

Clinical Research

Issues in Data Collection

James N. Weinstein, DO, MS, and Richard A. Deyo, MD, MPH

This focus issue of *Spine* is dedicated to measurement of clinical outcomes. At the practical level this is hard to do; however, at a scientific level it is essential. Those involved in clinical research or interested in collecting data in their own practices struggle with decisions about which instruments to use in collecting data and when to use them. Even when the relevant outcomes of a treatment are identified, collecting that information remains problematic. To be most efficient at designing and collecting the appropriate data elements, careful thought must be given to the exact questions they are to address. The formulation of a good research question, up front, informs the clinician or researcher about the most appropriate data elements to be collected.

In a clinical trial, the ability to determine the better of two treatments is often the product of the trial hypothesis or question, the data elements chosen to evaluate the treatments in question, and the magnitude of change in scores over time needed to consider one treatment preferable to the other. Even with these clarified, the appropriate sample size must be determined, given the smallest clinically relevant difference in outcome and the variability in measures. This is necessary to have sufficient statistical power to determine whether the efficacy of one treatment is superior to the other. Having determined this, the clinician or researcher can then consider the many practical aspects of actually performing a study or sophisticated clinical trial. This article focuses on data collection and the phases in a study or clinical trial that necessitate major considerations over time. There is a paucity of surgical trials from which to glean perspective on how best to organize and run such a trial. A basic understanding of issues related to data collection will be useful to investigators considering a trial.

■ Data Collection Issues

Decisions regarding study design and data collection may be the most important in planning a trial. The study design is the investigator's map from which data collection follows and which enables the investigators to thoughtfully produce the necessary data forms (Figure

1). The data collection schedule and related forms are among the most important aspects of clinical research (Table 1).⁴ They determine the amount and quality of data gathered. Similar issues pertain to office practice. If the researcher understands office flow and can systematize the follow up processes, then his or her office can map data collection in less cumbersome but no less useful ways (Figure 2).⁶

■ Creating Data Forms

Before data collection begins, investigators must agree on the details of the data collection items and the process by which data collection will occur. They must identify who will ensure enrollment and accurate collection of the necessary data elements. The study design provides a basic blueprint for answering the question(s) or hypothesis to be tested but does not provide for the actual running of the study. Investigators must define the schedule according to which patients are to be seen throughout the study and outline the specific data elements to be collected each time the patient is seen. This schedule should be the same in the control and treatment arms.

Form development is often challenging. It should be undertaken by investigators experienced in form construction and familiar with the methods of data collection, data processing, and content necessary for the study at hand. The development of forms and content is often facilitated by review of the literature for instruments used in similar studies. The researcher might contact the instrument developer to obtain a user manual and ensure that the most current version is at hand. Investigators must have a thorough knowledge of the epidemiology of the disease (*i.e.*, back pain, disc herniation, stenosis) and associated conditions that are likely to influence outcome measures independent of the primary diagnosis under study. It is helpful to review those instruments, the ways that subjects complete questionnaires (phone, paper and pencil, or computer), and how this information is transmitted for storage and analysis.

Data form development is a collaborative effort among the investigators and often takes months of planning and preparation. The investigators should allow adequate time to conduct pilot tests of data forms. This can take several additional months, depending on the burden of data to be collected. It may be best to enroll at least 25 test patients. Pilot testing data forms with real patients allows investigators to react to suggestions from patients as well as from staff and personnel and provides more realistic estimates of data collection times.

From the *Center for the Evaluative Clinical Sciences, Dartmouth Medical School and Dartmouth-Hitchcock Medical Center, Hanover, New Hampshire; and the ‡Center for Cost and Outcomes Research, Departments of Medicine and Health Services, University of Washington, Seattle, Washington.

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (U01-AR45444-01A1) and the Office of Research on Women's Health, the National Institutes of Health, and the National Institute of Occupational Safety and Health, the Centers for Disease Control and Prevention (JNW).

