

Introduction to Quantitative Research and Data

The foundation of any e-book analysis framework rests on knowledge of the general e-book landscape and the existing information needs of a local user community. From this starting point, quantitative methods, such as cost analysis, can provide evidence for collection development initiatives and demonstrate how they align with patrons' needs and the overarching goals of library administrators or funding agencies.

Essentially, “data stands in place of reality we wish to study. We cannot simply know a phenomenon, but we can attempt to capture it as data which represents the reality we have experienced . . . and are trying to explain.”¹ The data collected through quantitative investigations provides a baseline for future evaluation, evidence for when and how patrons make use of electronic collections, and promotes data-driven decisions throughout collection development departments. To get the most mileage out of the time and resources invested into quantitative investigations, it is essential to first understand what quantitative research is and what types of questions it can answer.

What Is Quantitative Research?

In the most basic terms, quantitative research methods are concerned with collecting and analyzing data that is structured and can be represented numerically.² One of the central goals is to build accurate and reliable measurements that allow for statistical analysis.

Because quantitative research focuses on data that can be measured, it is very effective at answering the “what” or “how” of a given situation. Questions are direct, quantifiable, and often contain phrases such as

what percentage? what proportion? to what extent? how many? how much?

Quantitative research allows librarians to learn more about the demographics of a population, measure how many patrons use a service or product, examine attitudes and behaviors, document trends, or explain what is known anecdotally. Measurements like frequencies (i.e., counts), percentages, proportions, and relationships provide means to quantify and provide evidence for the variables listed above.

Findings generated from quantitative research uncover behaviors and trends. However, it is important to note that they do not provide insight into *why* people think, feel, or act in certain ways. In other words, quantitative research highlights trends across data sets or study groups, but not the motivation behind observed behaviors. To fill in these knowledge gaps, qualitative studies like focus groups, interviews, or open-ended survey questions are effective.

Whenever I sit down to a new quantitative research project and begin to think about my goals and objectives, I like to keep a small cheat sheet on my desk to remind me of the trends quantitative data can uncover and the stories that I can tell with study conclusions. This serves as one quick strategy that keeps my thoughts focused and prevents scope creep as I discuss project plans with various stakeholders.

Quantitative Research Cheat Sheet

Six key characteristics of quantitative research:

- It deals with numbers to assess information.
- Data can be measured and quantified.
- It aims to be objective.
- Findings can be evaluated using statistical analysis.

- It represents complex problems through variables.
- Results can be summarized, compared, or generalized.

Quantitative findings can provide evidence or answers in the following areas:

- Demonstrate to what extent services and collection are used and accessed.
- Back up claims about use and impact.
- Provide evidence for how the budget is spent and whether adjustments should be made.
- Demonstrate return on investment when presenting budget figures.
- Inform decisions regarding packages and subscriptions that are or are not worth pursuing.
- Demonstrate evidence for trends and prove or discount what is known anecdotally.
- Provide a method to make information accessible to audiences.
- Provide evidence of success and highlight areas where unmet information needs exist.

Main advantages of quantitative research:

- Findings can be generalized to a specific population.
- Data sets are large, and findings are representative of a population.
- Documentation regarding the research framework and methods can be shared and replicated.
- Standardized approaches permit the study to be replicated over time.

Main limitations of quantitative research:

- Data does not provide evidence for *why* populations think, feel, or act in certain ways.
- Specific demographic groups, particularly vulnerable or disadvantaged groups, may be difficult to reach.
- Studies can be time consuming and require data collection over long periods of time.³

Quantitative Research in Information Management Environments

In the current information landscape, a wealth of quantitative data sources is available to librarians. One of the challenges surrounding quantitative research in the information management profession is “how to make sense of all these data sources and use them in a way that supports effective decision-making.”⁴

Most libraries pay for and receive materials through multiple routes. As a result, a quantitative

research framework for e-book collections often consist of two central components: an examination of resource allocations and expenditures from funds, endowments, or gifts; and an examination of titles received through firm orders, subscriptions, packages, and large aggregated databases.⁵ In many cases, examining funds and titles according to subject areas adds an extra layer of knowledge that can provide evidence for teaching, learning, or research activities in a specific field or justify requests for budget increases.⁶

