# A Comparison of Work-Sampling and Time-and-Motion Techniques for Studies in Health Services Research

*Steven A. Finkler, James R. Knickman, Gerry Hendrickson, Mack Lipkin, Jr., and Warren G. Thompson*

**Objective.** This study compares results and illustrates trade-offs between work-sampling and time-and-motion methodologies.

**Data Sources.** Data are from time-and-motion measurements of a sample of medical residents in two large urban hospitals.

**Study Design.** The study contrasts the precision of work-sampling and time-and-motion techniques using data actually collected using the time-and-motion approach. That data set was used to generate a simulated set of work-sampling data points.

**Data Collection/Extraction Methods.** Trained observers followed residents during their 24-hour day and recorded the start and end time of each activity performed by the resident. The activities were coded and then grouped into ten major categories. Work-sampling data were derived from the raw time-and-motion data for hourly, half-hourly, and quarter-hourly observations.

**Principal Findings.** The actual time spent on different tasks as assessed by the time-and-motion analysis differed from the percent of time projected by work-sampling. The work-sampling results differed by 20 percent or more of the estimated value for eight of the ten activities. As expected, the standard deviation decreases as work-sampling observations become more frequent.

**Conclusions.** Findings indicate that the work-sampling approach, as commonly employed, may not provide an acceptably precise approximation of the result that would be obtained by time-and-motion observations.

**Keywords.** Work-sampling, time-and-motion

How health workers spend their working time is of interest to health services researchers. Studies requiring such information range from evaluations of the use of physical therapy personnel time, through work measurement for nursing services, to Hsiao's work on developing a relative-value scale for physician services (Domenech et al. 1983; Hsiao et al. 1987). Two widely used techniques for collecting work activity information are work-sampling and time-and-motion. Both techniques are used frequently by industrial engineers. Each technique has strengths and weaknesses.

The work-sampling technique collects data at intervals of time. For example, data might be collected by determining exactly what a worker is doing four times each hour. Sometimes the data are collected by observing the worker in action at the point in time selected for the observation. In other studies the workers use logs to self-report their activity. In some cases the intervals between observations are of fixed duration. In other cases the observations occur at randomly chosen moments in time. Typically, an inference is made about the portion of overall work time spent on an activity, based on the percent of observations that relate to that activity.

In contrast to the work-sampling method, the time-and-motion technique uses an observer to record exactly how much time is being devoted to each task. This is a much more labor-intensive method of data collection, because it requires a one-on-one observation. Observers must follow the subject continuously for extended periods of time. Each activity and its duration must be recorded on a data collection instrument.

Both methodologies have advantages and limitations, some of which are functions of the type of observation done for each. Work-sampling studies that rely on self-reported logs are generally considered least reliable, since workers may not record activities in a timely fashion, and may not be totally frank concerning what activities were being

---

Steven A. Finkler, Ph.D., CPA is Professor of Public and Health Administration, and Gerry Hendrickson, Ph.D. is Assistant Research Professor, in the Health Research Program, Robert F. Wagner Graduate School of Public Service, New York University; James R. Knickman, Ph.D. is Vice President for Research and Evaluation, The Robert Wood Johnson Foundation, Princeton, NJ; Mack Lipkin, Jr., M.D. is Associate Professor of Clinical Medicine and Director, Division of Primary Care, Department of Medicine, New York University School of Medicine; and Warren G. Thompson, M.D. is Associate Professor of Medicine in the Department of Medicine, University of Tennessee, Knoxville, TN. Address correspondence and requests for reprints to Steven A. Finkler, Ph.D., CPA, Professor of Public and Health Administration, Robert F. Wagner Graduate School of Public Service, New York University, 600 Tisch Hall, Washington Square, New York, NY 10012-1118. This article, submitted to *Health Services Research* on November 1, 1991, was revised and accepted for publication on April 21, 1993.

done at the specified sampling times (Lurie, Rank, Parenti, et al. 1989). Work-sampling approaches that use an observer or observers to record the activities of several workers are used most frequently if workers are in a circumscribed area—for example, nurses on a unit, factory workers on a floor, or pharmacists in a pharmacy. If workers are not in a circumscribed area—for example, residents traveling throughout the hospital—then the time-and-motion approach of one observer for each subject may be more feasible.

Similarly, other data collection approaches such as the use of one-way mirrors and closed-circuit television are impractical for most observations of health care workers who move extensively around a large building such as a hospital or nursing home. If use of a videocassette recorder for continuous taping were feasible, such taping could be done for extended periods. One could then review the entire tape (equivalent to time-and-motion but less obtrusive) or choose random sampling as desired. For example, this might be useful for analyses of time spent by some researchers or administrators. However, one would still have to decide whether someone should continuously watch the videotape, mirror, or television, or whether instead a sampling approach will be sufficient, requiring that the worker or videotape be watched only at specified time intervals. Thus, we come back to the choice between continuous observation versus work-sampling.

Although both time-and-motion and work-sampling methods are vulnerable to error because the workers may change their behavior upon being observed, the problem is more severe for continuous observation. Work-sampling that uses one observer on a floor may allow the observer to blend in since he or she can usually be stationed at some distance from the worker being observed. On the other hand, time-and-motion observers shadowing workers are much more obvious and are more likely to disrupt the normal routine. It is more difficult for the worker to forget that he or she is being observed.

However, the distance from the observer creates limitations in *what* can be observed. In work-sampling the observer needs to make quick judgments about behavior for a number of workers. This means that it may not be possible to make fine-grained distinctions about differing behaviors from a great distance. It may not be possible to distinguish between "professional interaction" and "personal conversation." With small distances between the observer and subject more subtle characterizations of behavior can be ascertained. For example, the observer must determine if a physician is taking a patient's medical history, or simply having a personal conversation not directly related to the patient's care.

