

Estimating Lost Volumes in a University Library Collection

This study employed standard sampling theory to make a study of library book losses, but unlike previously reported studies, the investigators instituted periodic searches for volumes missing after the original search. Over a period of two years and nine months, the original loss figures were cut by more than 60 percent.

With the assumption that the loss was related to the size of the collection each year, thus taking into account the rapid growth in recent years, a rough estimate of the annual loss rate was obtained. This figure was adjusted to reflect known losses discovered annually (identification of which resulted in routine purging the shelflist of holdings) yielding an adjusted estimated loss rate of about one-third of 1 percent annually.

FOR SOME TIME THE LIBRARY ADMINISTRATION HAS BEEN CONCERNED with the problem of missing books in the Washington University's central library (the John M. Olin Library). In the background of our concern there was the desire to consider alternative security arrangements. The present security arrangements in Olin Library include using a single exit, where an inspector visually checks all briefcases and bundles and verifies that all Olin Library books have been properly charged out (by inspecting a date stamped on a "due date" slip pasted inside the back flyleaf of each volume).

Our concern led in the fall of 1970

Powell Niland is professor of management, Graduate School of Business and Public Administration, and William H. Kurth is university librarian, Washington University, St. Louis, Missouri. The authors express their thanks to Carlos Alexander and Virginia Secrest for their assistance in conducting the research reported here.

to a systematic attempt to determine the book losses being incurred. No inventory had ever been made of Olin's collections; our attempt would thus give us some conception of the books lost since the start of the collection in the latter part of the nineteenth century. Given the obvious difficulty in making a complete inventory—at the time there were approximately 850,000 volumes—we decided upon a sample study.

The basic sampling technique used in this study is the same as that employed in two other recent studies by Bommer and Ford and Clark.^{1,2} Moreover, the Bommer and Ford study had the same impetus, the analysis of the value of an electronic security system.

One of the principal differences between our study and both the Bommer and Ford and Clark studies lies in the repeated searches for missing volumes incorporated into our study, in an effort to refine the estimate of missing volumes. The need for doing this was the

subject of a letter commenting on the Bommer and Ford article.³ Another is the use of a different assumption to break down a cumulative loss estimate into an annual loss rate. A third important difference is the fact that we extended our analysis to a second, and quite different, collection. The contrasting results we achieved in these two studies, together with the different magnitudes of the estimates generated by the Bommer and Ford and Clark studies, led to our trying to analyze factors accounting for differences in loss rates among various libraries.

DESIGN AND METHOD OF DRAWING SAMPLE

It was decided to determine the proportion of missing volumes with an accuracy of about ± 10 percent, and with a confidence level in the range of 90 percent to 95 percent. The "best guess" in advance of the study was that the percentage missing would be about 10 percent of total volumes. Using the standard formula for a 95 percent confidence level,⁴ a sample of about 3,600 was indicated. At a 90 percent confidence level, this was 2,400. So, we targeted a sample in between, about 3,000. These were chosen as three subsamples, each of about 1,000.

The shelflist rather than the card catalog was used to draw the sample, and the items were chosen by inches of material, rather than a fixed number per drawer. Use of the shelflist facilitated retrieval of information from both the stacks and the library's daily computer printout of volumes charged out to borrowers, because all of these were arranged in the same sequence. Furthermore, the problem of cross-reference cards (present in the card catalog) was eliminated, along with the problem of cards in the catalog that referred to items located in departmental libraries (rather than in Olin Library itself).

Using inches rather than a fixed number of cards per drawer to determine the number of items chosen eliminated the bias against drawers containing a large number of cards. Under a fixed number per drawer method, for instance, we would pick two cards from a drawer containing 400 cards, and two cards from a drawer containing 1,000 cards. At the same time, the tedium of counting every card, necessary if a fixed number of cards has been used, was avoided.

Furthermore, the samples were accumulated in lots of fifty, the researchers both compiling a list of fifty volumes and checking status information on the items in that lot the same day. All this "work unit" represented about four hours of work. This procedure enabled the data gatherers to appreciate the value of their efforts and also provided for some flexibility in making assignments of employees to this task.

