

Direct Measurement of Health Care Costs

Mark W. Smith

Health Economics Resource Center and Center for Health Care
Evaluation, Veterans Affairs Palo Alto Health Care System

Paul G. Barnett

Health Economics Resource Center and Center for Health Care
Evaluation, Veterans Affairs Palo Alto Health Care System;
Department of Health Care Policy, Stanford University

Cost identification is fundamental to many economic analyses of health care. Health care costs are often derived from administrative databases. Unit costs may also be obtained from published studies. When these sources will not suffice (e.g., in evaluating interventions or programs), data may be gathered directly through observation and surveys. This article describes how to use direct measurement to estimate the cost of an intervention. The authors review the elements of cost determination, including study perspective, the range of elements to measure, and short-run versus long-run costs. They then discuss the advantages and drawbacks of alternative direct measurement methods such as time-and-motion studies, activity logs, and surveys of patients and managers. A parsimonious data collection effort is desirable, although study hypotheses and perspective should guide the endeavor. Special reference is made to data sources within the Department of Veterans Affairs (VA) health care system.

Keywords: *cost and cost analysis; cost-benefit analysis; research design; data collection*

A challenging element of cost-effectiveness analysis is the proper measurement of costs. Cost data typically come from the financial records of providers or insurers, but such administrative data are not sufficiently accurate for all studies. For example, costs borne by patients and unpaid caregivers are not

Medical Care Research and Review, Vol. 60 No. 3, (Supplement to September 2003) 74S-91S

DOI: 10.1177/1077558703257001

© 2003 Sage Publications

74S

represented. Administrative data also do not give the cost of innovative treatments and may not be sensitive to changes in resource use caused by an intervention. Moreover, data from one provider or insurer do not capture activities of other providers and insurers.

When administrative cost data will not suffice, researchers frequently employ direct methods of measuring cost, gathering data through surveys and observation. This article describes these methods, with examples from the health care system of the U.S. Department of Veterans Affairs (VA). It begins with a brief overview of cost determination elements, including the choice of perspective and the need to measure all economic costs. The second section describes methods for direct measurement of the use and cost of staff, supplies, equipment, and capital. A discussion follows of methods of surveying patients to find costs that they incur and the cost of care obtained in other health care systems. We then discuss characteristics of measurement such as accuracy, precision, and validity. The final section compares methods, offers guidance on their appropriate use, presents a brief discussion, and lists areas for further research.

NEW CONTRIBUTION

Although direct measurement methods are commonplace, there has been little guidance on using them. Direct measurement has been used in VA studies of mental health (Rosenheck, Neale, and Frisman 1995), geriatric management (Toseland et al. 1997), research (Barnett and Garber 1996), and home-based primary care (Hughes et al. 2000). An earlier overview of VA cost methods briefly describes direct measurement but provides little detail and no guidance on which methods are best (Barnett 1999). The federal task force on cost-effectiveness analysis described direct measurement of health care costs, but its report was not systematic and did not provide application to VA resources (Gold et al. 1996). To our knowledge, this is the first article to compare methods of direct measurement of health care and to offer specific guidance for their use in VA.

CONSIDERATIONS IN COST DETERMINATION

Since the choice of method depends on the range of costs to be measured and the perspective of the study, we first review some considerations that will affect the method and outcome of a cost analysis that uses direct measurement.

PERSPECTIVE

The perspective is the economic viewpoint from which an analysis is done, and it dictates the range of cost elements to include. Alternative perspectives include those of society, public payers, the VA, insurance companies, providers, and patients. The relation between perspective and cost elements is illustrated in Table 1, which shows cost elements that must be considered from the perspectives of society, the provider/payer (e.g., VA), and the patient. Luce et al. (1996) describe additional perspectives.

The top row of Table 1 relates perspective to costs that must be counted. From the patient's perspective, for example, only out-of-pocket payments for medical care are considered. A payer such as VA records the payments it makes for covered services, but not costs borne by others (e.g., the patient's out-of-pocket expenses or Medicare payments). Payments by all parties are included from society's perspective.

LONG-RUN COSTS

Many costs of an intervention vary with the number of people served. These may include medications, medical supplies, and staff time spent on direct care. Other elements such as administrative structures and capital (land, mortgages and leases, utility contracts, etc.) are fixed over a short time horizon. Even these costs may vary over the long term, however. A cost-effectiveness analysis that takes a long-term view must therefore estimate the change in administrative and capital costs that may accrue due to the intervention.

Consider the introduction of second-generation antipsychotics in the 1990s. VA patients using certain drugs in this class have substantially fewer inpatient days than do other patients (Rosenheck et al. 1999; Fuller et al. 2002). In the long run, this difference in inpatient utilization could lead to a reduction in inpatient psychiatric beds, thereby reducing the capital cost of psychiatric care. Other services may be affected as well. In the case of atypical antipsychotics, overall staffing levels may remain constant because the reduction in inpatient days is partly offset by an increase in outpatient visits.