A critical concern to health services researchers is the cost of the data collection. Work-sampling is usually less costly than continuous observation because fewer observations are made. Since continuous observation requires an observer for each subject, most studies using this methodology limit the number of subjects. On the other hand, work-sampling requires fewer observations but more subjects. Continuous observation yields a detailed description of the activities of a few workers, whereas work-sampling gives less detail, but for a larger sample of workers. The trade-off often is between depth and breadth.

One recent study followed eight residents for over 13,000 minutes using a time-and-motion approach (Knickman, Lipkin, Finkler, et al. 1992). Clearly, a weakness of this approach is that the activities of eight residents may well not be reflective of the full population of residents. It is a small sample. On the other hand, 13,000 minutes of observation is a large sample. Several authors of that study wondered if work-sampling would perhaps have given nearly as precise a result at lower cost, allowing expansion of the sample to a broader number of residents. That led to this article, which communicates some practical information for researchers considering the two approaches.

What must we consider in making the trade-off between the two approaches? The answer partly depends on how close the results of work-sampling come to approximating the results of time-and-motion techniques. That comparison is the focal point for this study. The researcher must decide the desired level of precision and how costly it will be to obtain that level using the alternative methods. (This assumes that a complete time-and-motion data collection generates an accurate portrayal of actual occurrence).

The techniques of work-sampling and time-and-motion are far from new. Textbooks in the field of industrial engineering discuss the two methods. Brisley's discussion of probability with respect to work-sampling gets to the core of this study (Brisley 1971). He indicates that if a task were to occupy 50 percent of a worker's time and we would be satisfied with a 10 percent precision level (i.e., 50% ±5%) in the work-sampling results, 400 observations would be adequate. However, if we wanted a 1 percent precision level for a task that occurred 50 percent of the time (i.e., 50% ±0.5%), we would need 40,000 observations. If the task only occurred 5 percent of the time, we would need 760,000 observations for a 1 percent precision (i.e., 5% ±.05%), and if the task occurred only 1 percent of the time, nearly 4 million observations would be required for a 1 percent precision (i.e., 1% ±.01%). Obviously, then, if we are willing to divide a subject's activities into two or three categories, work-sampling is more attractive than if we desire 100 different categories.

## WORK-SAMPLING AND TIME-AND-MOTION IN THE HEALTH SERVICES RESEARCH LITERATURE

Many health care studies have used one method or the other. In fact, one of the first examples of work-sampling took place in a hospital setting (Wright 1954). Work-sampling has been applied to nursing personnel (Abdellah and Levine 1954), midlevel health professionals (Reid 1975), emergency departments (Liptak et al. 1985), and other health services areas.

A review article in the *American Journal of Hospital Pharmacy* examined a variety of work measurement methodologies used in pharmacy research (Rascati, Kimberlin, and McCormick 1986). In pharmacy studies, work-sampling techniques have been used to measure the time taken to prepare mixtures, fill piggyback bottles, process prescriptions, and distribute controlled substance drugs, and the time spent on various clinical activities. Some authors noted that self-reported work-sampling using logs not only is less expensive than time-and-motion, but also avoids the problems of behavior changes due to being watched.[1] Work-sampling is often used in pharmacy studies that document baseline work load activities, or that compare work load activities under different conditions. Dickson (1978) looked at fixed-time intervals for work-sampling in contrast to random observations, and found no significant differences.

A number of time-and-motion studies are also available in the health literature. The National Center for Health Services Research (NCHSR) issued a report on care of the elderly that used time-and-motion to evaluate resource requirements in nursing homes (Roddy, Liu, and Meiners 1987). Gillanders and Heiman (1971) followed six interns for five days and recorded all of their activities.

None of these earlier studies have compared the same data base for both the work-sampling and the time-and-motion techniques. While the pros and cons are discussed in a number of the articles in a general way, to our knowledge no prior attempt has been made in the literature to compare the two methods directly.

## THE STUDY

A recent study (Knickman, Lipkin, Finkler, et al. 1992) examined the way that medical residents spend their time in hospitals. That study used a time-and-motion methodological approach. Although the sample was small (eight residents), each resident was followed for a substantial amount of time, and activities were recorded in great detail. The data offer a micro look at time allocations and represent systematic data in an area where careful measurement and systematic observations have been rare.

The small sample size was necessitated by limited resources for the study. The authors of that study chose the time-and-motion approach with limited sample size instead of a work-sampling approach that would have allowed for a larger number of residents observed. The present study takes the results from that earlier study and uses the data base to generate work-sampling observations. A comparison of the time-and-motion and work-sampling results is then made.

## METHODOLOGY

To assess each resident's allocation of his or her time, the time-and-motion study used 22 coders who were undergraduate premedical students. The coders received seven hours of training prior to starting the study. Included in the training was an orientation to the definition of activity codes and a "dry run" spent following a resident for a period of time with one of the research team members.

The coders were trained to record the start time and end time of each activity performed by the resident. When there was some question about the nature of an activity, the coder either asked the resident or wrote a detailed "comment" that could be interpreted later with members of the research team. The activities were coded in 67 distinct categories (see the Appendix).

The study data included information from two interns at Bellevue Hospital in New York, two interns at New York University (NYU) Medical Center, three residents at Bellevue, and one resident at NYU out of a population of approximately 600 residents. Each of the residents (we will refer to both interns and residents as residents) was in the internal medicine residency program. The two hospitals have one training program in internal medicine, and residents spend some time at NYU and some time at Bellevue. The reader is referred to the original study for additional description of the methodology used to collect the time-and-motion data (Knickman, Lipkin, Finkler, et al. 1992).