In this fashion, three samples (A, B, and C) of approximately 1,000 cards each were drawn from the shelflist in January and February 1971. In each case, the stacks were checked to establish the physical presence or nonpresence of each item on the shelf. Then each missing item was checked against the Circulation Department's computer printout of books on loan. Volumes that were not found on the shelves and not listed as charged out to borrowers were classified as missing.

INITIAL SAMPLING RESULTS

The study data on March 1, 1971, are shown in Table 1.

TABLE 1
STUDY DATA AS OF MARCH 1, 1971

Sample	Number in Sample	Found	Not Found	Percent Missing
A	971	924	47	4.84
B	887	845	42	4.73
C	1091	1026	72	6.59
Total	2949	2795	161	5.45

Employing the same formula used above, we solved for the accuracy of the 5.45 percent figure and found that at the 95 percent confidence level, the upper limit for the percentage of books missing would be 6.3 percent and the lower limit 4.7 percent. That is to say, if our sample was truly unbiased, the odds were 95 in 100 that the proportion of books missing in the library will fall somewhere in the range of 4.7 percent to 6.3 percent, with 5.45 percent being the "best guess" figure.

The figure of 5.5 percent was much less than our original estimate of 10 percent. But the volumes in Olin by this time were approaching 900,000, so 5.5 percent of these—49,500—was a significant total. Assuming a cost to acquire and catalog of only \$15.00 per volume—certainly minimal—the missing volumes represented a loss of at least \$750,000.

ADDITIONAL SEARCHES

Were all of the missing volumes permanently lost? Would some "turn up" later on? We decided to find out and arranged to check on the missing items five times over the following three years. By January 1972, ten months after the original searches, nearly half the missing volumes had been found, and only eighty-six (2.91 percent) were still lost. And by January 1974 only sixty-five, or 2.20 percent, still could not be located. The effects of these searches, therefore, were to cut the missing volumes to about 60 percent of their original proportion. Applying our formula again after the last of these searches, at the 95 percent confidence level, the lower limit to the estimated proportion of books missing in the entire collection would be 1.7 percent, the upper limit, 2.7 percent, and the "best guess," 2.2 percent.⁵

Figure 1 plots this information for the total sample averages on a time axis. Attempts were made to fit various curves

to these data points, including the Gompertz, logistic, exponential, and second-degree exponential. No completely satisfactory fits could be obtained, however, because of difficulties in reconciling the first (March 1971) and second (July 1971) data points. Curves that gave good fits to the second and later points gave estimated values for the first point that were substantially below its actual value. On the other hand, curves that obtained close fits through the first and second data points yielded unsatisfactory fits for the latter data points, notably for April 1973 and January 1974. For instance, a logistic curve (of the form

$$\frac{1}{Y_t} = \frac{1}{K} + AB^t,$$

where Y^t is the percentage of books missing at time t) could be fitted in a way that yielded an estimate for the first data point of 5.44 percent, only .01 percent deviation from the actual. But the estimate for the last data point (January 1974) using this curve would be 3.06 percent, which compared very unfavorably with the actual value of only 2.2 percent.

It seems very likely that the magnitude of the first data point (March 1971) reflects the effects of books in the process of being reshelfed at the time, i.e., a "float." This "float" was not checked. By the time of the second search, however, four months had elapsed, and all (or practically all) of the missing volumes that had been in the "float" were now either on the shelves or recorded as checked out to a borrower—i.e., they were no longer missing. From the second search on, then, the "float" did not significantly affect the number of missing volumes.

The magnitude of this float probably would vary from one library to another depending on the details of their reshelfing procedures. We, therefore, concluded that a simple exponential curve ($Y_t = AB^t$) fitted to points II through