Clinical Trial Processes

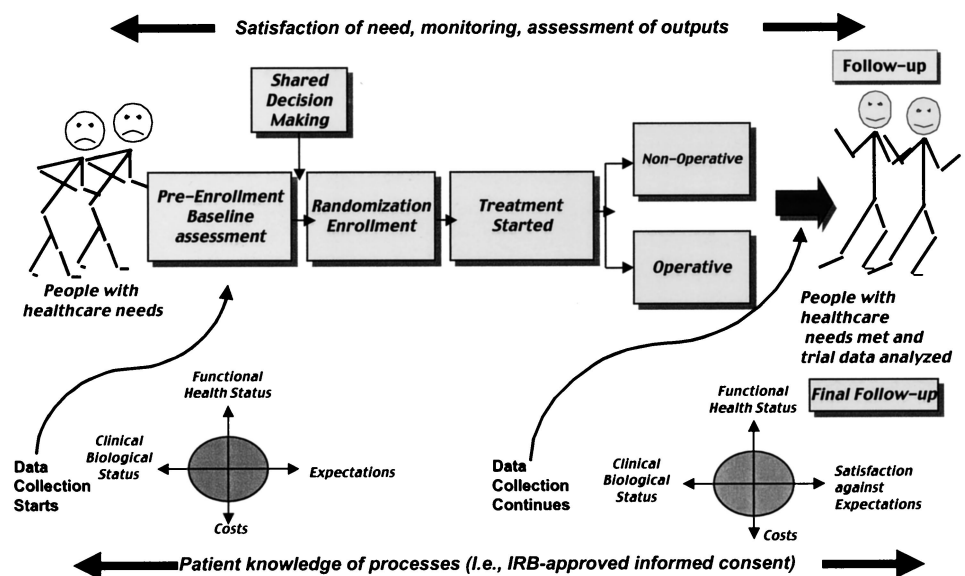


Figure 2. Clinical microsystem.

Batalden, Nelson. Adapted by Weinstein, with permission, 2000⁹

tered, number of patients randomized *versus* number observed).

8. Leave enough lead time (up to 1 year) for planning and development of the database system or project informatics.

In most cases, it is best to collect all the required initial data for a subject during a single clinic visit. Patients who are entering a study should be able to enter baseline data, be issued a treatment assignment, begin the study treatment, and make required follow-up arrangements in 1 day. The baseline clinic visit should:

1. Determine a patients eligibility for enrollment (inclusion and exclusion criteria)
2. Explain the purposes of the study to the patient, and obtain the patient's informed consent before participation in the trial
3. Provide baseline data (*e.g.*, generic [SF-36] or disease specific [Oswestry, Roland-Morris]) to asses changes after initiation of treatment
4. Result in a treatment assignment, after the necessary screening and diagnostic procedures to determine the patient's eligibility (history, physical examination, diagnostic imaging)


After enrollment, the patient generally makes a series of return visits, extending over a period of weeks, months, and even years, during which the assigned treatment is provided, and follow-up data are collected.

A potential study patient usually visits the office before randomization. This is referred to as the recruitment/pre-enrollment/randomization visit. Enrollment in the trial actually occurs at the time of randomization, which may be during the initial visit or a subsequent visit. This is marked by an explicit act of randomization which

traditionally involves opening a sealed manila envelope containing a treatment assignment. Today, a computerized randomization process is possible on site or through the Internet. After randomization, the patient is a member of a treatment group to which she or he has been assigned. Data collected at the time of prerandomization and at randomization are considered baseline data. Post-randomization visits are referred to as follow-up visits and are determined by the data collection schedule, which usually begins at the time of randomization.

Enrollment in prospective studies is nearly always slower and more difficult than investigators expect, because of several problems. First, although each exclusion criterion may seem likely to exclude very few patients, together the criteria often exclude more than anticipated. This is part of the problem of conducting fastidious clinical trials that enroll relatively homogeneous patients who are free of potentially confounding comorbid conditions and are likely to be available for follow-up. Second, a substantial number of patients refuse to enter studies, especially randomized trials. Perhaps most often, the patient simply does not want to be bothered with completing forms, making necessary visits, and taking the necessary time. However, some patients have strong preferences about receiving or avoiding one of the study interventions, in which case, refusal to enter is appropriate. Finally, treatment preferences aside, some patients simply are uncomfortable with being, as they see it, "guinea pigs."