Work-sampling data were derived from the raw time-and-motion data. In time-and-motion, it is possible to tally exactly the number of minutes spent on each type of task. Then percentages of time devoted to each task can be calculated. However, work-sampling requires extrapolations of data to determine how time is spent. In work-sampling, one has information only about the task being performed at the exact time of the observation. If half of the observations showed the subject to be in the process of performing tests, then half of the subject's total time would be assumed to be performing tests.

Work-sampling cannot determine the duration of a task. While time-and-motion measures the exact number of minutes, work-sampling only catches an instant in time. However, the more observations that are made, and the more closely together the observations are made in time, the more precise work-sampling should become. At an extreme, making an observation every minute under work-sampling should yield essentially the same result as time-and-motion continuous observations.

Health services researchers are often unclear, however, on how close one would have to come to "observations every minute" in order for the work-sampling results to closely approximate those of time-and-motion. To examine this issue, we generated work sampling data from our time-and-motion study. For work-sampling data we identified the activity that a resident was engaged in at specified moments in time. We generated hourly work-sampling data by noting what each resident was doing at 20 minutes after each hour according to information from the time-and-motion study. In addition, work sampling information was generated using half-hourly data (20 minutes after the hour and 10 minutes before the hour), and quarter-hourly data (5, 20, 35, and 50 minutes after each hour). Ideally, one might want to sample on a random basis. We chose to use set time intervals because that is the approach that has generally been taken in work-sampling studies in health services research. We stopped at quarter-hourly data because we reasoned that this time interval was the shortest interval used for most work-sampling studies. If, for example, observations were made twice as often, every 7.5 minutes, the time taken locating each resident to be observed would likely take the intervening time, and time-and-motion would be just as economical. In practice, one rarely sees work-sampling studies that sample more frequently than every 15 minutes.

Hourly sampling generated 223 observations; half-hourly sampling generated 446 observations; and quarter-hourly sampling generated 892 observations. The results of each of these work-sampling intervals could then be compared with the time-and-motion results to see if they adequately converge on the time-and-motion results.

One of the problems with work-sampling, as with any sampling technique, is the potential for systematic bias in the data collection. Suppose more administrative activities are done shortly before the hour ends because people may be completing their work before leaving for other assignments, meetings, classes, and so on.

We carefully attempted to avoid that problem. Our sample used 10 minutes before the hour and 5 minutes after the hour, to avoid bias due to activities that routinely occur very close to the hour. We did not want such points overrepresented in the sample. On the other hand, one could equally argue that by *not* sampling just before and just after the hour, we failed to reflect adequately the unusual peak in administrative activity. By avoiding such typical administrative times, we may have introduced bias that results in an underreporting of administrative activities. In other words, any work-sampling approach using time intervals for sampling that are not random does create the potential for systematic bias. On the other hand, using randomly determined time intervals makes scheduling the observations difficult. Some hours may have many observations—more than the observers can handle. Other hours may have few observations. Therefore, it is common for health services work-sampling studies to take observations at fixed periodic intervals. This problem is avoided with time-and-motion.

Our expectations were that the hourly work-sampling data would differ from the time-and-motion results. We also expected that as the number of observations per hour increased, the difference between the two methods would decrease.

## FINDINGS

Tables 1 and 2 present the basic results by task and type of function, respectively. Table 1 presents the percent of time from the time-and-motion results for each task, and the percent of observations (and therefore the inferred percent of time) from the work-sampling derivation for each task. Table 2 examines the differences when they are grouped into ten major categories: education activities, information gathering, personal time, testing, consultations, documentation, transit, procedures, patient interaction, and administration. The activities included in each of the groups are shown in the Appendix.

The reason for providing this second table is that studies often are

Table 1: Comparison of Time-and-Motion to Work-Sampling Based on Activity Code Information

