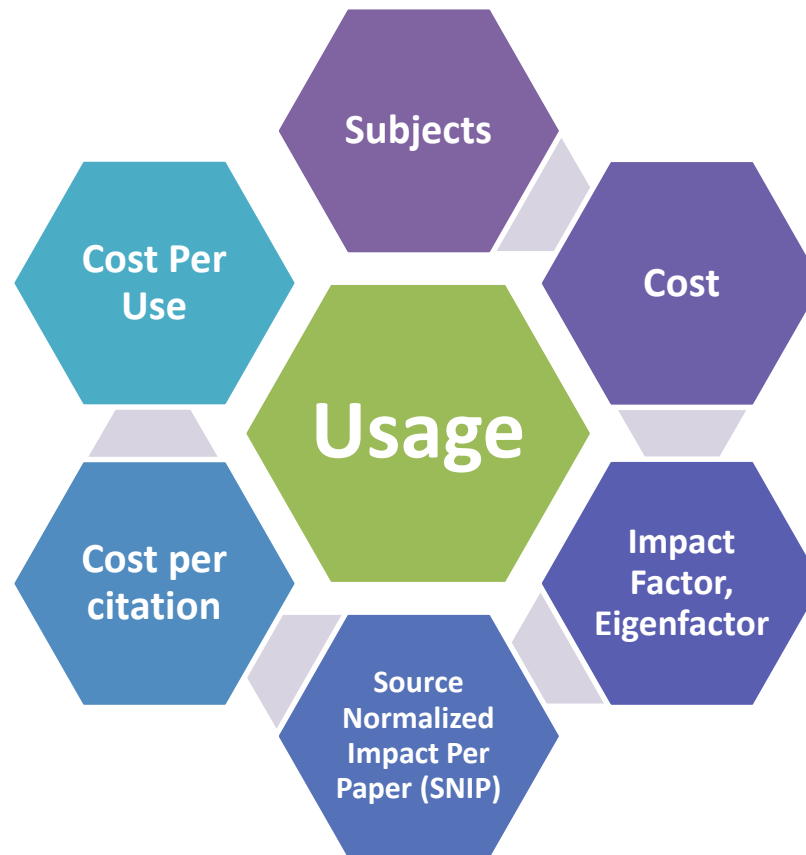
HTML + PDF

**EJOURNAL USAGE:
WHAT DO YOU COUNT?**

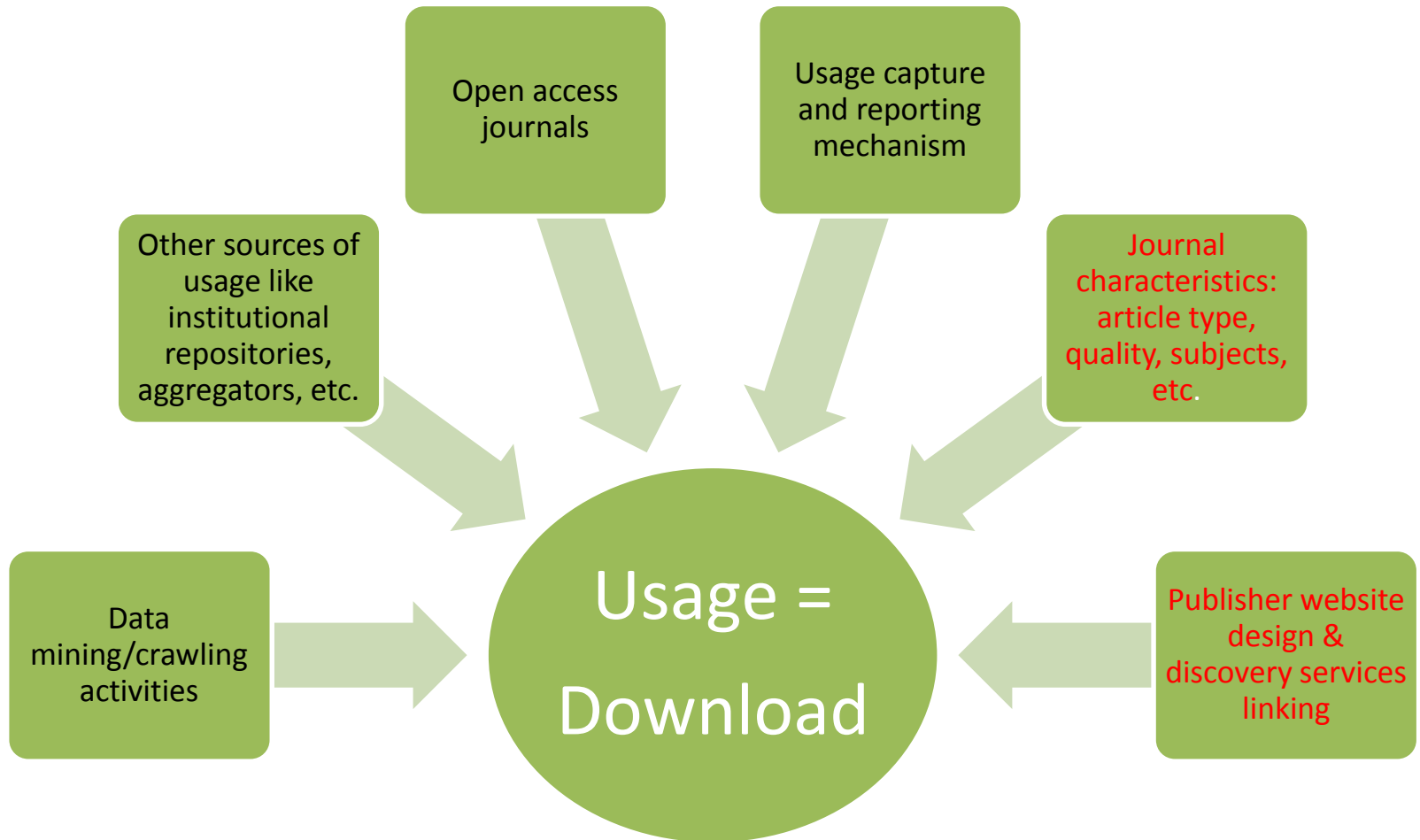
HTML

PDF

JOURNAL **USAGE** IS **CENTRAL** TO COLLECTION DEVELOPMENT AND ASSESSMENT

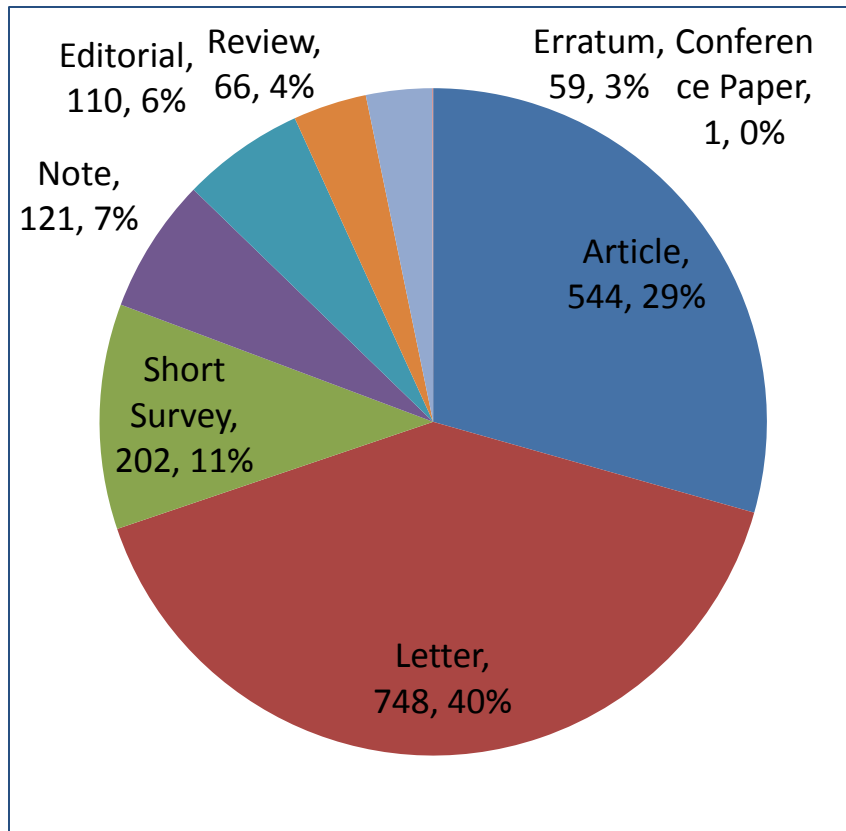


MANY FACTORS CAN AFFECT PUBLISHER E-JOURNAL USAGE REPORTING

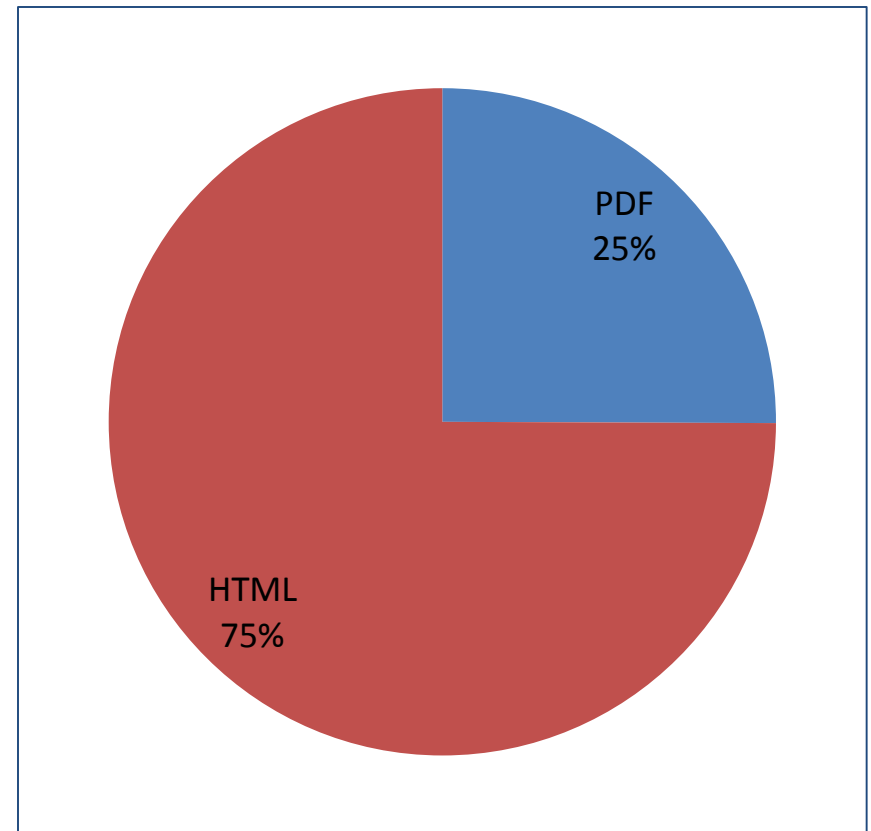


JOURNAL CONTENT TYPES CAN AFFECT USAGE

Content Type Distribution



Usage Distribution



New England Journal of Medicine 2013

PUBLISHER WEBSITE DESIGN & DISCOVERY SERVICES

LINKING CAN AFFECT USAGE REPORTS

Landing page: abstract, links to both HTML and PDF versions

Landing page: fulltext HTML

The screenshot shows the Science journal website interface. The article title is "Forecasting Regional Climate Change Flunks Its First Test" by Richard A. Kerr. The "Article Views" sidebar on the left contains three items: "Summary", "Full Text", and "Full Text (PDF)". Red arrows point to the "Full Text" and "Full Text (PDF)" links. The word "Abstract" is written in red over the main article text.

