The Problem

- Use statistics provide a way for libraries to assess the value of electronic journal subscriptions, but what types of use should be examined and how can use statistics be quickly and easily collected?
  1. From the vendor/platform?
  2. From the link-resolver?
  3. From citations of research publications?
Vendor/Platform Statistics

- Most publishers/vendors are COUNTER-compliant (Counting Online Usage of NeTworked Electronic Resources)

- Advantages:
  - COUNTER allows for standardized reporting of use across different publishers and platforms.
  - Successful Full-Text Article Requests (SFTARs) provide an article-level metric of use

- Disadvantages
  - Statistics often need to be collected from many different vendors in an idiosyncratic manner
  - For journals available from multiple vendors, statistics must be tabulated and merged to get a complete SFTAR count
  - The platform design can affect usage counts so that there is inconsistent reporting
  - Standardized Usage Statistics Harvesting Initiative (SUSHI) should help with collection, but implementation has not been consistent

Link-Resolver Statistics

- Count the number of click-throughs to a journal title from a link-resolver service, including click-throughs from A-Z lists and MARC records

- Advantages
  - Easy to collect from single source
  - Use on different platforms is brought together and managed in knowledge base

- Disadvantages
  - Does not measure actual article use
  - Once user leaves link resolver, additional use is no longer counted
  - A lot of use may occur through alternate paths
### Local Citation Data

- Reflects how many times a journal is cited by researchers at a specific institution
- **Advantages**
  - Use demonstrates clear value of content
- **Disadvantages**
  - Time-consuming to collect
  - Other uses, such as consultation, clinical use, student reference, etc., are not represented

### The Questions

- Do link-resolver statistics positively correlate with COUNTER-compliant vendor SFTAR statistics?
- Do link-resolver statistics reflect use patterns seen in vendor statistics?
- Does local citation analysis demonstrate different patterns not reflected in vendor or link-resolver statistics?
Methodology: Data Collection (1)

- Study university (UIC) is large urban Research 1 university with 6 health sciences colleges, a large urban medical center, and 3 regional medical campuses.
- Journal list compiled from ERMS, Serials Solutions
  - Used HILCC (Hierarchical Interface to Library of Congress Classification) to identify journals in the health sciences
  - Identified list of 3496 current titles
  - List narrowed to 2619 after removing titles missing use data
- Collected vendor COUNTER-compliant SFTARs for 2010
  - Went to individual website and downloaded Excel reports from over 20 providers
  - All were COUNTER-compliant except MD Consult

Methodology: Data Collection (2)

- Obtained link-resolver data for titles for 2010
  - Click-through counts through the OpenURL resolver
  - Click-through counts through the e-journal A-Z list
  - Click-through counts through the library catalog MARC records
- Collected citation data for 2010
  - Searched ISI's Web of Science by author affiliation
  - Although study focused on health sciences, citations for all researchers affiliated with the university were included
Matching and Cleaning the Data

- Matched link-resolver data to citation data
  - This was easy: 1 line of data per journal
- Matched link-resolver and citation data to COUNTER SFTARs
  - This was the hard part
- Challenges
  - A title might be listed in up to 10 different databases in ERMS
  - Automated matching of data by ISSN, but titles still needed to be examined by hand
  - Matching process was not 1 to 1
  - Some platforms lacked vendor statistics (e.g., open-access)
  - Some databases were listed multiple times in ERMS, but only once in platform statistics (e.g., one title could be in multiple EBSCOhost databases)
  - Errors in knowledge base or changes in providers could lead to missed matches
  - Unmatched data need to be manually searched by title

Title coding scheme

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Decision</th>
<th>Total Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>Each source in ERMS matches corresponding source for provider statistics</td>
<td>Used in analysis</td>
<td>1853</td>
</tr>
<tr>
<td>Perfect Match Plus Open Access</td>
<td>Each source in ERMS matches corresponding source for provider statistics AND title also available on OA platform</td>
<td>Used in analysis</td>
<td>392</td>
</tr>
<tr>
<td>Open Access Only</td>
<td>No vendor statistics available because only on OA platforms</td>
<td>Title removed</td>
<td>467</td>
</tr>
<tr>
<td>Perfect Match + Extra Data</td>
<td>Each source in ERMS matches corresponding source for provider statistics AND received additional provider statistics with no match in ERMS</td>
<td>Used in analysis</td>
<td>311</td>
</tr>
<tr>
<td>Perfect Match + Extra Data + OA</td>
<td>Each source in ERMS matches corresponding source for provider statistics AND title also available on OA platform AND received additional statistics with no match in ERMS</td>
<td>Used in analysis</td>
<td>63</td>
</tr>
<tr>
<td>Missing Data</td>
<td>Missing data in one or more sources in ERMS that was not found through individual title searching</td>
<td>Title removed</td>
<td>408</td>
</tr>
</tbody>
</table>

3494
Statistical Analysis

- Spearman rank order
  - Looks only at the rank of use between the three use factors and does not take into consideration the actual number of uses
  - Best for non-parametric data

Results - Descriptive Statistics for the Three Datasets

<table>
<thead>
<tr>
<th></th>
<th>Vendor COUNTER Successful Full-Text Article Requests</th>
<th>Link-resolver Click – Through Statistics</th>
<th>Local Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counts at the 25th Percentile mark of a ranked title list (663 titles)</td>
<td>68</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Median of all data</td>
<td>283</td>
<td>81</td>
<td>3</td>
</tr>
<tr>
<td>Counts at the 75th Percentile mark of a ranked title list (1302 titles)</td>
<td>757</td>
<td>215</td>
<td>13</td>
</tr>
<tr>
<td>Highest Count for Each Measure</td>
<td>71,326</td>
<td>11,761</td>
<td>1,784</td>
</tr>
</tbody>
</table>