| | | *Work-Sampling Percent of Observations* | | | | | |
|---|---|---|---|---|---|---|---|
| *Activity* | *Time-and-Motion Percent of Minutes* | *Qtr-Hourly* | *s.e.* | *Half-Hourly* | *s.e.* | *Hourly* | *s.e.* |
| 1 | 3.4% | 2.6% | 0.53% | 2.7% | 0.77% | 2.3% | 1.00% |
| 2 | 2.3 | 4.7 | 0.71 | 5.1 | 1.04 | 4.9 | 1.45 |
| 3 | 6.0 | 10.0 | 1.00 | 10.4 | 1.45 | 11.6 | 2.14 |
| 4 | 1.8 | 4.9 | 0.72 | 5.6 | 1.09 | 4.9 | 1.45 |
| 5 | 0.3 | 0.3 | 0.18 | 0.3 | 0.26 | 0.0 | 0.00 |
| 6 | 5.1 | 3.5 | 0.62 | 3.3 | 0.85 | 2.8 | 1.10 |
| 7 | 1.2 | 0.7 | 0.28 | 0.9 | 0.45 | 0.8 | 0.60 |
| 8 | 2.5 | 1.2 | 0.36 | 1.2 | 0.52 | 1.0 | 0.67 |
| 9 | 0.0 | 0.5 | 0.24 | 0.5 | 0.33 | 0.5 | 0.47 |
| 10 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 11 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 12 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 13 | 0.7 | 0.4 | 0.21 | 0.4 | 0.30 | 0.8 | 0.60 |
| 14 | 0.1 | 0.3 | 0.18 | 0.4 | 0.30 | 0.8 | 0.60 |
| 15 | 0.1 | 0.1 | 0.11 | 0.3 | 0.26 | 0.0 | 0.00 |
| 16 | 0.1 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 17 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 18 | 1.5 | 3.8 | 0.64 | 3.8 | 0.91 | 4.1 | 1.33 |
| 19 | 1.7 | 1.3 | 0.38 | 1.3 | 0.54 | 1.3 | 0.76 |
| 20 | 0.1 | 0.1 | 0.11 | 0.0 | 0.00 | 0.0 | 0.00 |
| 21 | 0.3 | 0.1 | 0.11 | 0.3 | 0.26 | 0.0 | 0.00 |
| 22 | 0.7 | 0.3 | 0.18 | 0.1 | 0.15 | 0.3 | 0.37 |
| 23 | 0.2 | 0.1 | 0.11 | 0.0 | 0.00 | 0.0 | 0.00 |
| 24 | 0.2 | 0.3 | 0.18 | 0.3 | 0.26 | 0.3 | 0.37 |
| 25 | 6.1 | 3.6 | 0.62 | 3.8 | 0.91 | 4.4 | 1.37 |
| 26 | 2.1 | 1.6 | 0.42 | 1.7 | 0.61 | 2.1 | 0.96 |
| 27 | 0.1 | 0.3 | 0.18 | 0.1 | 0.15 | 0.3 | 0.37 |
| 28 | 0.4 | 0.3 | 0.18 | 0.3 | 0.26 | 0.0 | 0.00 |
| 29 | 0.2 | 0.3 | 0.18 | 0.1 | 0.15 | 0.3 | 0.37 |
| 30 | 1.0 | 0.7 | 0.28 | 0.5 | 0.33 | 0.0 | 0.00 |
| 31 | 5.2 | 3.2 | 0.59 | 3.0 | 0.81 | 3.1 | 1.16 |
| 32 | 8.1 | 4.3 | 0.68 | 4.4 | 0.97 | 4.4 | 1.37 |
| 33 | 1.6 | 1.2 | 0.36 | 1.2 | 0.52 | 0.8 | 0.60 |
| 34 | 0.5 | 0.5 | 0.24 | 0.3 | 0.26 | 0.5 | 0.47 |
| 35 | 1.7 | 1.1 | 0.35 | 1.2 | 0.52 | 1.6 | 0.84 |
| 36 | 2.0 | 1.2 | 0.36 | 1.0 | 0.47 | 1.0 | 0.67 |
| 37 | 6.3 | 10.7 | 1.03 | 10.6 | 1.46 | 11.4 | 2.13 |
| 38 | 0.2 | 0.1 | 0.11 | 0.0 | 0.00 | 0.0 | 0.00 |
| 39 | 0.2 | 0.1 | 0.11 | 0.1 | 0.15 | 0.3 | 0.37 |
| 40 | 0.5 | 0.3 | 0.18 | 0.4 | 0.30 | 0.5 | 0.47 |
| 41 | 1.3 | 0.8 | 0.30 | 0.9 | 0.45 | 0.8 | 0.60 |
| 42 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 43 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |

*Continued*

Table 1: Continued

| Activity | Time-and-Motion Percent of Minutes | Work-Sampling Percent of Observations Qtr-Hourly | s.e. | Half-Hourly | s.e. | Hourly | s.e. |
|---|---|---|---|---|---|---|---|
| 44 | 0.1 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 45 | 0.1 | 0.1 | 0.11 | 0.0 | 0.00 | 0.0 | 0.00 |
| 46 | 0.1 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 47 | 0.4 | 0.7 | 0.28 | 0.9 | 0.45 | 0.8 | 0.60 |
| 48 | 1.0 | 0.6 | 0.26 | 0.5 | 0.33 | 1.0 | 0.67 |
| 49 | 5.9 | 3.9 | 0.65 | 3.5 | 0.87 | 3.1 | 1.16 |
| 50 | 0.3 | 0.2 | 0.15 | 0.3 | 0.26 | 0.0 | 0.00 |
| 51 | 0.0 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 52 | 6.0 | 6.9 | 0.85 | 7.0 | 1.21 | 7.5 | 1.76 |
| 53 | 3.4 | 1.4 | 0.39 | 1.6 | 0.59 | 1.6 | 0.84 |
| 54 | 1.1 | 0.8 | 0.30 | 0.8 | 0.42 | 1.0 | 0.67 |
| 55 | 2.1 | 1.2 | 0.36 | 1.2 | 0.52 | 1.0 | 0.67 |
| 56 | 0.8 | 0.5 | 0.24 | 0.8 | 0.42 | 0.8 | 0.60 |
| 57 | 2.4 | 1.6 | 0.42 | 1.6 | 0.59 | 0.5 | 0.47 |
| 58 | 5.6 | 7.6 | 0.89 | 6.9 | 1.20 | 6.7 | 1.67 |
| 59 | 0.1 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 60 | 0.2 | 0.2 | 0.15 | 0.4 | 0.30 | 0.3 | 0.37 |
| 61 | 1.5 | 4.3 | 0.68 | 3.8 | 0.91 | 3.6 | 1.25 |
| 62 | 0.2 | 0.1 | 0.11 | 0.0 | 0.00 | 0.0 | 0.00 |
| 63 | 0.1 | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 |
| 64 | 1.1 | 0.6 | 0.26 | 0.7 | 0.39 | 0.5 | 0.47 |
| 65 | 0.9 | 0.5 | 0.24 | 0.3 | 0.26 | 0.5 | 0.47 |
| 66 | 0.6 | 3.2 | 0.59 | 3.3 | 0.85 | 3.1 | 1.16 |
| 67 | 0.3 | 0.2 | 0.15% | 0.3 | 0.26% | 0.3 | 0.37% |
| Total* | 100.0% | 100.0% | | 100.0% | | 100.0% | |

*Deviations from 100.0% are due to rounding.

not concerned with the raw results as much as with their implications. Based on the probability information from Brisley (1971), discussed earlier, we know that it might take an enormous number of observations to render a high degree of precision for 67 different tasks, each representing a small portion of total work time. A smaller set of grouped tasks might be adequate for policy analysis. Such a smaller group would require fewer observations. Further, the smaller set looks like most work-sampling check sheets, which are usually shorter and simpler than what would be used for continuous observation because the former is what is practical with one observer and several subjects.