Estimating a long-run effect requires assumptions of clinical impact and projections of future caseloads and costs. There would be two types of costs to estimate: (1) direct, pertaining to clinical staff time, space, and materials; and (2) indirect, the same inputs from administrators. A reasonable source for clinical and staffing projections would be clinicians and managers in the affected units. In VA, the cost of administrative time may be assessed directly through

TABLE 1 Cost Elements under Alternative Perspectives

<i>Cost Element</i>	<i>Societal</i>	<i>Veterans Affairs</i>	<i>Patient and Patient's Family</i>
Medical care (total cost)	All costs	All covered costs	Out-of-pocket payments
Patient time for treatment	All costs	None	Patient's opportunity cost
Paid caregiving	All costs	All covered costs	Out-of-pocket payments
Unpaid caregiving	All costs	None	Opportunity cost of caregiver time
Transportation and nonmedical services	All costs	All covered costs (if any)	All costs
Sick/disability leave, transfer payments	Administrative costs only	Amount paid + administrative costs	Amount received (negative cost)

Source: Adapted from Luce et al. (1996), Table 6.1.

existing data sets. In other organizations, it may be necessary to use a survey to estimate the total labor cost (wages plus benefits) of administrators.

A second long-term consideration is the pricing of products used in the intervention. Some interventions, particularly those in the arena of medical technology, lead to patented products and services. In these cases, the long-term cost of adopting the intervention includes paying a royalty (Garber 2000). The prices of similar patented products offer a basis for estimating the total cost. The supplies required for an intervention (e.g., drugs, medical devices) may become less expensive if an intervention is adopted widely as manufacturers increase supplies of the needed products.

Clinical efficiency can also affect the cost of an intervention. The average cost of the intervention may fall over time as clinicians become more practiced at performing it (Rosenheck, Neale, and Frisman 1995). Clinicians in specialized facilities may be more efficient at providing care than those at typical hospitals and clinics. Finally, there may be returns to scale in providing an intervention as methods of care are adjusted within a facility. If this happens, long-run costs will fall below short-run costs measured during the study. As analyzing this is a venture into the hypothetical, sensitivity analysis can play an important role in facilitating the optimal use of the findings.

SCOPE OF MEDICAL COSTS

The range of items needing direct measurement will be affected by the impact an intervention has in other areas of medical care. For instance, prescribing clozapine for treatment-refractory schizophrenia patients was shown to reduce inpatient costs considerably relative to the use of older antipsychotics (Rosenheck et al. 1999). Thus, research on pharmaceuticals should track all types of medical treatment. If patients are likely to purchase nonprescription medications or supplies, these should be monitored through direct survey as well.

JOINT PRODUCTION

In some instances, a single product is produced simultaneously with other products. For example, research and education often coincide with the delivery of health care to patients. The accounting profession recognizes several methods for assigning costs to products that are jointly produced (Finkler 1992). The first allocates cost in proportion to each product's physical measure, such as its weight or volume. Another method allocates cost according to the proportion of total sales the two products will yield, less the cost of any processing beyond the point of joint production. These approaches do not apply easily to research and education, however; neither one is physical in nature, nor is research destined for sale.

Incremental cost is often a useful concept for isolating the impact of changes in activities. Incremental cost is the additional cost that results from the production of a good or service, holding the production of all other products constant. Consider an example concerning the nurse time in a clinical research trial (Barnett and Garber 1996). Suppose that patient care activities unrelated to an experimental intervention take up 25 percent of a nurse's time; activities that benefit research and patient care take 50 percent time; and activities needed only for the research protocol take the remaining 25 percent. In this scenario, the incremental cost of research is 25 percent of the nurse's time. Incremental costs must be stated in terms of a given level of production of other products. The extra cost from an intervention adds to total health care costs given current levels of patient care.

The process of studying an intervention may itself change the cost. Patients may need to travel farther to a study site than to their usual health care facilities, for instance. Likewise, time spent by clinicians or managers filling out data collection forms should not be counted as an intervention cost; it is a research production cost.

THE COMPARATOR

Whenever possible, the cost of an intervention should be measured against a comparator, whether placebo, usual care, or another new treatment. The choice of comparator will guide how costs should be measured. A finer level of detail may be needed when alternative treatments are close substitutes than when they are quite different. For example, a comparison of two surgical techniques for coronary bypass would require time in the surgical suite to be recorded in minutes to accurately capture important differences in the costs of the two procedures. If the comparison were between surgery and pharmacotherapy, however, capturing fine distinctions in surgery time may be unnecessary and a less precise method would probably suffice.

