Steps for Using the Worksheets for Evaluating Link Resolvers

Prepared by
Karen R. Harker, MLS, MPH
Collection Assessment Librarian

Being and Librarianship
University of North Texas Libraries
1155 Union Circle, #305190
Denton, TX 76203
940-565-4688

Simple Randomized Cluster Sampling
1. Select the cluster of ejournal titles
2. Select the articles in the sources from each title selected

Steps:
1. Determine the Sample Size using the Sample Size Calculator worksheet:

<table>
<thead>
<tr>
<th>Sample Size Calculator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instructions:</strong> Complete the fields in green with the necessary information. Then read through the steps below to attain your final sample size.</td>
</tr>
<tr>
<td><strong>Level of Confidence Measure</strong></td>
</tr>
<tr>
<td><strong>Margin of Error (MOE)</strong></td>
</tr>
<tr>
<td><strong>Expected Success Rate</strong></td>
</tr>
<tr>
<td><strong>Design effect (Deff)</strong></td>
</tr>
<tr>
<td><strong>Expected Response Rate</strong></td>
</tr>
<tr>
<td><strong>Number of Databases or Sources</strong></td>
</tr>
</tbody>
</table>

Adapted from World Health Organization: [http://www.who.int/chp/steps/resources/sampling/en/index.html](http://www.who.int/chp/steps/resources/sampling/en/index.html)

a. Read each variable and make any changes, as is necessary. The most common changes will be:
   i. **Expected Success Rate** - this is the estimated success rate of getting to the full-text directly through the link resolver. This could be considered the "minimally tolerable rate" or it could be a desired rate of success.
   ii. **Expected Response Rate** - this is the rate of titles that will be found in each database searched. This should be an average across all databases.
   iii. **Number of Databases or Sources** - this is the number of databases you plan on testing.

b. The "Final Sample Size" should be rounded to a workable number.

2. Select the journal titles to search
   a. Download list of all ejournals that you want to study.
      i. Recommendation: download only those that have been **used at least once** in last twelve months. This will ensure that your sampling frame is of those titles that are used by your clients.
      ii. The fields should include:
         1) Journal Title
         2) ISSN
         3) Publisher or Source of full-text access (if this is of interest to you)
   b. Copy or import this data into the **Ejournals** worksheet.
   b. Remove duplicates
      i. Data->Remove Duplicates->Select Type, Title & ISSN (or other fields to match on, as appropriate).
iii. Copy or import this data into the **EJournals** worksheet.

b. Remove duplicates
   i. Data->Remove Duplicates->Select Type, Title & ISSN (or other fields to match on, as appropriate).

   c. Remove titles for which you cannot search (e.g. non-Latin characters)
   d. Insert a column at the far-left and label the column, **ID**.
   e. Populate the **ID** cell for first 10 rows with the numbers 1...10. This will seed the trend for the remaining numbers.
   f. Fill the remaining cells:
      i. Highlight all the cells in that column to the last row.
      ii. Select Fill->Series

   g. Take a random sample using Excel’s Sampling method
      i. [Here are the detailed instructions](#)
      ii. Requires the Data Analysis Pack ([instructions from Microsoft](#))
      iii. Highlight the **ID** cells for all the populated rows.
      iv. Select Data ribbon tab -> Data analysis -> Sampling
v. Refer to the **Sample Size** calculation for the **Number of Samples** value.
vii. In the **Sample Titles** worksheet, copy cells C2 and D2 and paste in the remaining rows. This grabs the journal title & ISSN of the selected titles.

ix. Copy the table in Sample Titles worksheet and **paste as values**. This will effectively fix the selected sample.

h. Populate the data worksheets (**Data for DB1...**) for each resource:

   i. Copy the cell A2 (DB) in the **Sample Titles** worksheet and paste to the last row of titles. This randomly assigns each title a number between 1 and the total number of databases identified in the Sample Size Calculator worksheet. This randomly divides the titles into the different groups.
   1) Check the randomization by hitting the “Refresh” button next to the **Random Sampling Table**.
   2) Each time you push the “Refresh” button, the assignment of the numbers changes.
   3) Hit “Refresh” until all of the groups are pretty much evenly distributed.