Prudent investigators should anticipate enrollment problems by pilot testing study procedures to assess their acceptability. They must be prepared to extend the duration of enrollment, if necessary, to consider advertising the study in an appropriate manner, and even to consider



SPORT Screening Questions

SPORT

1. Have you ever been diagnosed with any of the following:
 - Herniated Disc
 - Spinal Stenosis
 - Degenerative Spondylolisthesis
2. Do you have low back pain which radiates into your buttock, thigh or leg? If YES, ask:
 - 2a. Do you have this pain when you walk?
 - 2b. Do you have this pain when you sit?
3. Do you have any loss of sensation or weakness in your leg(s)?

4. Have you ever had surgery on your back?
5. Are you pregnant?
6. Are you under 18 years of age?




Figure 3. Sport inclusion and exclusion criteria.

alternative or additional study sites. Charlson and Horwitz¹ have demonstrated that enrollment sites should have at least twice the volume of appropriate patients needed in the trial to be confident of enrolling the planned number.

Physician bias regarding randomization and enrollment can play a major role in the ability to recruit patients into a study. Depending on the physician's beliefs, patients may be inappropriately biased away from randomization.

Informed Consent

After IRB approval, there must be sufficient time for the informed consent process and assessment of the inclusion and exclusion criteria before patients are enrolled in the trial. The goal is the simplest procedure with the least nuisance and lowest risk to the patient. Eligible patients should be enrolled with minimal encumbrances, and ineligible patients should be spared the inconvenience of the more complex and time-consuming procedures of actual enrollment. A checklist of inclusion and exclusion criteria may make it immediately obvious to study personnel whether a patient is eligible. This also permits later tabulation of the reasons for exclusion and the number of patients excluded (e.g., sport inclusion and exclusion criteria; Figure 3).⁴

The authors prefer the term "informed choice," which implies that consent is not simply a one-sided decision but that the patient is a partner in the decision to participate in a study. Even if the patient refuses randomization, informed choice carries forward as the means by which patients and their doctors decide on treatment.^{2,3,5}

Follow-Up

Required visits should be scheduled at predetermined times after the randomization visit. Clinical trial visits are often scheduled to collect information on the treat-

ment process, complications, the outcomes measures (e.g., SF-36, Oswestry, Roland-Morris, EQ 5D, Health Utilities Index) and use of services (e.g., hospital admission, care by other providers, additional tests,) needed for evaluation of the specific treatment(s) under study. The timetable of these visits may be dictated by requirements for treatment and the expected time course of response. In today's world of managed care, follow-up for research endeavors has become difficult. Some visits not required by the protocol must occur on an *ad hoc* basis if a specific problem arises. In such instances, data collection for the trial is often unnecessary. Data collection should capture scheduled visits and distinguish those from nonscheduled visits.

Visit Intervals

Ideally, all visits, from baseline through follow-up, should occur at precise time points relative to the time of randomization. However, such precision, in the real world, may be neither possible nor necessary for the observations required in a typical research study. The usual approach is to have a scheduled visit within a defined interval of time on either side of the desired time. This window of time depends on the required data collection items, the system by which these data are to be collected, and timing of the visit. As an example, any visit between 2 and 4 weeks may be acceptable as a 3-week follow-up, and any visit between 11 and 13 months as 12-month follow-up. The coordinating center for the study should provide clinics with appointment schedules that indicate the preferred dates and permissible time windows of each required follow-up visit.

Data Requirements by Visit Type

The purpose of each visit must be outlined, and there must be agreement among the investigators on the specific procedures to be performed at each step. To guard

against errors, specification of the items of information collected during each clinic visit is essential. Some items appear only at baseline or the initial visit. Others appear under several categories and are repeated over time.

Ensuring complete follow-up is challenging in any study. Completeness of follow-up is important, because withdrawals may bias the study results. Results in several studies of back pain therapy have demonstrated that patients who withdraw from the study tend to have less favorable baseline characteristics than the patients who complete follow-up, potentially creating a falsely favorable outcome result. Substantial resources are required to assure good follow-up, and therefore time and expense should not be underestimated. Even if patients change treatment arms or withdraw altogether, sequential follow-up must be continued.

Several steps and design features are recommended to optimize follow-up rates.

1. Minimize the respondent burden by keeping follow-up questionnaires and tests to a minimum.
2. Exclude subjects with no telephone or who plan to move in the near future.
3. Collect contact information at baseline for relatives or friends who always know where the subject lives.
4. Update contact information at every visit or data collection.
5. Provide reminders for appointments.
6. Plan multiple efforts at phone contact, both during and after working hours.
7. Provide a financial incentive, to acknowledge the patient's investment of time and effort and to defray possible travel costs for visits.