The screenshot shows the New England Journal of Medicine website. The article title is "The Politics of Emergency Contraception" by Alastair J.J. Wood, M.D., Jeffrey M. Drazen, M.D., and Michael F. Greene, M.D. The word "Fulltext HTML Article" is written in red over the article text. A red arrow points to the "PDF" link in the "TOOLS" sidebar on the right.

PUBLISHER WEBSITE ISSUE

- **Usus:** an independent community-run website for those interested in the usage of online content.
- **Issue Report:** *“I believe there is a problem with some publishers double-counting use in JR1s. Some publishers link directly to HTML via a link resolver (1 use) and when the same user clicks on the PDF version moments later (2nd use) a second use is counted.”* April 13, 2015
- **Usus Response:** *“COUNTER investigated this issue a few years back and addressed concerns raised about the inflation of full text request counts due to the **“interface effect”** where the users are automatically presented with the HTML version and access the PDF from the HTML view. COUNTER’s investigation into the situation resulted in the introduction of separate PDF and HTML counts in COUNTER JR1 and JR1 GOA reports...”*

RESEARCH PLAN

RESEARCH QUESTION
IDENTIFICATION

```
graph TD; A[RESEARCH QUESTION IDENTIFICATION] --> B[LITERATURE REVIEW]; B --> C[DEFINING THE LANDSCAPE: PUBLISHER SAMPLE ANALYSIS]; C --> D[DATA COLLECTION & REGRESSION ANALYSIS]; D --> E[RESULTS & REPORT WRITING];
```

LITERATURE REVIEW

DEFINING THE LANDSCAPE:
PUBLISHER SAMPLE ANALYSIS

DATA COLLECTION & REGRESSION
ANALYSIS

RESULTS & REPORT WRITING

RESEARCH QUESTION IDENTIFICATION

- Do the following affect HTML and PDF usage reports:
 - Discovery services linking
 - Document type
 - Journal citation value
 - Disciplinary differences
 - Publisher platform



DEFINING THE LANDSCAPE: PUBLISHER SAMPLE ANALYSIS



Journal Articles

- 40 Publishers
- From each publisher, we randomly selected articles published in 2000, 2009 and 2014