Researchers must scrutinize data collection methods to avoid bias that might favor one treatment arm, given possible incentives for patients or providers. For instance, suppose that a new drug treatment program aims to reduce VA hospitalizations. If it is likely to simultaneously lead to greater use of non-VA services, the cost estimation method should be able to account for both VA and non-VA services with similar levels of accuracy. If the control arm uses more VA care than the experimental arm, then bias could be introduced by relying on more accurate methods for VA services but less accurate methods for non-VA services that tended to underestimate the cost.

LEVEL OF ANALYSIS

The costs of an intervention may be analyzed at many different levels: the cost per intervention, per clinic visit or hospital stay, per patient contact, per day, and so on. The aggregation level will guide the choice of data collection methods.

The choice of analysis level should be guided by the researcher's ability to collect data with accuracy and precision, and by the cost of data collection. For example, cognitive impairment may prevent patients from completing self-reports accurately (without bias) or precisely (with sufficient detail), but self-reports may be necessary to track at-home care because sending an observer to scores of patient homes could be prohibitively costly and intrusive. Data collection methods are infeasible if potential patients find them intrusive and refuse consent. This suggests focus groups and field tests of primary data collection strategies may be necessary to support cost identification. Examples of such primary collection methods are available for modification to the specific application under scrutiny (Medical Expenditure Panel Survey [MEPS] interviews, National Center for Health Statistics [NCHS] Long-Term Care Survey, VA Cooperative Studies Program (CSP) Non-VA Use Survey, etc.).

COST DETERMINATION METHOD

STAFFING COSTS

This section describes common methods of direct measurement of staff costs. These methods include traditional time-and-motion studies, in which someone observes the process of care; activity logs, in which providers monitor their own time; and surveys of managers and patients.

Whatever method is chosen, the local institutional review board (IRB) must approve the data collection method as part of the overall study protocol. The IRB submission will include consent forms and data collection instruments. Data confidentiality and human rights considerations embodied in federal laws such as the Health Insurance Portability and Accountability Act (HIPAA) and the Privacy Act may affect the feasibility of each method.

TIME-AND-MOTION STUDY

In this approach, the analyst directly observes the staff members and keeps track of the time spent on each activity throughout the day. Observing staff members can yield precise results but is costly because observers must be paid for their time in training and data collection.

It is unnecessary and prohibitively costly to have someone observe clinicians over a very long period of time. An accurate estimate of the average time needed for a procedure or service can be obtained through observations at a sample of times that vary by time of day, day of week, and so forth. If more than one observer will be used, testing should be done to assure interrater reliability. Retraining may be necessary if data collection occurs over a long time period.

It is essential to secure the support of the clinical staff to be observed. At an administrative level, staff permission may be needed to obtain access to clinical areas. And without assurances to the contrary, some clinicians may assume that observational data will be shared with supervisors and form the basis of performance rankings. This is one source of Hawthorne effects, in which clinicians change their behavior when they know they are under observation. Moreover, they should not know the intended hypothesis, as a biased effect will be derived. Observational data must include notes about the exogenous environment, and the potential for confounding effect modifiers (e.g., tension over Joint Commission on Accreditation of Healthcare Organizations [JCAHO] accreditation readiness or the presence of state certifiers in the building).

The timing of time-and-motion studies is another important consideration. Often there are daily, weekly, or monthly patterns in the types of patients seen or procedures performed. The average time spent waiting in a hospital emergency room, for instance, will vary considerably by hour, day of the week, and facility. To obtain the most representative sample of outcomes, ask the clinical staff about changes in case-mix over time. Among the studies using work sampling or comparing results of work sampling to time-and-motion studies are Reid (1975); Brock et al. (1990); Finkler et al. (1993); Guarisco, Oddone, and Simel (1994); and Oddone and Simel (1994).

ACTIVITY LOGS

A second approach is to have employees keep daily activity logs for a sample of survey dates. The staff members record activities during an interval of work (e.g., 10-, 15-, or 30-minute periods) and characterize whether the activities involve the intervention being studied or some other activity. A prime benefit of activity logs is precision. They are also likely to be more accurate than post-hoc surveys, although they are subject to Hawthorne-type effects because the staff members know they are being monitored. Activity logs carry additional administrative burdens as well: developing and pretesting the survey instrument with allowance for staff members' input, training staff members to use the logs, and following up to ensure that logs are completed and gathered. It may be necessary to survey program managers beforehand to learn which staff members will need to complete logs.

As with direct observation, it may not be necessary to use activity logs for every day of an intervention, particularly if it extends for weeks or months. A random sample of days or hours within a day will suffice, but the sampling frame must be designed with care. If an intervention becomes less intensive over time, for instance, basing an estimate on activity logs from the early days of the intervention would lead to an overestimate of total time spent.