The coders measured both the activities that occurred and the time spent on each activity. Thus, it was possible to calculate actual time spent on an activity for the time-and-motion data. In total 13,383

Table 2: Comparison of Time-and-Motion to Work-Sampling Based on Type of Function

| | | *Work-Sampling Percent of Observations* | | | | | |
|---|---|---|---|---|---|---|---|
| *Percent of Minutes* | *Time-and-Motion* | *Quarter-Hourly* | *s.e.* | *Half-Hourly* | *s.e.* | *Hourly* | *s.e.* |
| A. Education activities | 20.7% | 15.7% | 1.22% | 15.9% | 1.73% | 15.8% | 2.44% |
| B. Information gathering | 13.7 | 22.4 | 1.40 | 23.9 | 2.02 | 23.8 | 2.85 |
| C. Personal | 13.2 | 7.6 | 0.89 | 7.6 | 1.25 | 7.8 | 1.80 |
| D. Testing | 12.0 | 10.5 | 1.03 | 10.5 | 1.45 | 10.6 | 2.06 |
| E. Consultation | 12.0 | 14.7 | 1.19 | 14.2 | 1.65 | 15.2 | 2.40 |
| F. Documentation | 9.8 | 6.7 | 0.84 | 6.6 | 1.18 | 6.7 | 1.67 |
| G. Transit | 8.2 | 9.5 | 0.98 | 8.9 | 1.35 | 7.5 | 1.76 |
| H. Procedures | 5.4 | 6.5 | 0.83 | 6.6 | 1.18 | 7.2 | 1.73 |
| I. Patient interaction | 3.1 | 2.1 | 0.48 | 1.9 | 0.65 | 1.8 | 0.89 |
| J. Administration | 1.8 | 4.3 | 0.68% | 3.8 | 0.91% | 3.6 | 1.25% |
| Total* | 100.0% | 100.0% | | 100.0% | | 100.0% | |

*Deviations from 100.0% are due to rounding.

minutes of activities were observed across the eight residents. During these minutes, the residents were involved in 1,726 distinctly coded activities. Activities averaged 7.75 minutes with a standard deviation of 14.15 minutes.

As one examines Table 1, no consistent pattern is revealed. As expected, the actual time spent on different tasks does differ from the percent of time projected by the work-sampling results. As one looks at the hourly results, followed by the half-hourly and then the quarter-hourly, in some cases there is a definite movement toward the time-and-motion results, which is what we would expect. For example, looking at activity 49, the expected trend is apparent. The time-and-motion result indicates that 5.9 percent of all time was spent on this activity. The hourly work-sampling observations indicate that 3.1 percent is spent on the activity; half-hourly indicates 3.5 percent; quarter-hourly observations indicate 3.9 percent. The standard deviation decreases as the work-sampling observations become more frequent.

In some cases, however, such as activity 19, no movement takes place. And in some cases, the trend is the opposite of that expected, such as activity 26, where the more frequent the work-sampling observations, the less similar the result is to the time-and-motion finding.

In 29 of the 67 cases the quarter-hourly data are an improvement over the hourly data. In 23 cases the quarter-hourly data are the same as the hourly data (however, in seven of those cases no observations were recorded in either the time-and-motion or work-sampling data, so the data were identical). In 15 cases the quarter-hourly work-sampling deviated more from the time-and-motion data than did the hourly data.

Table 2 offers similar results. For example, for the procedures category, more frequent work-sampling observations brought the results closer to the time-and-motion finding. On the other hand, administration shows the reverse trend. For four categories the quarter-hourly work-sampling data better approximated time-and-motion results than did the hourly data, one category showed no change, and in five categories the quarter-hourly data showed a poorer approximation.

## DISCUSSION

The difference between the time-and-motion and quarter-hourly work-sampling results is 20 percent or more of the estimated value for eight of the ten activities in Table 2. Are the results of the work-sampling approximation sufficiently close to the time-and-motion results? This depends on the ultimate use for the data.

Earlier we discussed the use of a 10 percent precision level. This implies that the activity would be estimated, *plus or minus 10 percent of the activity's frequency*. An activity occurring 15 percent of the time would be evaluated at a level ±1.5 percent. Some investigators might be satisfied instead with a constant interval such as 5 percent. For example, an activity that occurs 40 percent of the time would be assessed ±5 percent (instead of ±4 percent), and an activity that occurs 1 percent of the time is also measured ±5 percent (instead of ±0.1 percent). Thus, an activity that our estimation shows occurring 1 percent of the time, might actually be occurring 6 percent of the time, or 500 percent more frequently than thought. This would not require nearly as large a sample as a 10 percent precision. However, for some purposes this range may be too broad to be useful.

Researchers generally realize that they are possibly prone to some error if their sample sizes are too small. However, the extent of that error may not be apparent. In this study, because we used the same data for the time-and-motion and the work-sampling, it has been possible to see exactly how greatly the results differ when a sample size smaller than desirable is used.