Discovery Services Linking

- Google Scholar
- Publisher website
- WorldCat Local
- PubMed



Publisher Landing Page

- Abstract, HTML Option, PDF Option
- Full-text: HTML or PDF version

DISCOVERY SERVICES & PUBLISHER WEBSITE LINKING

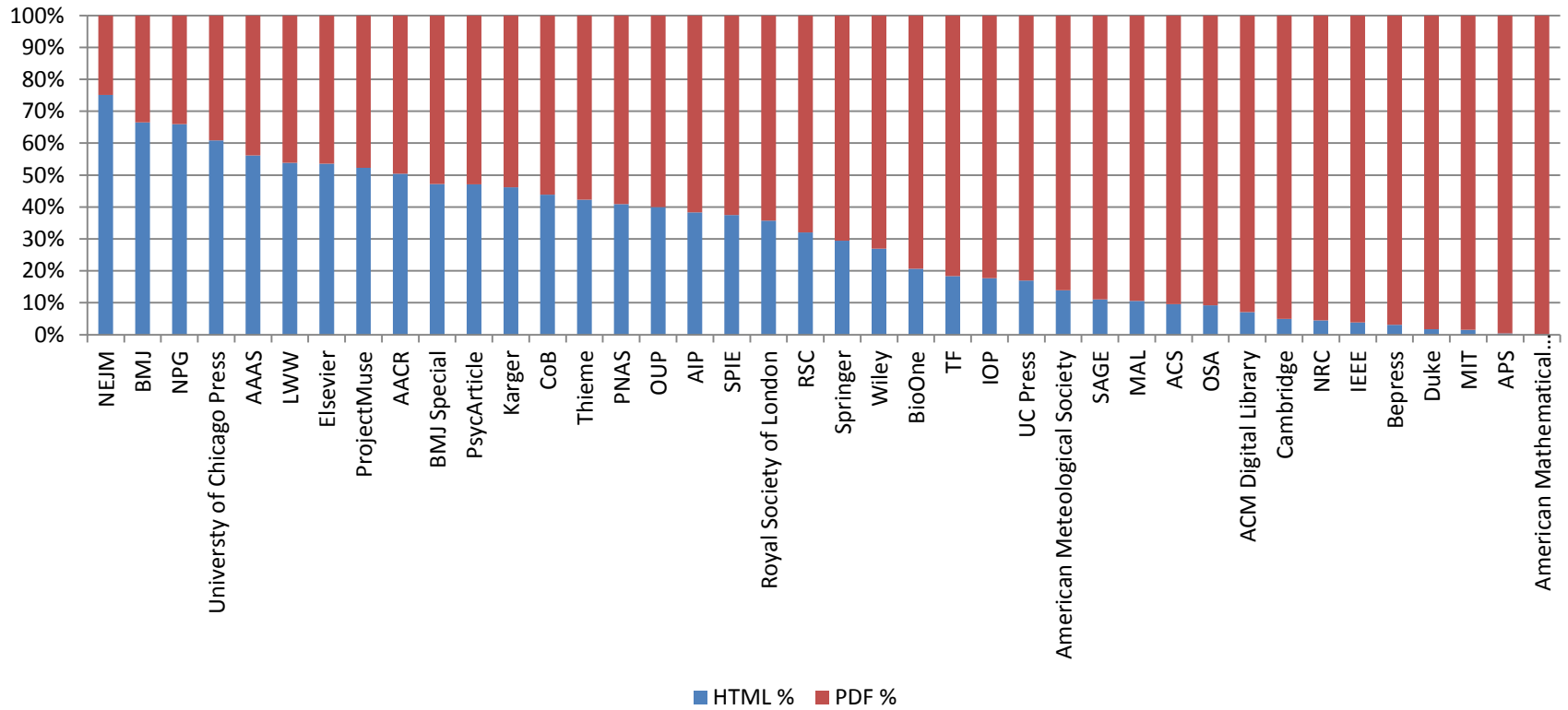
Linking Bias Group: 23 publishers link to full-text HTML articles as the default from some of the search portals below.

No Linking Bias Group: 17 publishers link to article abstracts and options to download either HTML or PDF full-text articles as the default from all of the search portals below.