MANAGER SURVEY

A third method for gathering staff data is to survey managers. The surveys can collect two types of information: the number of full-time-equivalent employees involved in the intervention, and the number of hours spent on the intervention per day or per week. To calculate staff compensation costs accurately, separate responses should be obtained for each category of employee involved: registered nurses, physicians, lab technicians, and so on. Finer detail may be needed if experienced or specially trained providers predominate, as in a neonatal intensive care unit.

Manager surveys are common because they take less time to prepare or complete. A single manager can report on activities of many staff members, and so another advantage is the relatively small number of people who must be surveyed. The primary drawback of manager surveys is a relative lack of accuracy and precision. Managers may have a good sense of the number of days spent on the intervention in a week, for example, but probably will not be accurate at the level of hours or half-hours. The quality of data from manager surveys depends on the effort of the managers themselves. Manager surveys are not advisable when high precision is needed or when many managers would have to be surveyed to cover the actions of all staff members involved.

CALCULATING EMPLOYMENT COSTS

Once time spent on an activity has been determined, the next step is to assign a cost to that time. Although hourly or annual earnings may be obtained through surveys, they will not be accurate guides to the total employment cost. Benefits, taxes, and time spent on overhead activities are all parts of the true employment cost, yet employees may have little information on costs incurred by their employer.

It is straightforward to determine hourly employment costs. The first step is to determine annual labor costs, including wages and benefits, assuming a 40-hour work week. To find the raw hourly cost, divide the annual cost by 2,088, the number of hours in a 52-week work year. The raw figure includes time spent on activities other than patient care such as vacation, sick leave, and administrative work. Because such nonapplied time must be spent in support of carrying out an intervention, it is necessary to adjust the hourly cost to reflect this extra cost.¹

There are two sources of VA employment costs. The first is the VA payroll system, known by the acronym PAID. Access to PAID is limited to VA employees who can substantiate a need for employee-level detail. The second source is the Financial Management System (FMS), also known as the VA general ledger. FMS data are available to employees who have access to the VA's Austin Automation Center. FMS reports all labor costs, including benefits and employer contributions to taxes. The data are arranged into subaccounts, of which 72 correspond to occupation classifications. Data are reported separately for each VA facility, allowing calculation of local as well as national average costs. A guide to using FMS to determine employment costs is available from the authors.

Check the employment costs source before designing data collection instruments. It will greatly ease the process of assigning costs if the data collection forms use the same occupation categories as the cost data. One difficulty

faced by VA researchers is the small number of FMS subaccounts for administrative work. If more specificity is needed than FMS can provide, one can collect employment cost data through surveys of employees or their managers.

SUPPLY, EQUIPMENT, AND CAPITAL COSTS

The costs of supplies and equipment may be gathered through manager surveys or by contacting manufacturers. Two caveats are in order. First, supply and equipment costs may fall if a new intervention is widely adopted. Both competition and economies of scale in production can lead the price of goods to fall as the number of items produced rises. Second, the list price of a good may greatly overstate the cost of supplies and equipment because large providers like VA frequently negotiate substantial discounts. The average wholesale price of pharmaceuticals, for instance, is often substantially higher than negotiated rates available to VA (Smith and Joseph 2003).

There are several data sources for VA capital costs. The first is the VA Cost Distribution Report (CDR), which provides the depreciation on VA buildings and equipment, but omits the cost of financing (Barnett 1999). Like FMS, the CDR may be accessed by VA employees with access to the Austin Automation Center.

Although from an accounting viewpoint VA buildings are completely depreciated after 30 years, they still have economic value. The value of the next-best alternative use can be determined by the cost of renting similar facilities or by the replacement cost of the VA facilities used in the study. Estimates of land values and rental rates for medical office space may be obtained from real estate agents or other local sources.

The replacement cost of current VA facilities may be estimated through a combination of VA financial data and proprietary commercial data (Rosenheck, Frisman, and Neale 1994). They suggest two alternative methods, one based on rental rates for similar properties and another based on replacement costs. Unfortunately, these methods can lead to very different conclusions. Across nine VA facilities, Rosenheck and colleagues found the capital cost based on rental rates to be nearly 40 percent lower than costs based on replacement. There are no a priori grounds for preferring one method to the other.

OTHER COSTS

Other types of costs that need to be measured include the cost of care provided in other health care systems, out-of-pocket costs incurred by patients, including the travel cost and nonprescription medications, and the value of

patients and informal caregivers' time. This section describes methods for estimating each of these.

Estimates of these costs are often based on surveys. Patients are asked about medication use, care received from other physicians and hospitals, the cost of travel to providers, and time spent seeking health care. The validity and reliability of such survey instruments have not been extensively studied. Length of accurate recall is an important issue that is frequently ignored. Recent research on factors affecting recall include Simmons and Schnelle (2001), Clegg et al. (2001), and Nicholson et al. (2000). Methods for reducing errors include asking patients to keep daily or weekly logs, bring in prescription bottles or papers, and provide bills from inpatient visits. Proxies may be necessary for patients who have cognitive or physical impairment, introducing another source of possible bias.