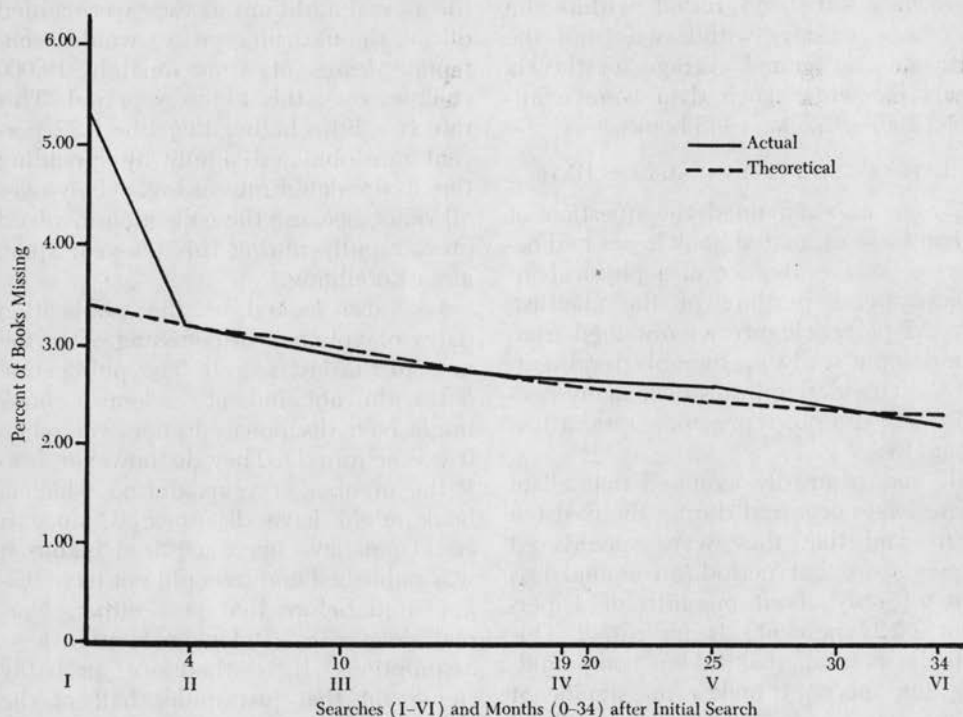


Fig. 1
Percent of Books Missing, at Various Intervals

VI (omitting I) was the most appropriate general indicator of the effect of successive searches. The formula for the curve that does this is:

$$Y_t = 3.378 (0.9868^t)$$

where t = serial number of months (1, 2, 3, . . . 34) from first search. It was calculated by using a computer program for fitting regressions and is plotted on Figure 1.

SHELFIST DELETIONS

This turned out to be just the first step in estimating Olin Library's losses, since even if we accept the validity of the sampling results showing accumulated losses equal to 2.2 percent of the entire collection deposited there, all this discloses is the discrepancy between the books the shelflist showed were in the collection in March 1971 and the vol-

umes actually there (or charged out to borrowers). Naturally, this raised the question of what deletions in the shelflist had been made over the years, for one reason or another. Investigation determined that none of the existing staff knew of any physical inventory and, in particular, no actions since the library moved into its new building in 1962, that might have resulted in a wholesale revision of the shelflist.

One regular type of adjustment was made to the shelflist that bore on the question of book losses. This related to books that library users could not find themselves and that the Circulation Department, after being asked to help, also could not find. The standard procedure calls for three searches to be made for items of this type, the last taking place one year after the first. If the

item has not been found within the year, the card is withdrawn from the catalog. The annual average for the six years for which such data were available, 1967-1972, was 445 books.

ESTIMATING THE ANNUAL LOSS RATE

There also remained the question of when these estimated book losses had occurred. In the absence of a physical inventory and purging of the shelflist, the 2.2 percent figure we obtained from our sample study presumably represented the cumulative losses over many decades and did not represent a rate of loss annually.

If one arbitrarily assumed that all of these losses occurred during the past ten years and that they were spread out evenly over that period, an annual loss rate of only about one-fifth of 1 percent (0.22 percent) is indicated. This rate is so small that taking compounding into account makes no significant difference—the loss rate would still be 0.22 percent per year.