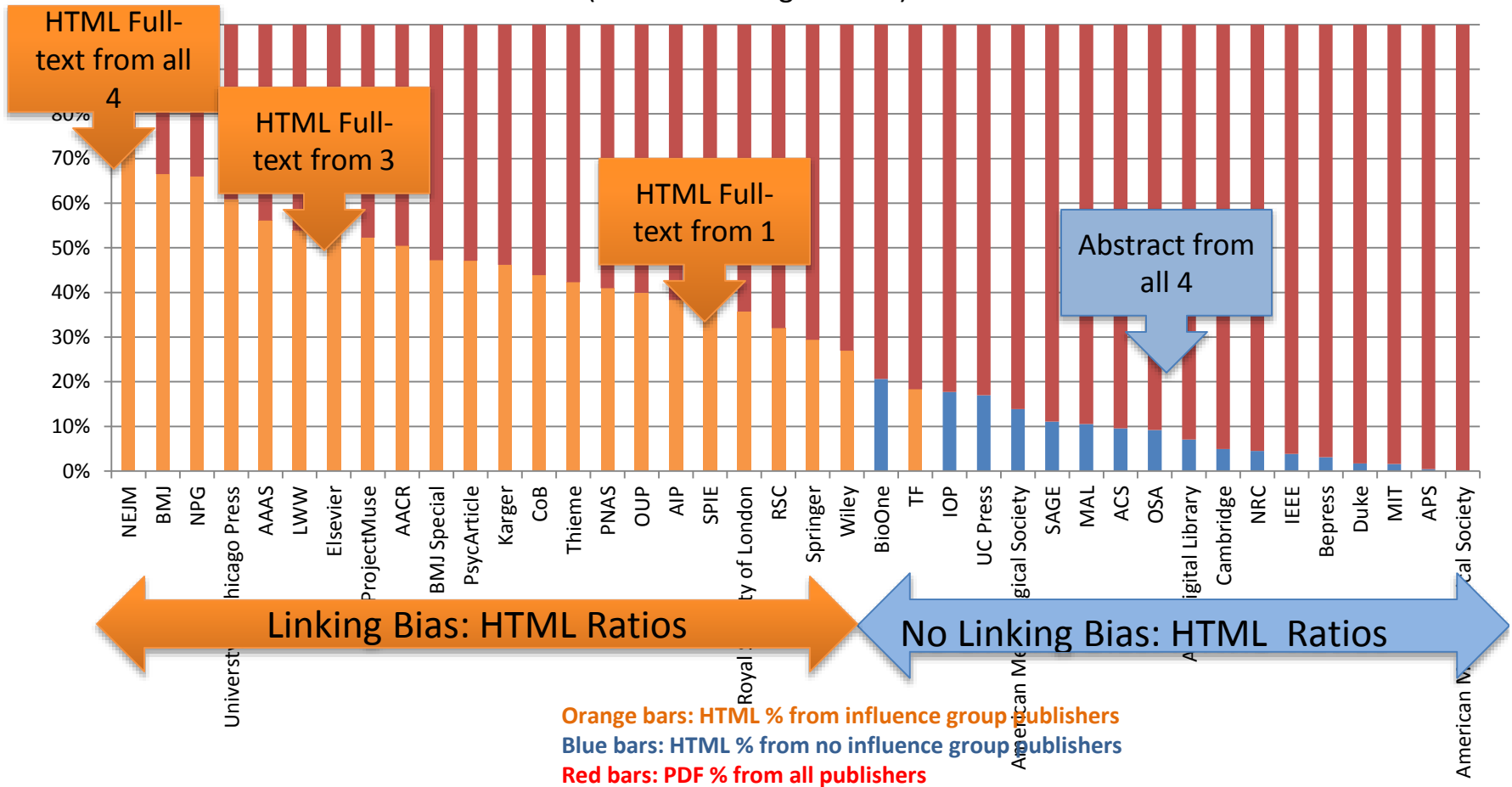
HTML AND PDF USAGE RATIOS VARY GREATLY AMONG PUBLISHERS

UC 2013 HTML Usage To Total Ratio by Publisher
(sorted from high to low)



HTML USAGE IS GREATLY IMPACTED BY DISCOVERY SERVICES LINKING & PUBLISHER WEBSITE DESIGN

UC 2013 HTML Usage To Total Ratio by Publisher
(sorted from high to low)



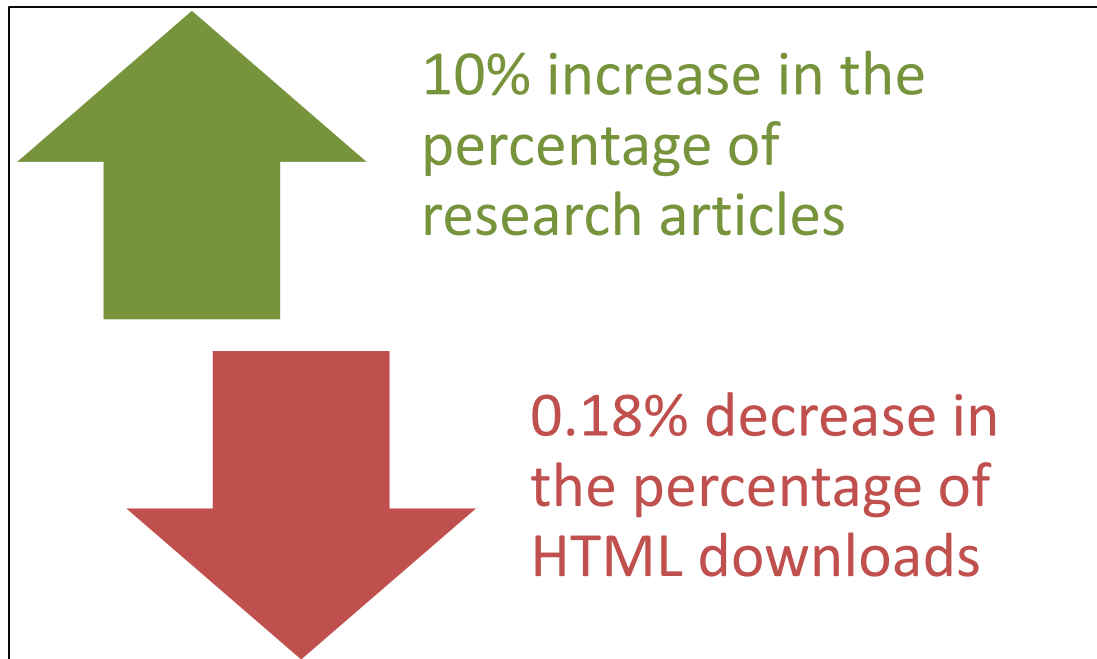
DATA COLLECTION & REGRESSION ANALYSIS

- Analysis sample:
 - 20+ journal publishers with 4,500 journals
- Methodology
 - Regression analysis to estimate the effect of various factors on HTML/PDF usage
- Factors Analyzed
 - Document Type
 - Journal Citation Value
 - Disciplinary Differences
 - Publisher Platform