NON-VA CARE

Patients may obtain health care beyond the institution where an intervention occurs. In theory, it could account for a substantial proportion of health care spending. As noted earlier, inpatient cost and utilization is best captured by asking patients to submit logs of outside care and then writing to providers for details. If patient surveys are not feasible or do not produce adequate information, other sources may be consulted. These include VA administrative files (for VA-funded care at non-VA facilities), Medicare files, and national surveys.

VA researchers may turn to administrative sources that report payments to non-VA providers for care given to veterans. The Fee Basis files contain the cost of inpatient and outpatient services provided to VA patients by contract providers, and by noncontract providers who gave care on an emergent basis. The quality and completeness of the Fee Basis data have not been determined, however. The VA discharge files, known as the Patient Treatment Files (PTF), include non-VA inpatient stays provided under contract to VA. The PTF reports discharge date, length of stay, and Diagnosis Related Group, but not the cost of these stays.

Medicare is a prominent source of non-VA care for veterans. For a fee, researchers can request Medicare utilization and cost data from the Center for Medicare & Medicaid Services (CMS). VIREC is engaged in a project to merge Medicare data with standard VA utilization data (Hynes, Cowper, and Stefos 1999).

There are other sources of person-level cost data, although they will not be linkable to VA patients. These include national surveys such as MEPS, the

Healthcare Cost and Utilization Project (HCUP), and the Medicare Provider Analysis Review (MEDPAR), surveys carried out by professional societies, and private firms that manage health care claims.

Two articles in this supplement illustrate how costs may be estimated through a combination of internal and external data sources. Wagner, Chen, and Barnett (2002) used MEDPAR and VA information to estimate costs of VA medical-surgical inpatient stays. Phibbs et al. (2003) estimated VA ambulatory care costs using Medicare payment scales (Resource-Based Relative Value scale [RBRVS]) and others. The studies conclude that combining utilization and cost data from separate sources requires particular care. Costs in one source may refer to utilization categories that do not match those in other sources. Arbitrary simplifications are often necessary.

TRAVEL AND TIME

Analyses from a societal viewpoint include patient travel cost. Patients may be surveyed about the specific mode of conveyance and the number of miles traveled. This adds considerable complexity and may not be worthwhile if patient-incurred travel costs are a small fraction of total costs. An alternative approach is to calculate the straight-line distance from the patient's residence to the health care provider and then apply a standard mileage rate, such as the amount allowed by the U.S. Internal Revenue Service for business expenses. Without much loss of accuracy, this may be further simplified by estimating travel cost using the distance between the geographic center of the postal zip code of the patient's residence and that of the provider's location (Phibbs and Luft 1995).

Beyond the direct cost of travel is the implicit value of time spent traveling. Patients also spend time obtaining care. Society values this time, and so it must be assigned a cost in an analysis from a societal viewpoint. Analyses from the viewpoint of a payer such as VA or Medicare would not include patient time.

There are several approaches to valuing patients' time. For employed persons, the hourly wage is a reasonable measure of time cost. Many veterans and their caregivers are retired, however, and so a current wage will be unavailable. A standard practice is to assign either the minimum wage or the national average wage for home health care workers, the latter available from the U.S. Department of Labor or in the Statistical Abstract of the United States (U.S. Census Bureau 2002). An alternative to valuing time directly is to factor it into the change in quality-adjusted life years due to an intervention; Garber et al. (1996) provide an overview of this concept and recommendations.

DATA QUALITY

Ensuring data quality is as important as choosing a good method for collection. This section briefly discusses precision, accuracy, and reliability, each of which directly affects data quality.

The level of precision needed in data collection will depend on the intervention. Consider an intervention that takes 15 minutes to perform. In an outpatient setting, it may be sufficient to use a survey that records time in 15-minute intervals. The overhead cost of surgery suites is typically billed by the minute, however, and so for surgical interventions the instrument would need to record time at the minute level. Billing methods are thus a second guide when designing survey instruments.

Accuracy in gathering data is an important consideration because even small errors in reporting can accumulate from repetition. A rounding error of 1 to 2 minutes per event may become large when separate measurements are accumulated. Solutions include improved training of survey staff and pilot testing and revising data collection instruments. In some cases it may be necessary to change the basic method of data collection. If pilot testing reveals that nurses feel too busy to keep accurate time logs, for instance, direct observation by a third party could be used instead. Motivation, training, and clinician input into the observation method can all improve accuracy.