Complicating the determination of annual losses was the fact that this collection had doubled during the decade, 1962-72, reaching 953,809 volumes by 1972. We speculated that losses each year bore some relationship to the number of volumes on hand, with larger losses being incurred when the collection was (relatively) large (i.e., the more recent years) than when it was small. We, therefore, sought a level annual rate of loss that, applied to the starting inventory ten years earlier and to each year's additions beginning when they occurred, would yield the losses accumulated by early 1971. The accumulated loss for ten years was assumed to be approximated as 2.2 percent of the June 30, 1970, perpetual inventory (844,301 volumes), or about 19,000 volumes. By trial and error, we found that a rate of loss of about 0.3 percent, compounded annually (applied to the book inventory ten years earlier and picking up each of

the annual additions as they were added during the next nine years) would accumulate losses of approximately 19,000 volumes over this ten-year period. This rate is a little higher than the 0.22 percent rate obtained simply by spreading the accumulated rate of loss evenly over all years, because the collection involved grew rapidly during this ten-year span, almost doubling.

We also looked at the publication dates of volumes still missing after the next to the last search. The publication dates do not indicate when a book might have disappeared (nor even when it was acquired). They do, however, limit the number of years during which a book might have disappeared, since it could not have been acquired before it was published and so could not have disappeared before that year either. Normal delays in cataloging fortify these assumptions. It is, therefore, probably significant that just under half of the seventy-five books missing at the time of the second to last search (April 1, 1973) had been published during the last decade, 1961-1970. This fact lends some support to the assumption that the cumulative losses might be spread approximately over the last ten years with some, but not great, overstatement. The increasing numbers shown over the individual years from 1950 through 1970 also roughly parallel the growth in the collection during this period, although it should be reiterated the publication date does not indicate when a volume disappeared, only the earliest time it could have disappeared.

To the estimate of a 0.3 percent annual rate of loss based on our sample study we added an adjustment for the rate at which the lost volumes were being purged from the shelflist, since their having been purged precluded any of them from being included in the sample. As was stated earlier, an average of 445 lost volumes were purged annually for the six years that such data

were available. Since the average size of the collection during those six years was 816,800 volumes, this amounted to a loss rate of 0.05 percent annually. Summing the two loss rates yielded a total estimated annual loss rate of about 0.35 percent. Subject to the evidence that may be developed in the future showing this estimate to be very wide of the mark, losses at this rate did not seem to us to justify further consideration of an electronic security system for this collection, so we did not pursue that objective any further.

A SECOND STUDY: THE ART AND ARCHITECTURE LIBRARY

In May 1973 a second study was begun in the Art and Architecture Library, one of the university's several departmental libraries. This collection is located in a different building, near the schools that account for the major use of this material. It includes 50,202 volumes. Unlike Olin Library, it uses a manual system to charge out books to borrowers, and it has no form of exit control.

The sample chosen here was smaller, 594 volumes, and, on the average, the searches were conducted at shorter intervals than in the Olin study. The first, conducted on April 30, 1973, showed 153, or 25.5 percent, of the sample to be missing. Eight months (and fifteen searches) later, in December 1973, the missing total had been reduced to 55 (9.2 percent). By May 19, 1975, two years and twenty-two searches later, the lost volumes numbered only 33 (5.5 percent). The curve of declining losses was similar to that found in the Olin collection, bulking relatively large at the time of the first search but shrinking by about one-half after four months had elapsed. Beginning with the fourth month, an exponential equation, $Y = .1262 (.9691^t)$, described subsequent reductions very well ($R^2 = .95$), just as it did in the Olin case.

On the other hand, the magnitude of the losses in the Art and Architecture Library study is much greater than in the Olin study. The last estimate, 5.5 percent accumulated losses, is more than twice as high as Olin's was two years after the original search. At the 95 percent confidence level, the upper limit for the estimate of Art and Architecture books missing would be 8 percent, and the lower limit 4 percent. Since the staff of the Art and Architecture Library had made a complete inventory of the collection in the summer of 1971 and had brought the shelflist and inventory into agreement by the time the sample was drawn, the estimated losses of 5.5 percent had accrued over a time span of only two years. Therefore, an annual rate of about 3.75 percent was implied. Volumes requested by patrons that could not be located following searches by the staff were routinely deleted. (As in Olin, staff searches—six in the case of this collection—were made over a one-year period before the items were deleted.) The record of such items for the past four fiscal years showed an average of 309 per year, or an additional loss rate of about 0.6 percent. (A few books deleted represent items lost by borrowers who make payment for the book, but these do not materially affect the net losses; although historical records of such items have not been kept, in the year ending June 30, 1975, for instance, there were only thirty-five items like this.) Adding this to the sample loss rate yielded an overall loss rate of 4.35 percent per year, a level substantially higher than the 0.35 percent estimate for the Olin collection. The disparity in the two rates is still wide even if the lower limit of 4 percent for the sample study result is used; this implies an annual loss rate of about 2 percent, increasing to 2.6 percent per year when the losses identified through patrons' requests are added. Actions, including the establishment of stronger exit controls, have

been taken to reduce losses from the Art and Architecture collection.