One might question whether the aggregation process or the observation methodology is the source of the discrepancy between work-sampling and time-and-motion results. For example, what if observers categorized an observation in the wrong activity? Or for Table 2,

Figure 1: How Confidence Interval Varies in Relation to Percent Occurrence of the Element and Sample Size at Various Time Intervals

| | Confidence Interval (95%) | | |
|---|---|---|---|
| **Percentage Occurrence (Frequency) of Element** | **Hourly** $n = 223$ | **Half-hourly** $n = 446$ | **Quarter-hourly** $n = 892$ |
| *50* | ±*6.7* | ±*4.7* | ±*3.3* |
| *40* | ±*6.6* | ±*4.6* | ±*3.3* |
| *30* | ±*5.1* | ±*4.3* | ±*3.1* |
| *20* | ±*5.4* | ±*3.8* | ±*2.7* |
| *10* | ±*4.0* | ±*2.8* | ±*2.0* |
| *5* | ±*2.9* | ±*2.1* | ±*1.5* |
| *1* | ±*1.3* | ± *.9* | ± *.7* |

suppose that the activities were combined incorrectly. This cannot be the cause of the difference between the two methods. The work-sampling data were drawn from a population totally defined by the earlier time-and-motion data collection. Every activity observed during the time-and-motion study was assigned an activity number (1 to 67). All activities assigned a value from 1 to 10 were education activities; all activities assigned 11 to 15 were information gathering activities (see Appendix). This was true when the original observations were made, and the categories were maintained when the work-sampling was done.

One might contend that an observation may initially be categorized incorrectly. If so, that creates an inaccuracy in the original time-and-motion data collection. However, the same error would have to be made when the already collected time-and-motion data were used for the work-sampling calculations. If the error exists for one method it exists for both, and it therefore cannot be used to explain any discrepancy between the methods.

Where does this leave us? First, we can definitely see that the results differ for the time-and-motion and hourly work-sampling observations, at least with the number of observations made. Were the hourly, half-hourly, or even quarter-hourly number of observations sufficient? Figure 1 yields some insight on this problem.

Given the number of observations we had, if an activity is expected to occur 50 percent of the time, then the half-hourly observations would yield a confidence interval of ±4.7 percentage points at a 95 percent confidence level, which we might consider to be acceptable. The ±6.7 percentage point confidence interval for hourly observations might be inadequate for us to conclude that our work-sampling results reasonably approximate time-and-motion results.

However, in Table 1 there is not a single activity that time-and-motion observed to occur even 10 percent of the time, let alone 50 percent of the time. Looking at the aggregated Table 2, the most frequent activity was Education, which consumed 20.7 percent of the residents' time. Given the number of hourly observations used in the work-sampling reported in this paper, the confidence interval for a 20 percent activity is ±5.4 percentage points. For quarter-hourly observations it would be ±2.7 percentage points. Given these confidence intervals, the results reported in Table 2 do not seem too bad. Work-sampling is approximating time-and-motion as would be expected given the precision level associated with the work-sampling sample size.

However, if one were to undertake a work-sampling study expecting 20 activities, each taking approximately 5 percent of the workers' time, one would need 7,600 observations for a confidence interval of ±0.5 percentage points (i.e., 10 percent of the time spent on that activity). See Figure 2. This is a substantial number, perhaps more than many health services researchers might expect. Would 7,600 observations suffice? Only if in fact the population were equally divided at 5 percent of time spent on each activity.

Consider the results reported in Table 2. Although several activities are around 5 percent, some are more and some are less. Looking at each result, how many observations would actually have been needed to have a confidence interval that is 10 percent of the frequency for each item in the table? (See Figure 3.)

Note that for administration one would have needed to make over

Figure 2: How Number of Required Observations Varies in Relation to Percentage Occurrence of the Element and Relative Precision Desired in Work Sampling Results

| | Number of Observations Required | |
|---|---|---|
| **Percentage Occurrence (Frequency) of Element** | **±5% of Frequency** | **±10% Frequency** |
| *50* | *1,600* | *400* |
| *40* | *2,400* | *600* |
| *30* | *3,738* | *930* |
| *20* | *6,400* | *1,600* |
| *10* | *14,400* | *3,600* |
| *5* | *30,400* | *7,600* |
| *1* | *158,400* | *39,600* |

Figure 3: Required Number of Observations for a Confidence Interval Equal to 10 Percent of Element Frequency

| Activity | Frequency | Number of Observations for Confidence Interval ±10% of Frequency |
|---|---|---|
| *Education* | *20.7%* | *1,532* |
| *Information gathering* | *13.7* | *2,520* |
| *Personal* | *13.2* | *2,630* |
| *Testing* | *12.0* | *2,933* |
| *Consultation* | *12.0* | *2,933* |
| *Documentation* | *9.8* | *3,682* |
| *Transit* | *8.2* | *4,478* |
| *Procedures* | *5.4* | *7,007* |
| *Patient interaction* | *3.1* | *12,503* |
| *Administration* | *1.8* | *21,822* |

20,000 observations. To make matters worse, administration is one of the specific areas for which we noted that systematic bias from fixed periodic sampling is likely. The sample sizes being reported here assume that work-sampling is done at random time intervals. If in fact a fixed periodic schedule is used, such as every 15 minutes (as is often the case), the results are less reliable, even if the first starting point is randomly chosen.

Clearly, the work-sampling sample sizes used in this study were inadequate. Using the Table 1 data, the hourly work-sampling results differed by more than 100 percent from time-and-motion results in approximately one-third of the activities, and in over two-thirds the results differed by more than 30 percent. Table 2 indicates that the difference exceeds 100 percent in only one case out of ten, but for six of the ten cases the difference exceeds 30 percent.