When data are collected through direct observation, the accuracy of the data will rely on the people collecting it. There are several steps that can be taken to increase reliability. Data collectors must be trained to ensure that they understand the collection forms. Retraining is advisable during lengthy collection periods. The degree of consistency between collectors—known as interrater reliability—is an important measure (Dunn 1992; Kelsey et al. 1996). It can be assessed by comparing the results when two or more people collect data from the same source.

In a famous study of General Electric's Hawthorne plant, researchers determined that employees were becoming more productive not from repeated changes in the work environment but from the knowledge that they were being carefully watched (Franke and Kaul 1978). The same issue can arise in clinical studies. Patients under study may be more likely to take medications; clinicians may work more slowly to avoid accidents or, conversely, they may work more quickly to appear more efficient. Regardless of the direction of effect, Hawthorne effects will bias study results because they will not appear under normal circumstances if the intervention is adopted widely. Researchers collecting data by direct observation can reduce the probability of Hawthorne effects by making the observation process as unobtrusive as possible. For example, recording an intervention on film and later assessing the

time spent would be less intrusive than standing at bedside with a stopwatch and a clipboard.

SUMMARY

A plan for collecting data on each cost element should be determined during the planning phase of a study. Clinical input will be essential in choosing a method and setting a timetable for collection. Here we summarize the major elements that enter into the choice of direct measurement method.

Beyond the choice of intervention and any comparators, the range of costs to be measured will rest largely on the study perspective. The standard for cost-effectiveness analyses is the societal viewpoint, although others may be useful for comparison. The process of care must be understood to distinguish the intervention itself from actions taken only to study the intervention. Costs such as royalties and capital expenses that count in a long-run timeframe may not be applicable to short-run analyses.

There are three basic methods of direct observation: time-and-motion studies, activity logs, and surveys of patients, providers, and managers. The most appropriate one will depend on clinical input, the location of intervention activities (e.g., at-home vs. in-hospital), and the presence or absence of appropriate administrative data. Administrative data will most likely be needed to estimate any long-term capital costs arising from the intervention. If costs for supplies and equipment are not available from administrative data, it will be necessary to measure them directly through surveys or direct contact with suppliers and manufacturers. Patient-incurred costs for over-the-counter supplies and medications, time spent obtaining care, and travel will require patient surveys.

Because surveys play an essential role in direct measurement, proper survey design is essential. Recent books on survey methodology include Fowler (1995), Converse and Presser (1996), Aday (1996), Rea and Parker (1997), and Dillman (2000). Published studies can offer guidance based on clinical experience. Some studies compare the reliability and validity of competing approaches, such as self-reports versus administrative data extracts (Korthuis et al. 2002) or self-reports versus proxy reports (Grootendorst, Feeny, and Furlong 1997). There is a vast literature describing the design and testing of survey instruments for particular subgroups; see, for instance, Field et al. (2002), Kahn et al. (2002), and Kressin et al. (2002). The VA Measurement Excellence Initiative has developed detailed reviews of the properties of many survey instruments. The reviews appear on its Web site (www.measurementexperts.org). A third source of design advice are research teams who have done similar studies. They may have knowledge to share on the success of particular

survey instruments in their own research. One can also obtain input on survey design from those who will complete them. Pilot testing can reveal confusing questions, poor graphic design, and other problems.

The data collection method must be feasible and feature clinical, financial, and temporal dimensions. It must not interfere with the giving of care. It must be able to distinguish costs of intervention and comparator. And it must be possible to accomplish within an acceptable timeframe and within financial means. Although using electronic sources of administrative data can save considerable time and energy, the accuracy of administrative data can vary considerably by source, across facilities, and over time. Validation studies are useful guides for determining which sources are reliable.

It is often advisable to use two or more methods in the same study to save money while obtaining an acceptable level of precision and accuracy. Consider a study comparing surgical and drug treatment. An analyst might use staff surveys or study logs to determine the cost of the initial treatment. A less precise but less costly method such as average costing could be employed to determine the cost of subsequent health care (Swindle et al. 1999). The tradeoff is between precision and implementation cost: more precise methods are typically the most labor intensive and hence the most costly to carry out.

In general, one can use direct measurement for elements most important to the study outcome and average costing for elements that are less central. It is often sensible to use average-cost methods for inpatient care when the treatments being studied are unlikely to have an impact on inpatient utilization. Wagner et al. (2002) and Phibbs et al. (2003) describe average-cost data sets created for the VA system.

AREAS FOR FUTURE RESEARCH

Every kind of direct measurement requires a data collection instrument. Survey design issues are rarely acknowledged in cost-effectiveness research, a needless omission considering that fundamental characteristics like validity and reliability are straightforward to test. From the VA perspective, psychometric testing in an elderly population would be of great value. More research is also needed into the accuracy of self-reported costs, such as over-the-counter and travel costs.