Many of the quantitative research projects that I have conducted over the past four years are in direct response to an inquiry from library administrators. In most cases, I have been asked to provide evidence for collection development activities that support expressed information needs, justify expenditures, or project annual increases in preparation for a new fiscal year. Study results are often expected to describe or weigh several courses of action in the short and long term. Essentially, my work is categorized into three basic concepts related to library management:

- Distinguish between recurrent and capital expenditure and projects, and between past, present, and future states.
- Accommodate priorities and determine how resources are spread across collections.
- Indicate the ways of allocating resources at input, monitor performance, and assess performance at output.⁷

To assist in my prep work for a quantitative research project, I put together a file of background information about my library system and local user community to ensure that the project supports institutional goals and aligns with the general direction of programs and services on campus. Below are seven categories of information that I have on file at all times:

- the institutional identity of the library
- the stakeholder groups to be served
- collection resources
- financial resources
- library personnel
- facilities and equipment
- the various programs and services related to the quantitative investigation⁸

Typically, I take a day or two at the beginning of each fiscal year to update this information and ensure that it accurately reflects the landscape of collections and services available at CUL. From this starting point, it is simple to look at new project descriptions and think about the data required to support high-level decisions regarding the allocation of resources,

to assess the effectiveness of collections and services, or to measure the value and impact of collections.

A wealth of local and external data sources is available to librarians, and each one can be used to tell a story about collection size, value, and impact. All that is required is an understanding of what the data measures and how different sources can be combined to tell a story about a user community.

Definitions of Local and External Data Sources

The remaining sections of this issue of *Library Technology Reports* discuss how I use quantitative data, what evidence I have uncovered to support e-book collection decisions, and how I apply quantitative findings in practical library settings. For the purposes of these discussions, I will use the following terminology:

Bibliographic record: A library catalog record that represents a specific title or resource.

Catalog clickthroughs: Counts of patron use of the catalog to access electronic full texts.

Citation analysis: Measurement of the impact of an article based on the number of times it has been cited.

Consortia reports: Consolidated usage reports for consortia. Often used to view usage linked to each individual consortia member.

COUNTER (Counting Online Usage of Networked Electronic Resources): An international initiative to improve the reliability of online usage statistics by providing a Code of Practice that standardizes the collection of usage data. It works to ensure vendor usage data is credible and comparable.

Cost data: Factual information concerning the cost of library materials, annual budget allocations, and general acquisitions budget.

FTE (full-time equivalent): The number of full-time faculty and students working or studying at a specific institution.

IP (Internet Protocol) address: A numerical label usually assigned to a library router or firewall that provides access to a private network (e.g., school or library network).

Link resolver statistics: Information regarding the pathways users take to access electronic resources.

Overlap data: Measurement of the degree of duplication across a collection.

Publication analysis: Measurement of impact by counting the research output of an author. Metrics include the number of peer-reviewed articles, coauthor collaborations, publication patterns, and extent of interdisciplinary research.

Title lists: Lists of e-book titles available in subscriptions, databases, or packages. These lists are generated and maintained by vendors and publishers.

Turnaway statistics: The number of patrons denied access to a specific title.

Vendor use data: Electronic use statistics provided by vendors.

Indicators and Performance Measures That Support Quantitative Research

I regularly use several indicators and performance measures to analyze e-book collections. Local and external data sources (listed in the section above) inform these investigations and provide the necessary “ingredients” to conduct cost analysis, examine return on investment, or measure the value of e-book collections to the community at CUL. Below is a breakdown of how I classify data and relate it to different indicators.⁹

Input Cost Measures

Data source: Cost data pulled from Voyager reports (or your institution’s ILS system).

In general, cost data demonstrates how funds are allocated across a budget. Analysis can identify areas where additional resources are required, monitor cost changes over time, and flag collection areas where funds can be pulled (e.g., overbudgeted funds, subject areas that no longer support the curriculum, etc.) and “reinvested” in the collection to support current information needs.

Each of the investigations described in the following chapter began with a review of cost data. I relied on a basic knowledge of how e-book acquisition budgets are distributed across subject areas or pooled to purchase interdisciplinary materials. Essentially, these investigations involved the identification of fund codes linked to subject areas, expenditures across set date ranges (e.g., calendar years, fiscal years, academic years), and bulk versus long-tail purchases.