We can also see that simply moving to a more frequent work-sampling observation pattern, but still with fewer than the required observations, does not seem to resolve the differences adequately. Using quarter-hourly observations and the data from Table 1, the work-sampling results differed from time-and-motion results in approximately 20 percent of the activities by more than 100 percent, and in approximately three-quarters of them by more than 30 percent. Using Table 2 information, the difference in one case still exceeds 100 percent, and half of the ten categories differ by more than 30 percent.

It is also clear that results are more consistent between the two methods at the aggregated level, as shown in Table 2. However, even when the various tasks were aggregated for policy analysis, large percentage differences persisted. Even with education (Table 2) as the single largest category for time-and-motion, consuming 20.7 percent of the work time, the quarter-hourly work-sampling result of 15.7 percent was about 24 percent less than the time-and-motion result.

## POLICY IMPLICATIONS

Are these differences large enough to be important? Suppose that the results of the study are to be used to make policy decisions. These decisions might concern how much of an alternative work force would be needed to substitute for residents if the number of residents were altered. Some of the residents' activities would require a physician, and other activities could be replaced by lesser trained individuals. An error rate of over 30 percent in half of the categories would have to be considered to be substantial. It likely would have an effect on the decisions made, and could possibly lead policymakers to create incentives or regulations based on erroneous conclusions. Using a precision level of 10 percent would allow an activity that occurs 15 percent of the time to vary by 10 percent in either direction (i.e., $\pm 1.5$ percent). It might take up anywhere from 13.5 to 16.5 percent of the worker's time. On the other hand, using a confidence interval of $\pm 5$ percent would allow an activity that occurs 15 percent of the time to vary by 33 percent, anywhere from 10 to 20 percent of the worker's time. In this specific study, 10 percent precision is probably better than the more traditional $\pm 5$ percent confidence interval.

The goal of this article has been to contrast work-sampling methodology to the time-and-motion technique. The data from an earlier time-and-motion study were used, and work-sampling results were derived from those data. The time-and-motion method is weak because study cost constraints often result in few workers being observed out of a large population. The work-sampling method is weak because even if observations are made from almost the entire population of workers, their extrapolation may be inaccurate due to systematic bias introduced during the sampling process, or due to use of too small a sample size.

Sampling is two-staged. First, individuals and a time block are selected and, second, moments in time are selected within this first stage. For example, if work-sampling were used in the study of residents' time, residents and days would be selected first, and then the moments in time would be sampled within this time frame. The first stage is crucially important and the fact that it is expandable gives it its advantage. For example, if 400 observations are needed to yield the desired precision, then these 400 observations can be spread over four days and ten residents at the rate of ten per day, or four days and 20 residents at a rate of five per day. If the observations are at close intervals and the number of residents few (say ten residents, each observed 40 times in a given day) then work-sampling approximates continuous observation and has all of its advantages and disadvantages. On the other hand, if the numbers of days and residents are large, then accuracy improves, but costs also go up.

Our study was limited because of the size of the initial study and data base. As a result, our work-sampling quarter-hourly observations consist of only 892 data points, whereas sampling theory indicates that perhaps 22,000 observations were needed for 10 percent precision on all grouped activities. Note from Figure 3 that administration occurred only 1.8 percent of the time, requiring 21,822 observations at a .10 precision level. And those observations must be random. Time-and-motion represents a 100 percent data collection for the periods observed (13,383 resident minutes). Therefore, the data concerning the specific 13,383 minutes in the original study may be viewed as the true value of the eight residents' time allocation to tasks. As a result, the fact that the original study's sample was small has little if any bearing on this comparison of work-sampling and continuous observation.

By definition, the data contained in our time-and-motion results represent true data for purposes of this study. The data are not necessarily a perfect reflection of how all residents spend their time. They are not necessarily generalizable even to the hospitals where the data were collected. However, our results do represent an accurate picture of the data base against which the work-sampling took place and against which it was compared. Generally, when work-sampling is done, there is no way to know how precise the results are, because the true underlying values for the population are unknown. In this study, however, since the work-sampling was done against the known time-and-motion data, the true values of the population are known. To the extent that work-sampling is an accurate method, the data sampled from our time-and-motion data base should be able to approximate that data base, if the sufficient number of observations are present.

We believe our results indicate the need for a high number of work-sampling observations if the work-sampling approach is to reasonably approximate the time-and-motion results in many health services research studies. By "a high number" we mean a number of observations that is greater than would be undertaken in many of the work-sampling studies actually done and reported on. It is possible that researchers sometimes erroneously calculate sample sizes using unrealistic expectations of the frequency with which activities take place.

Suppose that there are 20 activities, and a priori, none is known to occur more or less frequently than any other. Each therefore might be expected to occur approximately 5 percent of the time. Would the researcher be content with a result of 5 percent $\pm 5$ percent, or would the researcher want a confidence interval of 5 percent $\pm 0.5$ percent? This is very situational, and depends substantially on the ultimate use of the calculated result.

The intent of this article is not to critique the work-sampling method, but rather to highlight its inappropriate use. Sample sizes are in fact critical in this method—more critical than one might expect. The simple response to the problem would be to use a larger sample size. However, financial constraints often limit that alternative, because in fact, the sample sizes needed for accurate results are potentially enormous, as we have seen.

Because in this study we did not randomly sample for the work-sampling data points, nor establish the random nature of occurrences, nor take a large enough number of observations for 10 percent accuracy, an industrial engineer would not find it surprising that the work-sampling results were not adequate for approximating time-and-motion results. This does not controvert well-accepted statistical theory. Nevertheless, given several hundred, or perhaps several thousand observations, many health services researchers might believe intuitively that work-sampling, while not perfect, would be an adequate approximation of time-and-motion. However, this article serves as a concrete demonstration of the potentially inaccurate results from even large numbers of observations, if they are not as large as statistically required.