A second design issue, the ability of patients and providers to recall events, appears to get even less attention but may greatly impact the accuracy of survey measurement. A fruitful area for future research would be the development of standardized questions for the gathering of cost and utilization data. The field of cost-effectiveness research would benefit greatly if surveys of this

kind were given the kind of careful attention paid to psychometrics and quality of life.

NOTE

1. Suppose that 90 percent of an employee's time is applied to patient care and other intervention-related activities, and 10 percent to overhead. Multiply the raw hourly cost by (1.00/0.90), or 1.11. If 15 percent of the employee's time were spent on overhead activities, the adjustment would be (1.00/0.85), or 1.18.

REFERENCES

- Aday, L. A. 1996. *Designing & conducting health surveys: A comprehensive guide*. 2d ed. New York: John Wiley.
- Barnett, P. G. 1999. Review of methods to determine VA health care costs. *Medical Care* 37 (4): 9-17.
- Barnett, P. G., and A. M. Garber. 1996. The cost of VA-sponsored research. *Academic Medicine* 71 (10): 1074-78.
- Brock, D. M., C. S. Scott, T. W. Pendergrass, and S. C. Macdonald. 1990. Sampling clinicians' activities using electronic pagers. *Evaluation and the Health Professions* 13 (3): 315-42.
- Clegg, L. X., A. L. Potosky, L. C. Harlan, B. F. Hankey, R. M. Hoffman, J. L. Stanford, and A. S. Hamilton. 2001. Comparison of self-reported initial treatment with medical records: Results from the prostate cancer outcomes study. *American Journal of Epidemiology* 154 (6): 582-87.
- Converse, J. M., and S. Presser. 1996. *Survey questions: Handcrafting the standardized questionnaire*. New York: Sage.
- Dillman, D. A. 2000. *Mail & Internet surveys: The tailored design method*. 2d ed. New York: John Wiley.
- Dunn, G. 1992. Design and analysis of reliability studies. *Statistical Methods in Medical Research* 1 (2): 123-57.
- Field, T. S., C. A. Cadoret, M. L. Brown, M. Ford, S. M. Greene, D. Hill, M. C. Hornbrook, R. T. Meenan, M. J. White, and J. M. Zapka. 2002. Survey physicians: Do components of the "total design approach" to optimizing survey response rates apply to physicians? *Medical Care* 40 (7): 596-606.
- Finkler, S. A. 1992. Allocating joint costs. *Hospital Cost Management and Accounting* 4 (1): 1-4.
- Finkler, S. A., G. Hendrickson, M. Lipkin Jr., and W. G. Thompson. 1993. A comparison of work-sampling and time-and-motion techniques for studies in health services research. *Health Services Research* 28 (5): 577-97.
- Fowler, F. J. 1995. *Improving survey questions: Design and evaluation*. New York: Sage.
- Franke, R. H., and J. D. Kaul. 1978. The Hawthorne experiments: First statistical interpretation. *American Sociological Review* 43:623-43.