Three reasons for the difference in losses between the Art and Architecture and the Olin collections seem plausible. The first is the relatively greater attractiveness of the volumes in art and architecture on the average both because of their contents and their higher prices. The second is the lack of formal exit control, although some minimal security was obtained by the surveillance of staff, the circulation desk being located near the library's exit. A third is possibly the manual system for charging out volumes to borrowers which, if not well maintained, may impair timely follow-up on past due items and be more liable to errors or losses of data.

EVIDENCE BASED ON OTHER WASHINGTON UNIVERSITY COLLECTIONS

We have not made sample studies of other departmental collections at Washington University, but evidence is available from some of them that tends to help validate our estimate of losses for Olin, low though it appears. The biology collection (27,218 volumes) conducted full inventories at a two-year interval (1971-1973) and found losses of 1.39 percent, or about 0.70 percent per year on the average. The Chemistry Library (16,719 volumes) took inventory annually, and for the year ending October 1973 found losses only 0.18 percent. The earth sciences collection (19,632 volumes) showed 0.54 percent missing between its 1972 and 1973 inventories, while the East Asian collection (60,108 volumes) showed a loss rate of 0.19 percent during its first year in new quarters (it was removed physically from Olin in 1972). Bearing in mind that, although these collections usually receive close attention from their professional staffs, none of them has the kind of exit control that Olin has and that all use manual systems for charging out volumes, we think their

loss data lend additional credence to our estimate of 0.35 percent per year for the Olin collection.

FACTORS AFFECTING BOOK LOSS RATES

Based upon our results, it seems quite likely that the kind of books comprising a collection can have an important effect on its loss rate—expensive, attractive volumes of broad appeal are more likely to disappear than the average document. Second, the existence of an exit control, the nature of the exit examination, and the thoroughness with which those examinations are conducted undoubtedly affect the loss rate, probably significantly. In the case of Olin, all persons exit through one entrance. All briefcases and bundles are inspected. Library-owned volumes are easily identified, and the fact that a borrower has properly charged out a book is indicated by a date stamp on a slip pasted to the rear flyleaf of each volume. When the volume is returned, this date stamp is overprinted by a "returned" stamp. Finally, although the standard procedures may be followed in the breach at times, by and large we believe that inspections are thorough. In effect, ours is a manual system that closely approximates an electronic system when the standard procedures are followed, although it is perhaps more subject to human frailties. The contrast in loss rates between the Olin and Art and Architecture collections clearly demonstrated the importance of these two factors in our opinion.

We have also mentioned a third factor that we think probably has some effect on reducing losses: the electronic system Olin uses for charging out volumes, maintaining a record of the books on loan and automatically dispatching recall notices for past due documents. We think a well-designed system of this type is less liable to errors, of both omission and commission, or to misuse, than at least a great many man-

ual systems. Fourthly, the loss rate in any library is obviously affected by the proportion of its volumes that circulate compared to those that must be used within the library. Less obvious, however, is another factor: the proportion of patrons' voluntary use of documents within the library, rather than withdrawing them for use elsewhere. Especially in a university environment, the availability of sizable, attractive space designated for this purpose for undergraduates, graduate students, and faculty, including appropriate and sufficient study carrels, probably cuts down on the external use and thus on losses.

Other factors remaining constant, it also seems probable that increased borrowing—i.e., an increase in circulation for external use—would also raise the loss rate, although whether this would be proportional or not, we are unprepared to say. We also suspect that changes in the social environment, and specifically the social unrest of the late 1960s, may also be an influence on the loss rate. Finally, of course, there may be a significant accumulated "paper loss" of volumes—items physically present within the four walls of the library but unlocatable except by chance—as when books are mis-shelved either by staff or patrons, or when there are improper additions to or deletions from the shelflist and public catalog.