Tip: When working with cost data and examining input cost measures, I have found it helpful to categorize data by fund type. E-book collections at CUL are often built with general income (GI) funds, endowments, and gifts. Policies and procedures regarding how funds can be transferred and what materials can be purchased impact how resources are allocated to build e-book collections. Before beginning a cost analysis project at your institution, it may be helpful to review the policies in place and determine how they relate to overarching institutional goals and collection priorities.

Collection Output Measures

Data sources: Cost data, title lists, overlap data, bibliographic records (particularly subject headings).

Collection output measures are related to the quantity and quality of output. Examples include the number of e-book titles included in a subscription or package deal acquired by a library, the number of e-book records acquired over a given period of time, the number of publishers and unique subject areas represented in an e-book collection, the currency of information (e.g., publication year), and the percentage of title overlap, or duplication, within a collection.

At this stage in my cost analysis projects, it is often necessary to combine data to create a snapshot of how funds flow in and out of subject areas to acquire research and teaching materials. For example, many of our large e-book packages are interdisciplinary. By pulling cost data, I can determine how the total cost was split across subject divisions based on fund code counts. Then, I break title lists apart by subject to determine what percentage of total content relates to each library division. By comparing the cost breakdown and title list breakdown, it is possible to determine what percentage of total content each library division receives and if it is on par with the division's financial contribution.

Effectiveness Measures and Indicators

Data sources: Cost data, title lists, COUNTER reports, vendor reports, consortia reports, resolver statistics, turnaway statistics, Google Analytics.

Examining input and output measures is an effective way of determining how budgets are allocated and the quantity and quality of materials available to patrons. To develop a quantitative baseline for the general value of e-book collections, measures like rate of use, cost per use, and turnaway rates can be very effective.

Again, this form of analysis relies on data from multiple sources. The ability to combine cost data, title lists, and COUNTER data (or vendor data) has yielded actionable results at my library. For instance, I combine data from these three sources to measure the value of databases. By pulling cost data covering three fiscal years and matching title lists against COUNTER reports, I have been able to examine trends in annual increase rates, examine overlap between subscriptions in the same subject area, and calculate cost per use to determine what percentage of the user community makes use of subscriptions.

Finally, by looking at turnaway statistics (also found in COUNTER data), it is possible to determine if sufficient access is provided to users. For instance, I look at turnaway statistics to evaluate if e-books listed on course reading lists provide sufficient access to a class of students over a semester. In cases where access is limited to a single user, I may look at the budget to find areas where funds can be shifted to

purchase simultaneous usage instead.

Together, the data sets mentioned above provide evidence for how funds are invested, if they are invested in materials that are heavily used by patrons, and if access models are suited to the needs of the local user community.

In some cases, particularly when dealing with foreign language materials, I have encountered challenges because COUNTER data is not provided, and in some cases, it is difficult to obtain vendor reports as well. In the absence of usage data, I have experimented with link resolver statistics to determine what information they provide about user activities and the value of e-book materials.

Link resolver statistics provide information about the pathways users take to access electronic resources.¹⁰ Resolver statistics show that a patron made a “request” via the link resolver and started the process of trying to view a full text. If the patron successfully accesses the full text, this is counted as a “clickthrough.”

It is important to note that link resolver statistics and usage statistics (like COUNTER) are not comparable because they measure different activities. Link resolvers measure attempts to connect while usage data measures usage activity. However, comparing sets of link resolver statistics against each other may provide insight into which resources patrons attempt to access most frequently. This can provide a ballpark idea of resource value in cases where usage statistics are not available.

Domain Measures

Data sources: FTE (full-time equivalent), IP address, demographic information.

Domain measures relate to the user community served by a library. They include total population, demographic information, attributes (e.g., undergraduate level, graduate level), and information needs.