N=2,619
Number of Titles Per Usage Quartile

Results – Correlation Coefficients

<table>
<thead>
<tr>
<th>Data Subsets</th>
<th>Link Resolver/ Vendor</th>
<th>Link Resolver/ Citation</th>
<th>Vendor/ Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Journals</td>
<td>2619</td>
<td>.843</td>
<td>.752</td>
</tr>
<tr>
<td>1-25th Percentile of titles (Link-resolver use ≤26)</td>
<td>663</td>
<td>.454</td>
<td>.392</td>
</tr>
<tr>
<td>25th-75th Percentile of titles (Link-resolver use ≥27-≤214)</td>
<td>1302</td>
<td>.563</td>
<td>.408</td>
</tr>
<tr>
<td>75th -100th Percentile of titles (Link-resolver use ≥215)</td>
<td>654</td>
<td>.586</td>
<td>.511</td>
</tr>
<tr>
<td>25th -100th Percentile of titles (Link-resolver use ≥27)</td>
<td>1956</td>
<td>.703</td>
<td>.581</td>
</tr>
</tbody>
</table>

All correlations were significant at p<.01.
Sample of titles with vendor use higher than ERMS, ERMS higher than citations

<table>
<thead>
<tr>
<th>Title</th>
<th>Link-resolver</th>
<th>Vendor</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Emergency Medicine</td>
<td>795</td>
<td>992</td>
<td>37</td>
</tr>
<tr>
<td>Annals of Internal Medicine</td>
<td>4,161</td>
<td>7,581</td>
<td>155</td>
</tr>
<tr>
<td>Circulation</td>
<td>2,545</td>
<td>14,134</td>
<td>472</td>
</tr>
<tr>
<td>Develop. &amp; Comparative Immunology</td>
<td>12</td>
<td>4,069</td>
<td>0</td>
</tr>
<tr>
<td>Diabetes Care</td>
<td>1905</td>
<td>730</td>
<td>190</td>
</tr>
<tr>
<td>Journal of Biological Chemistry</td>
<td>2,280</td>
<td>45,526</td>
<td>1,784</td>
</tr>
<tr>
<td>Journal of Family Practice</td>
<td>808</td>
<td>1,365</td>
<td>8</td>
</tr>
<tr>
<td>Nature</td>
<td>4,420</td>
<td>41,428</td>
<td>1,015</td>
</tr>
<tr>
<td>PNAS</td>
<td>2,830</td>
<td>35,944</td>
<td>1,372</td>
</tr>
<tr>
<td>Science</td>
<td>5,706</td>
<td>40,155</td>
<td>997</td>
</tr>
</tbody>
</table>
Sample of titles with anomalous results

<table>
<thead>
<tr>
<th>Title</th>
<th>Link-resolver</th>
<th>Vendor</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amer. Journal of Clinical Nutrition</td>
<td>1,295</td>
<td>0</td>
<td>140</td>
</tr>
<tr>
<td>Biochemistry (Easton)</td>
<td>935</td>
<td>0</td>
<td>364</td>
</tr>
<tr>
<td>Infection and Immunity</td>
<td>142</td>
<td>0</td>
<td>122</td>
</tr>
<tr>
<td>Internet J. of Peds. &amp; Neonatology</td>
<td>0</td>
<td>301</td>
<td>0</td>
</tr>
<tr>
<td>J. For Specialists in Ped. Nursing</td>
<td>0</td>
<td>799</td>
<td>0</td>
</tr>
<tr>
<td>Critical Care</td>
<td>665</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Diabetes Care</td>
<td>1,905</td>
<td>730</td>
<td>190</td>
</tr>
<tr>
<td>Journal of Studies on Alcohol</td>
<td>520</td>
<td>496</td>
<td>38</td>
</tr>
<tr>
<td>British Journal of Ophthalmology</td>
<td>508</td>
<td>512</td>
<td>35</td>
</tr>
<tr>
<td>Gerontologist</td>
<td>590</td>
<td>608</td>
<td>26</td>
</tr>
</tbody>
</table>

Anomalous Use Patterns

- Anomalous use patterns can point to problems with links or vendor use statistics accounts
- Anomalies can easily be discovered using Excel by ratios testing or subtraction to discover where the data for a title show unexpected results
“80/20” Calculations

- Link-resolver Data: 80/29
  - 0.8% of titles had no click-throughs
- Vendor Data: 80/24
  - 4.5% of titles had no SFTARs
- Citation Data: 80/17
  - 27% of titles received no citations in 2010
- Use is the most concentrated for citations

Conclusions

- Link-resolver data correlates with vendor use statistics, and thus can aid with collection development decisions, especially if vendor data cannot be obtained.

- Link-resolver data can be used to identify journals that need further evaluation for retention decisions. If a journal has a high-link resolver count, no further evaluation is necessary.

- Citation data, which identifies titles used in faculty research, displays a similar pattern of use compared with other measures.
Areas for further study

- Repeat the study in the humanities, social sciences, and basic sciences to see if use patterns differ.

- ERMS and vendor data were not always proportional in use, the ratio of ERMS to vendor data varied widely, even within expected parameters. Is that effect related to the platform the journal is on? The type of journal? Or is it related to user behavior?

Next Steps at UIC

- UIC subscribed to 360 COUNTER to assist with collecting and merging vendor statistics
  - First year collected was 2012
  - Still clean-up/matching issues for the subscribing library to address
  - Data out is only as good as the data in
  - Need to evaluate SUSHI option

- Evaluate options to calculate cost-per-use
Sample JR1 report from 360 COUNTER

Thank you!

Questions?
Sandy De Groote sgroote@uic.edu
Deb Blecic dblecic@uic.edu
Kristin Martin kmarti@uchicago.edu