This leaves researchers with a difficult choice. It is less expensive to get many work-sampling points than to do time-and-motion analysis. If our resources are devoted to time-and-motion studies, our sample size will likely be limited. We will have very good information about those subjects and activities that are observed, but we may not have sufficient volume to base a generalization on.

However, this would still appear to be preferable to working with many work-sampling data points and having the ability to generalize about conclusions—but with a degree of precision regarding the data and conclusions that would be subject to question. It is possible that many health services researchers using work-sampling have an intuitive belief that the approach provides a highly accurate picture of how time is spent. The limited findings of this study raise a strong note of caution. Unless an extremely large number of work-sampling observations are made, it is quite possible that work-sampling inferences will be inaccurate.

# APPENDIX

## Activity

*A. Education Activities*

1. work management
2. work management, student
3. grand rounds
4. attending rounds
5. other conferences
6. teaching
7. chief resident rounds
8. reading medical literature
9. morning report
10. travel to education activities

*B. Information Gathering*

11. history
12. physical
13. chart review (not lab tests)
14. phone calls (not lab tests)
15. searching for medical records

*C. Personal*

16. sleeping
17. other (eating, talking, etc.)

*D. Testing*

18. blood tests (venous)
    arrange
    perform
19. blood tests (arterial)
    arrange
    perform
20. x-rays
    order and arrange—UH
    order and arrange—BH
    interpret results
21. lumbar puncture
    order and arrange
    interpret results
22. thoracentesis
23. paracentesis
24. arthrocentesis
25. EKG
    perform
    order and arrange
    interpret results
26. urinalysis
    order and arrange—UH
    order and arrange—BH
    interpret results
27. gram stain or AFB stain
28. blood smears
    interpret results
    order and arrange
29. serum ketone determination
30. special tests

*E. Consultation*

31. specialists/attendings
32. nurses
33. other hospital staff
34. sign-out rounds
35. other residents

*F. Documentation*

36. making chart notes
37. writing orders
38. issuing verbal orders
39. making discharge summaries
40. making personal notes
41. filling out other forms

*G. Transit*

42. transporting patients
43. for work-related activities
44. for personal activities

*H. Procedures*

45. Swanz-Ganz catheter
46. arterial line
47. central line
48. venous line
49. intubation
50. ventilator management
51. suture and wound care
52. dressing changes
53. hazardous waste disposal
54. blood transfusion
55. nasogastric tube
56. urinary catheter
57. cardiac arrest
58. intravenous medications

*I. Patient Interaction*

59. instruction or counseling: patient
60. obtain informed consent: patient
61. conversation: patient
62. instruction or counseling: family
63. obtain informed consent: family
64. conversation: family

*J. Administration*

65. phone calls
66. discharge planning
67. hospital management

## ACKNOWLEDGMENT

The authors are indebted to the anonymous reviewers for their valuable comments on an earlier draft of this article.

## NOTE

1. This may be only partly true. When subjects are recording their activities they may either lie about what they are doing or change normal routine in order to provide observations they believe are in some sense more desirable.

## REFERENCES

Abdellah, F., and E. Levine. "Work-Sampling Applied to the Study of Nursing Personnel." *Nursing Research* 3, no. 1 (June 1954): 11-16.

Brisley, C. L. "Work Sampling." In *Industrial Engineering Handbook*. Edited by H. B. Maynard. New York: McGraw Hill, 1971.

Dickson, W. M. "Measuring Pharmacist Time Use: A Note on the Use of Fixed Interval Work Sampling." *American Journal of Hospital Pharmacy* 35, no. 10 (October 1978): 1241-43.

Domenech, M., O. Payton, J. Hill, and R. Shukla. "Utilization of Physical Therapy Personnel in One Hospital." *Physical Therapy* 63, no. 7 (July 1983): 1108-1112.

Gillanders, W., and M. Heiman. "Time Study Comparisons of Three Intern Programs." *Journal of Medical Education* 46 (February 1971): 143-49.

Hsiao, W. C., P. Braun, E. R. Becker, and S. R. Thomas. "The Resource-Based Relative Value Scale." *Journal of the American Medical Association* 258, no. 6 (14 August 1987): 799-802.

Knickman, J. R., M. Lipkin, Jr., S. A. Finkler, W. G. Thompson, and J. Kiel. "The Potential for Using Non-Physicians to Compensate for the Reduced Availability of Residents." *Academic Medicine* 67, no. 7 (July 1992): 429-38.

Liptak, G. S., D. M. Super, N. Baker, and K. J. Roghmann. "An Analysis of Waiting Times in a Pediatric Emergency Department." *Clinical Pediatrics* 24, no. 4 (April 1985): 202-209.

Lurie, N., B. Rank, C. Parenti, T. Woolley, and W. Snoke. "How Do House Officers Spend Their Nights? A Time Study of Internal Medicine House Staff on Call." *New England Journal of Medicine* 320, no. 25 (22 June 1989): 1673-77.

Rascati, K., C. Kimberlin, and W. McCormick. "Work Measurement in Pharmacy Research." *American Journal of Hospital Pharmacy* 43 (October 1986): 2445-51.

Reid, R. "A Work Sampling Study of Midlevel Health Professionals in a Rural Medical Clinic." *Medical Care* 13, no. 3 (March 1975): 241-49.

Roddy, P., K. Liu, and M. Meiners. *Resource Requirements of Nursing Home Patients Based on Time and Motion Studies.* Report #PHS873408, National Center for Health Services Research. Washington, DC: Government Printing Office, April 1987.

Wright, M. J. *Improvement of Patient Care.* New York: G. P. Putnam's Sons, 1954, pp. 109–125.