In my work, domain measures impact subscription or package costs because campus-wide access is often priced according to FTE. Due to the size of CUL's student body, access to essential collections can become extremely expensive and fall outside of the budget range. When this occurs, examining patron access by IP address has opened the door to negotiation, particularly when dealing with content that is discipline-specific. For instance, when negotiating subscription prices for science materials, IP data provided evidence that usage is concentrated at the library router located in the Science and Engineering Library. This allowed science selectors to negotiate pricing models based around the FTE of natural science programs as opposed to the campus community as a whole.

Cost-Effectiveness Indicators

Data sources: COUNTER reports, vendor reports, turnaway statistics, citation analysis, publication analysis.

Cost-effectiveness indicators are related to measures like cost per use and ultimately examine the general return on investment. They evaluate the financial resources invested in a product and determine if the investment brings added value to the existing collection.

In my work, I often combine cost data with usage data to calculate cost per use and also capture usage trends spanning at least three calendar years. The results provide a benchmark regarding whether the financial investment in the product is equivalent to its general “demand” within the user community. A recent project with colleagues at the science and medical science libraries has examined how to use citation and publication data to determine general impact of electronic resources.

Challenges Presented by Quantitative Research

One of the challenges surrounding quantitative research in library environments is a lack of standardization across data sets, particularly vendor reports. The general situation has improved in recent years due to widespread compliance with the COUNTER Code of Practice, but there is still work to be done. It is difficult to interpret the meaning of vendor usage data that is still not COUNTER-compliant because clear definitions of *use* do not exist. This can create significant roadblocks when running quantitative projects that examine multiple e-book collections to get a sense of comparative value.

Also, usage data is generated outside of libraries by publishers or aggregators and vendors. Factors like turnover, company mergers, or password changes result in significant time lags between when usage statistics are generated and when libraries receive them. Also, some vendors pull down usage statistics after a period of months. In most cases, librarians need statistics captured over two or three years to meet reporting requirements, and data dating back this far can be difficult to obtain. Finally, annual usage statistics are provided according to calendar year. However, librarians look at usage by fiscal year and academic year as well. In many cases, this means that multiple usage reports have to be stitched together in order to capture the appropriate timeframe for reporting purposes. This process is labor intensive and takes a considerable amount of time to complete.

These challenges emphasize an ongoing need to build positive working relationships with publishers, aggregators, and vendors to discuss challenges and

develop solutions that benefit all stakeholders. It is important to note that libraries have valuable information that is not available to content providers, namely how e-books are discovered and used. Strong relationships allow for the transparent exchange of information between all parties, which ultimately benefits patrons by providing a seamless e-book experience.

Designing a Quantitative Research Framework

As mentioned earlier in this chapter, data stands in place of a reality we wish to study, quantify, and explain. In order to prevent scope creep and pull together bodies of data that add value to local work environments, it is essential to begin any quantitative research project with a set of clearly defined objectives, a strong understanding of the stakeholder group or audience, and knowledge of local information needs. These bits of information serve as markers to measure progress and ensure the project stays on track.

It is tempting to dive straight into a project and investigate if anecdotal information or assumptions are correct, but time spent developing a project outline is never wasted. The development of a successful plan requires “a clear idea of what it is to be achieved among the stakeholders. Clearly articulated objectives are the engine that drives the assessment process. This is one of the most difficult but most rewarding stages of the assessment process.”¹¹ Creating a roadmap for research projects can save countless hours down the line and ensures the correct quantitative method is selected. The plan also provides focus when the analysis phase of a project begins. Keep in mind that the data set you end up working with will be large; approaching it with stated goals and objectives saves significant amounts of time, which is especially important when working under a tight deadline!

Below is a checklist that I use at the beginning of any research project. It is based on recommendations made by Bakkalbasi, Sundre, and Fulcher.¹²

1. Project goals, objectives, and desired results
While goals and objectives are closely related, they are not the same. Project goals should state exactly what you hope to learn or demonstrate through your research. Objectives state what you will assess or measure in order to achieve your overarching project goal.
Example of a project goal:
 - Gain insight into how local patrons use the library to support teaching and learning needs.Example of project objectives:
 - To learn what activities local patrons engage in when using library facilities.