- Fuller, A. M., K. M. Shermock, M. Secic, J. S. Laich, and M. B. Durkin. 2002. Service use and costs among VA patients with schizophrenia taking risperidone or olanzapine. *Psychiatric Services* 53 (7): 855-60.
- Garber, A. M. 2000. Advances in cost-effectiveness analysis of health interventions. In *Handbook of health economics*, edited by A. J. Culyer and J. P. Newhouse. New York: Elsevier Science.
- Garber, A. M., M. C. Weinstein, G. W. Torrance, and M. S. Kamlet. 1996. Theoretical foundations of cost-effectiveness analysis. In *Cost-effectiveness in health and medicine*, edited by M. R. Gold, J. E. Siegel, L. B. Russell, and M. C. Weinstein. New York: Oxford University Press.
- Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein, eds. 1996. *Cost-effectiveness in health and medicine*. Oxford, UK: Oxford University Press.
- Grootendorst, P. V., D. H. Feeny, and W. Furlong. 1997. Does it matter whom and how you ask? Inter- and intra-rater agreement in the Ontario Health Survey. *Journal of Clinical Epidemiology* 50 (2): 127-35.
- Guarisco, S., E. Oddone, and D. L. Simel. 1994. Time analysis of a general medicine service: Results from a random work sampling study. *Journal of General Internal Medicine* 9 (5): 272-77.
- Hughes, S. L., F. M. Weaver, A. Giobbie-Hurder, L. Manheim, W. Henderson, J. D. Kubal, A. Ulasevich, and J. Cummings. 2000. Effectiveness of team-managed home-based primary care: A randomized multicenter trial. *JAMA* 284 (22): 2877-85.
- Hynes, D. M., D. Cowper, and T. Stefos. 1999. *Medicare data for Veterans Health Administration: A proposal for acquisition, maintenance, and analysis of data*. Undersecretary for Health, U.S. Department of Veterans Affairs. Grant No. XVA69-001 (9/01/99-9/30/02).
- Kahn, K. L., J. L. Malin, J. Adams, and P. A. Ganz. 2002. Developing a reliable, valid, and feasible plan for quality-of-care measurement for cancer: How should we measure? *Medical Care* 40 (6): 73-85.
- Kelsey, J. L., A. S. Whittemore, A. S. Evans, and W. D. Thomson. 1996. *Methods in observational epidemiology*. New York: Oxford University Press.
- Korthuis, P. T., S. Asch, M. Mancewicz, M. F. Shapiro, W. C. Mathews, W. E. Cunningham, J. A. McCutchan, A. Gifford, M. L. Lee, and S. A. Bozzette. 2002. Measuring medication: Do interviews agree with medical record and pharmacy data? *Medical Care* 40 (12): 1270-82.
- Kressin, N. R., J. A. Clark, J. Whittle, M. East, E. D. Peterson, B. H. Chang, A. K. Rosen, X. S. Ren, L. G. Alley, L. Kroupa, T. C. Collins, and L. A. Petersen. 2002. Racial differences in health-related beliefs, attitudes, and experiences of VA cardiac patients: Scale development and application. *Medical Care* 40 (1): 62-71.
- Luce, B. R., W. G. Manning, J. E. Siegel, and J. Lipscomb. 1996. Estimating costs in cost-effectiveness analysis. In *Cost-effectiveness in health and medicine*, edited by M. R. Gold, J. E. Siegel, L. B. Russell, and M. C. Weinstein. New York: Oxford University Press.
- Nicholson, J. M., D. J. Hennrikus, H. A. Lando, M. C. McCarty, and J. Vessey. 2000. Patient recall versus physician documentation in report of smoking cessation counseling performed in the inpatient setting. *Tobacco Control* 9 (4): 382-88.

- Oddone, E., and D. L. Simel. 1994. A comparison of work-sampling and time-and-motion techniques for studies in health services research. *Health Services Research* 29 (5): 623-26.
- Phibbs, C. S., and H. S. Luft. 1995. Correlation of travel time on roads versus straight line distance. *Medicare Care Research and Review* 32 (4): 532-42.
- Phibbs, C. S., W. Yu, F. A. Lynn, and P. G. Barnett. 2003. *HERC's outpatient average cost dataset for VA care: Fiscal years 1998-2001*. Menlo Park, CA: VA Health Economics Resource Center.
- Rea, L. M., and R. A. Parker. 1997. *Designing & conducting survey research: a comprehensive guide*. 2d ed. New York: John Wiley.
- Reid, R. A. 1975. A work sampling study of midlevel health professionals in a rural medical clinic. *Medical Care* 13 (3): 241-49.
- Rosenheck, R., J. Cramer, E. Allan, J. Erdos, L. K. Frisman, W. Xu, J. Thomas, W. Henderson, and D. Charney. 1999. Cost-effectiveness of clozapine in patients with high and low levels of hospital use: Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. *Archives of General Psychiatry* 56 (6): 565-72.
- Rosenheck, R., L. Frisman, and M. Neale. 1994. Estimating the capital component of mental health care costs in the public sector. *Administration and Policy in Mental Health* 21 (6): 493-509.
- Rosenheck, R., M. Neale, and L. Frisman. 1995. Issues in estimating the cost of innovative mental health programs. *Psychiatric Quarterly* 66 (1): 9-31.
- Simmons, S. F., and J. F. Schnelle. 2001. The identification of residents capable of accurately describing daily care: Implications for evaluating nursing home care quality. *The Gerontologist* 41 (5): 605-11.
- Smith, M. W., and G. Joseph. 2003. Pharmacy data in the VA health care system. *Medical Care Research and Review* 60 (3 Suppl.): 92S-123S.
- Swindle, R., C. VanDeusen-Lukas, D. Alexander-Meyer, P. G. Barnett, and A. M. Hendricks. 1999. Cost analysis in the Department of Veterans Affairs: Consensus and future directions. *Medical Care* 37 (4): 3-8.
- Toseland, R. W., J. C. O'Donnell, J. B. Engelhardt, J. Richie, D. Jue, and S. M. Banks. 1997. Outpatient geriatric evaluation and management: is there an investment effect? *The Gerontologist* 37 (3): 324-32.
- U.S. Census Bureau. 2002. *Statistical abstract of the United States 2002*. Washington, DC: Government Printing Office.
- Wagner, T. H., S. Chen, W. Yu, and P. G. Barnett. 2002. *HERC's inpatient average cost data sets for VA care, Version 3: Fiscal years 1998-2001*. Menlo Park, CA: VA Health Economics Resource Center.