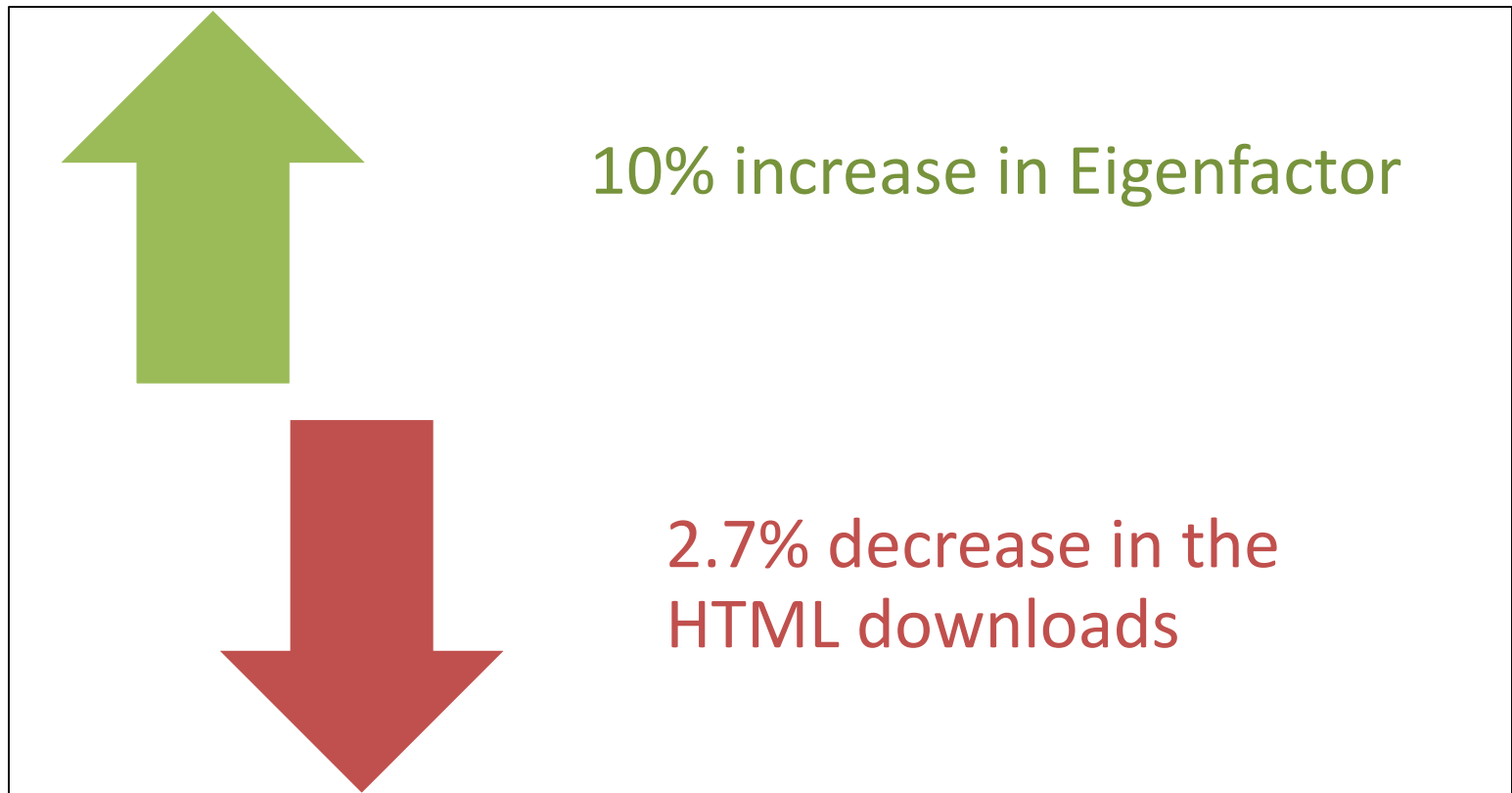
DOCUMENT TYPE HAS **LIMITED EFFECT** ON HTML USAGE

All Publishers



Note: Research article is defined as the combination of article, review and conference papers. The document type data is from Scopus.

**JOURNAL CITATION VALUE EFFECT:
HIGHLY CITED JOURNALS HAVE FEWER HTML
DOWNLOADS AND MORE PDF DOWNLOADS**



Data source: Eigenfactor: <http://www.eigenfactor.org/faq.php>

**DISCIPLINARY
DIFFERENCES OVERALL
HAVE **LIMITED EFFECT**
ON HTML USAGE**

**More HTML
Usage %**

Nursing

Social
Science

Neuroscience

Computer
science

**More PDF
Usage %**

Mathematics

Veterinary

Dentistry

Economics

Journal discipline categories are from Scopus

More HTML Usage %

NEJM

SPIE

NATURE

Elsevier

More PDF Usage %

Springer

NRC

MAL

OUP

JOURNAL PUBLISHER PLATFORM HAS THE **MOST SIGNIFICANT** EFFECT ON HTML USAGE

RESULTS & REPORT WRITING

- Discovery services linking options and publisher website choices have significant effect on reader behavior
- Document type, journal citation and disciplinary differences have relatively limited effect on usage data