- To assess the degree to which patrons make use of subscribed e-book content.
 - Consider how results may support improvement of collection development initiatives or lead to evaluation of existing workflows, policies, and procedures.
2. List of key stakeholders
 - What questions and/or evidence are required by stakeholders?
 - What information do stakeholders require to make decisions?
 - How will results support the improvement of collection development initiatives?
 - How will results be made accessible to stakeholders?
 - Are the results intended for internal use, or will they be shared with the professional community?
 - Will findings be used to support grant or funding applications?
 3. Project timeline
 - Is there a stated project deadline? What methods or resources will allow you to collect data, conduct analysis, and provide findings within the stated timeframe?
 - Does the project coincide with other activities that may require your attention (e.g., fiscal year, subscription renewal period)?
 - Are there colleagues at the library who may be able to provide assistance given the timeline of the project?
 4. Confidentiality
 - What data collected through the study cannot be shared with external stakeholders (e.g., cost data, FOIP compliance, etc.)?
 - Are there any permissions required before study results can be disseminated to external stakeholders?
 - Is clearance required to collect data from a user community?
 5. Data collection process
 - What data sources are most valued and meaningful to your library?
 - What data sources will allow results to be applied at your library?
 - What data collection methods will be most effective?
 - What data collection methods will provide valid and reliable results?
 - Are there parameters such as specific fiscal years, calendar years, or academic years that you are required to report on?
 6. Data analysis¹³
 - How will data be summarized and described?
 - What features of the data set are most relevant to project objectives and goals?

- What are the relationships between different data sets?
7. Presentation of results
 - How is data evaluated?
 - How is data interpreted into meaningful results and conclusions?
 - What are the recommendations for action or improvements?
 - How will findings be communicated to stakeholders?

The data sets collected through quantitative methods are large and can easily be examined from a variety of perspectives. As the project develops, mentally frame emerging trends into a story that can be shared with stakeholders. This process determines how results will ultimately be applied to collection development initiatives. Background knowledge of the local patron community and institutional goals serves as a compass; use it to shape results that bring value to your library or the greater professional community.

From my experience, each quantitative project that I work on allows me to expand my skill sets and understand how I can structure my daily activities to support overarching institutional goals. During many projects, I have encountered unexpected challenges or had to improvise when quantitative methods did not yield expected results (e.g., low survey response rates). However, each challenge equipped me to take on larger projects, better understand how our budget is structured, or build stronger relationships with patrons and colleagues.

One skill that has been invaluable to my work is the ability to develop a quantitative research plan. I hope that by sharing this structure, along with performance measures and data sources that I use, readers have a behind-the-scenes view of my process and all of the moving parts that I work with to conduct e-book collection analysis. And of course, now to the fun part! It is time to get down to the nitty-gritty and demonstrate how I conduct analysis to inform budget decisions and collection development activities at CUL.

Notes

1. Bob Matthews and Liz Ross, *Research Methods: A Practical Guide for the Social Sciences* (Harlow, UK: Pearson Education, 2010), 45.
2. *Ibid.*, 465.
3. Based on information provided by Stephen A. Roberts, *Financial and Cost Management for Libraries and Information Management Services* (London: Bowker-Saur, 1998), 140–41.
4. Darby Orcutt, *Library Data: Empowering Practice and Persuasion* (Santa Barbara, CA: Libraries Unlimited, 2009), 106.

5. Northwestern University Libraries, "DataBank: How to Interpret Your Data: Financial Support," LibGuide, last updated December 8 2015, <http://libguides.northwestern.edu/c.php?g=115065&p=748741>.
6. Ibid.
7. Roberts, *Financial and Cost Management*, 132.
8. Ibid.
9. For further information regarding indicators and performance measures, please see Roberts, *Financial and Cost Management*, 140–41.
10. Orcutt, *Library Data*, 107.
11. Nisa Bakkalbasi, Donna Sundre, and Kenton Fulcher, "Assessing Assessment: A Framework to Evaluate Assessment Practices and Progress for Library Collections and Services," in *Proceedings of the 2012 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment, October 29–31, 2012, Charlottesville, VA*, ed. Steve Hiller, Martha Kyrillidou, Angela Pappalardo, Jim Self, and Amy Yeager (Washington, DC: Association of Research Libraries, 2013), 538-545.
12. Ibid.
13. Based on information provided by Matthews and Ross, *Research Methods*, 345.