Clearly, there is a complex of factors that bear on the losses incurred by any particular library collection, and it is no easy task to identify and measure each. Moreover, the circumstances vary so much from one collection to another that simple comparisons of loss rates usually will not be very fruitful. Perhaps to oversimplify a bit, a library requires a sound circulation system, trained operating personnel, and effective supervision, all in terms of its particular circumstances, to properly control book losses. We would suggest that it also needs periodic inspections and

sample studies like ours to verify that the system is, indeed, functioning effectively.

CONCLUSION

We believe that the results of our study were important on three counts. First, the study reiterated the usefulness of sampling theory in studying book losses. Second, and probably more significant, is the discovery that there can be substantial returns from periodic searches for missing volumes. Over a two-year span, these cut the original losses by more than half in each of the two collections sampled. While we certainly do not suggest that the loss shown after an original search is always twice the real losses, our results do suggest that first search results are likely to be quite misleading and that additional periodic searches are essential to obtain a reasonably accurate estimate.

Finally, our study rather strongly supports the thesis that two factors significantly affecting book losses are the nature of the collection and the nature of the exit control. More sample studies, relating to several years, are needed before a more sophisticated model of the loss process can be developed. We plan to do additional sample studies over the next few years to obtain additional insights into the matter of losses.

REFERENCES

1. Michael Bommer and Bernard Ford, "A Cost-Benefit Analysis for Determining the Value of an Electronic Security System," *College & Research Libraries* 35:270-79 (July 1974).
2. Jay B. Clark, "An Approach to Collection Inventory," *College & Research Libraries* 35:350-53 (Sept. 1974).
3. F. H. Ayres, Letter to the Editor, *College & Research Libraries* 36:153 (March 1975).

$$4. SP = 2 \sqrt{\frac{(P)(1-P)}{N}}$$

where P = percent books missing
 S = accuracy of P , expressed as a percent of P
 N = sample size

For a 90 percent confidence level, the "2" is changed to "1.64." See John E. Freund and Frank J. Williams, *Elementary Business Statistics* (Englewood Cliffs, N.J.: Prentice-Hall, 1964), p.233ff., or any elementary text on statistics.

5. It is interesting to note, in a 1969/70 study at Cambridge University (England) of patrons discovering that books in the catalog they wanted were unavailable, permanently missing books accounted for 2.7 percent of the failures in one term's data and 2.0 percent for another term's. It was the custom for all borrowers to return all books at the

end of each term, and missing volumes were determined after this had been done. See John A. Urquhart and J. L. Schofield, "Measuring Readers' Failure at the Shelf," *Journal of Documentation* 27:276 (Dec. 1971). Great weight cannot be attached to the close correspondence of these proportions to those obtained in our study because their sample was of readers' requests that could not be filled, biased toward popular volumes in many cases, rather than being a random sample of all volumes. Furthermore, we know nothing of how often that library had taken full or partial inventories and adjusted their card catalog accordingly.

ON OUR COVER

The opening in 1873 of the Chancellor Green Library for the College of New Jersey at Princeton symbolized a new freedom of access that was beginning to prevail among college libraries. The building was considered one of the finest in the country at the time. Its 18,000-volume collection began to grow at a rapid pace under the hands of its new librarian, Frederic Vinton. Given the title of the distinguished jurist, Henry Woodhull Green, to make clear that it was named for him and not the donor, the building was the gift of his brother, John Cleve Green, who provided the \$120,000 fund. The Victorian Gothic octagonal structure, flanked by two small outlying octagons, was sixty-four feet in diameter and fifty feet in height, with the second floor made of perforated iron to permit the librarian to see everyone in the library from his place in the center of the reading room. (The exterior is shown in our January issue, p.38). The ornate showplace served the college for some twenty years when, with the construction of the connecting Pyne Library with a capacity of one million volumes, the old library was transformed into reading rooms for the library of the newly named Princeton University. At last in 1948, with the construction of the Harvey S. Firestone Library, the old buildings were converted to other uses. The Chancellor Green Library itself is now the student center.—
W. L. Williamson, Professor, University of Wisconsin-Madison