APPLYING RESEARCH FINDINGS TO LIBRARY USAGE ASSESSMENT

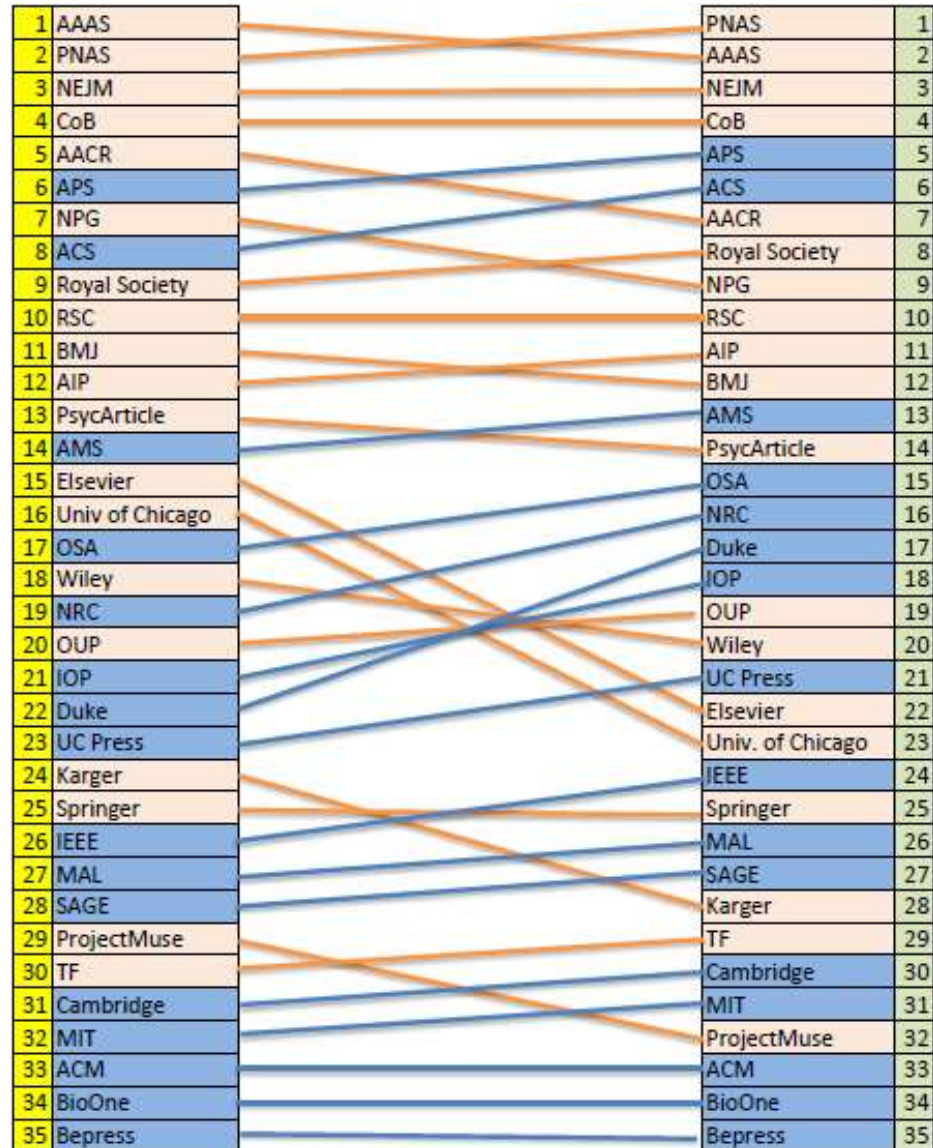
Our research indicates:

- Counting **total usage**, which combines PDF and HTML is not a reliable metric, because of the various factors studied
- Counting **HTML only** is incomplete and skewed
- BEST Options:
 - Count **PDF only**
 - Use a **range from total usage to PDF usage**