   ii. Highlight the A column, copy, and paste **as values in the same place**. This will fix the random assignment.

   iii. Populate the “**Data for DB...**” worksheets:
   1) Filter the **Sample Titles** worksheet by the “DB” column to show only where DB=1.
   2) Copy the cells in the B:D (highlighted in blue) for one DB number at a time (e.g. all the “1’s”).
   3) Paste these in starting at cell A3 in the appropriate worksheet for that database (e.g. “**Data for DB1**” for the “1’s”).
   4) Repeat this paste on each of the remaining “**Data for DB...**” worksheets.
   5) Make as many worksheets as are the number of resources included in your study. Delete any extraneous worksheets.
   6) This will ensure that the summary calculations are preserved for each database.

   i. Adjust the Summary worksheet based on the number of databases, adding or deleting rows, as appropriate.

3. Collect data
   a. For each database or source:
      i. For each title selected:
         1) Search for articles from that title within the subscribed date range.
         2) On the spreadsheet for that database, enter the total number of articles found for that title.
         3) If no articles were found, skip to the next title.
         4) A special problem occurs when the database provides the full-text article AND there is no button or link for the link-resolver. This can occur within EBSCO. In this case, enter “0” for Number of Articles Found and skip to the next title.
         5) If articles were found:
            a) Sort the list by author last name
            b) Enter the total number of articles found in the **Random Article Selector**.
            c) Note the random number generated.
            d) Return to the list and go directly to that record number that matches the random number generated.
            e) Click on your link-resolver button/link.
            f) Note the success of that link:
               i) Enter “Yes” in the column that best represents the result.
      ii. Repeat for each journal title in the sample.
   b. Repeat for each database or source in your study.

4. Summarize the data
   a. The Summary sheet provides the basic summaries of the data.
   b. It counts the number of “Yes” responses in each of the result options.
   c. There are also ratios using the total number of titles found in each database as the denominator.
   d. Totals, averages and standard deviations are calculated for each database.
5. Deriving meaning from the data
   a. There are tests of the count of "Any Full-Text" items (# of citations that the Find Full-Text link goes directly to the full-text or directly to the correct abstract of the article). These are located under the Summary Table. This data is tested against two values:
      i. The number of citations that would be expected to go to full-text. This is calculated by multiplying the # of titles found in each databases by the expected success rate that you entered on the Sample Calculator worksheet. This is used to determine if there is a significant difference between the overall success rate and the expected rate.
      ii. The mean number of successful links. This is used to determine if there is a significant difference between all of the databases.
   b. 2x2 Tables
      i. These enable you to compare any two resources, or to compare one resource against the average.
      ii. The data used here are derived from the Summary table.
      iii. To use, enter the row number on the Summary table that is for the database of interest.
      iv. To compare against the average, use the row number for the Means row in the Summary table.
      v. Chi-square tests are automatically computed for each table with significance stated based on the Confidence Level set in the Sample Size Calculator worksheet.
   c. View the Dashboard
      i. These are graphs that show the results of the charts on the Summary worksheet.
      ii. These can be copied & pasted into presentations for your reports.
6. Variations
   a. Use the entire ejournals list as the sampling frame - this is not recommended because of the tremendous number of items that are not indexed in multiple sources. But if your access to ejournals is limited and largely through direct subscriptions (versus full-text databases), this may be appropriate. Besides the source of your ejournals, no other changes to the process is necessary.
   b. Testing different link resolvers, if you're able to have a trial of two different link resolvers, which one is more successful.
      1) This would require testing titles from fewer sources but effectively doing it twice.
      2) This analysis is essentially a "matched-case", which tests the rate of discord between the two resolvers.
   c. Comparing different publishers:
      1) Export your list of titles for each full-text source separately
      2) Choose fewer sources to compare.