WHAT MIGHT HAPPEN IF YOUR LIBRARY CHANGED ITS JOURNAL USAGE COUNTING METHODOLOGY

TOTAL USAGE VS PDF USAGE: PUBLISHER RANKINGS VARY

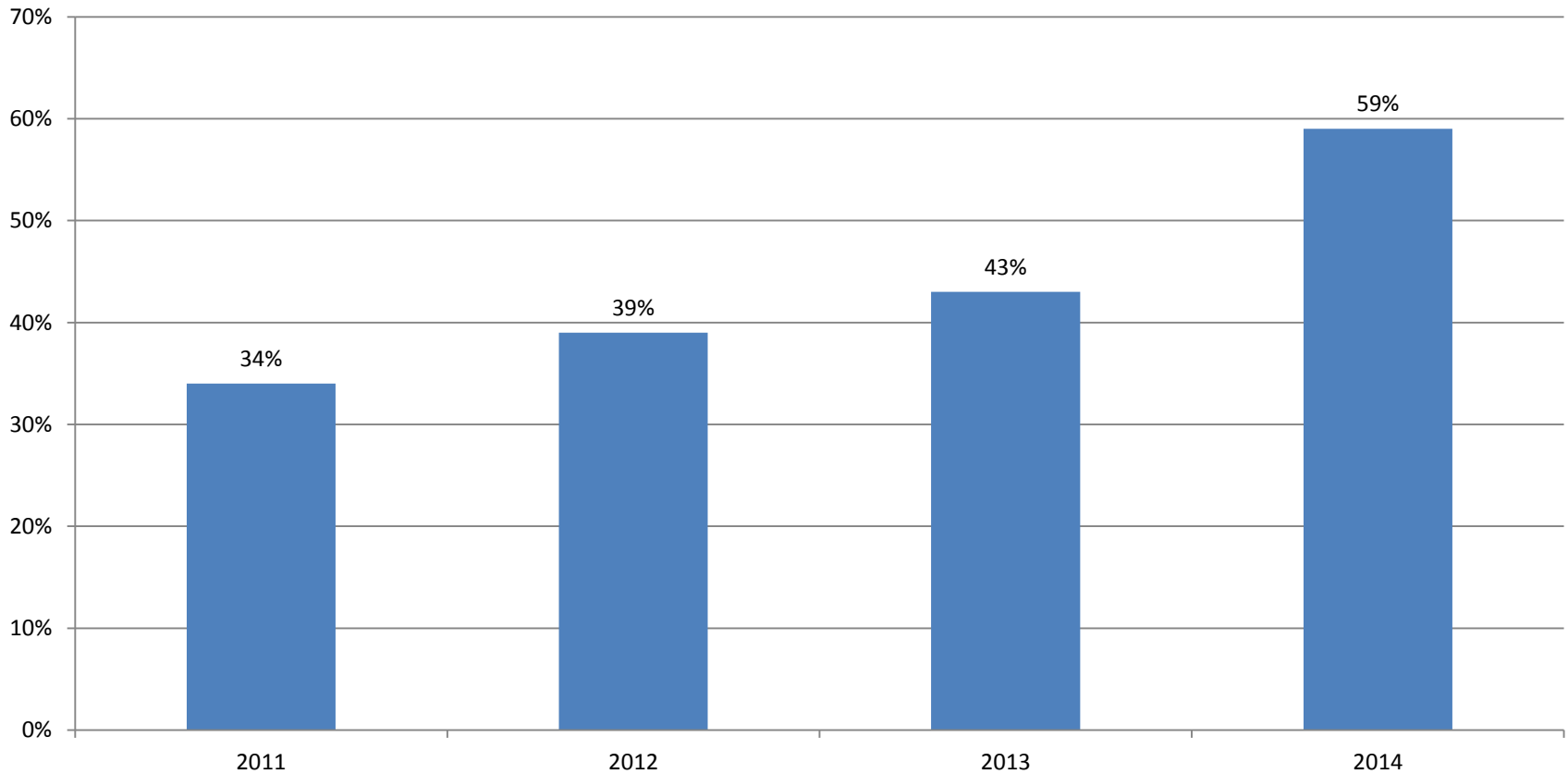
Publisher Ranking by Total Usage



Publisher Ranking by PDF Usage

HTML USAGE RATIO: **SIGNIFICANT GROWTH OVER TIME**

Change in UC's HTML Usage Ratio 2011-2014



REASONS FOR HTML GROWTH: TRENDS IN PUBLISHER WEBSITE DESIGN: ENHANCED HTML-- PRESENTATION, CONTENT AND CONTEXT

PDF (37 pages) E-mail Export More Display mode

Copyright © 2014 Elsevier B.V. All rights reserved.

Take our survey

Outline Show thumbnails [Article top](#)

Abstract



- Vibrancy of the field of leadership—looking backward
- Journal reputation of LQ
 - Table 1
 - Table 2
- Content analysis of LQ publications: 2000–2009
 - Table 3



3.1. LQ publications: What is published and by whom

- 3.1.1. Number and type of publications
- 3.1.2. Author information



The Leadership Quarterly

Volume 21, Issue 6, December 2010, Pages 922–958



Scholarly leadership of the study of leadership: A review of The Leadership Quarterly's second decade, 2000–2009

William L. Gardner ^a, Kevin B. Lowe ^b, Todd W. Moss ^a, Kevin T. Mahoney ^c, Claudia C. Coglisier ^a

^a Corresponding author. Institute for Leadership Research, Rawls College of Business, 15th & Flint, Texas Tech University, Lubbock, TX 79409, United States. Tel.: +1 806 742 1055; fax: +1 806 742 2308.

^a Texas Tech University, United States

^b University of North Carolina-Greensboro, United States

^c Louisiana Tech University, United States


Research highlights

- 353 articles published in the Leadership Quarterly (LQ) in its second decade were analyzed using multiple methods:
 - Impact factor analysis indicates that LQ is comparable in impact to other high quality international business, management and psychology journals. 36 of the top 50 most cited articles were published in the 2000's.
 - The proportion of articles that reflect new directions in leadership theory rose from 14% during the journal's first decade, to 44% in its second decade.
 - The vast majority (79.9%) of empirical research published was conducted in field settings with samples drawn from private organizations (46%); survey research methods (17%) were most prevalent.
 - The majority of articles employed quantitative (82.4%) as opposed to qualitative (24.1%) and mixed (11.6%) methods.

Sidebar content: Overview

Browse individual items

Access to ScienceDirect provided by:




Your Library

Explore

- Related articles
- Your applications (3)

Sponsored links



The Art of Leadership
Self Direction
September, February, 12, 2011

Additional article content and features

- Article information
- Journal information
- Author information (5)
- Related articles (6)
- Tables (15)
- Videos (9)
- Data analysis (7)
- References (141)
- Highlight links

INFLATED JOURNAL VALUE RANKINGS: PITFALLS YOU NOW KNOW ABOUT HTML AND PDF USAGE

Acknowledgements:

Research partners:

*Nicole Contaxis, MLIS 2015, School of Education and Information Studies, UCLA

*Alex Wood-Doughty, Ph.D. in progress, School of Economics, UC-Santa Barbara

*Significant earlier research on usage data problems by Terry Bucknell, Philip M Davis, and Jason Price

Chan.Li@ucop.edu

Jacqueline.Wilson@ucop.edu