With the patient's permission, a newer method of tracking is the use of information from the Internet, which often includes data on residence, home ownership, telephone numbers, car registration, and other public records.

In addition to baseline and follow-up patient data, information regarding treatment must be collected. In surgical studies this often requires entire sets of forms for characterizing the surgical process as well as the actual surgical procedure(s) (Table 1).

Follow-up forms should include information about treatment compliance, the exposure of patients to various operative and nonoperative treatments, and information and data about the extent of crossover treatment (receiving the unintended treatment). Operative patients also have nonoperative treatments that must be identified. Follow-up forms must also include data regarding side effects and complications of treatment (*e.g.*, monitored events), and whether they are related to the study treatment(s). Adjudication of these events may be necessary through an independent board.

■ Errors in Data Collection

One common cause of errors is haste in the development of data forms (*i.e.*, in identifying the data elements and in

the construction and testing of forms). Form development takes time and patience, and efforts to shorten these processes often lead to unwise decisions. An important problem is the desire of clinicians to create forms to meet the research goals of the study, but also to provide data for routine patient care. This is usually a case of trying to do too much. Certain measurements needed for routine patient care are not justifiable for research forms, and vice versa. The researcher should be clear about the data necessary for assessing the primary and secondary outcomes of the study without overlying the data necessary for day-to-day patient care.

In designing data elements and forms, it is necessary to perform an item-by-item review of each form against the study goals and objectives. Data items that cannot be justified should be deleted. Even published instruments must be checked for accuracy. Many instruments have multiple versions that vary in subtle ways.

■ Data Linkage

The data management group (those responsible for data retrieval and processing) needs to link records by using a unique identifier for each patient. For example, they might use the patient hospital identification (for purposes of confidentiality, patient identification can be deleted later) plus check digits that identify the patient, the center from which the patient comes, and the type of visit.

■ Data Storage

Finally, this valuable data bank must be kept in a locked, fireproof, protected location with adequate back-up. Data collection is too difficult and expensive to repeat, and patients' time is too valuable to have data lost or destroyed.

■ The Data Safety Monitoring Board

The data safety monitoring board (DSMB) is usually composed of five voting members from outside the study investigative group with experience in the relevant statistical and scientific issues. This committee reviews summaries of safety, the accrual and progress of the trial, the quality of the data, and blinded interim efficacy and effectiveness analyses and reports its findings to the principal investigator and the executive working group. Overall responsibility for interpreting data on adverse side effects is given to the data safety monitoring committee. This group reviews the data every six months and makes recommendations to the executive working group regarding actions to ensure that subjects are not exposed to undue risks. The mandate of the committee complies with the July 1, 1999, release of the National Institutes of Health (NIH) Policy for Data and Safety Monitoring, with its primary function being to "ensure the safety of participants and the validity and integrity of the data."

■ Steering Committee or Scientific Group: Study Coordination

The coordinating center for a given trial provides materials, training, and support for each of the sites. Major study decisions are decided by the scientific group, which meets monthly and consists of the principal investigator, physician investigator, coprincipal investigators, project coordinator, and others. Site investigators communicate with the coordinating center. The coordinating center brings relevant issues to the attention of the funding body, and the DSMB as needed. There must be no communication between DSMB members, site investigators, and other personnel except as otherwise stated in the DSMB charter. The DSMB has direct communications with the funding agency and can stop a trial when public health or safety is at risk.

■ Conclusions

Conducting a clinical study is a major undertaking, and the fiduciary responsibilities are weighty and extensive. It is difficult to gather the necessary data elements at the appropriate times while avoiding missing data. The idea of conducting good research is appealing, but the reality is often onerous. It is very different from searching a secondary database of previously collected information or conducting retrospective chart reviews, which are often subject to biased interpretations (by no fault of anyone involved). It is even more difficult to collect primary data in a prospective randomized trial according to a very strict protocol, wherein chance, bias, and confounding can be addressed. Data mining of secondary data sources and retrospective chart reviews do not interfere with clinical practice and minimize issues of informed consent or compliance with treatment. Good primary research calls for constant dedication by practicing phy-

sicians and patients willing to participate for the sake of knowledge and better treatment of future patients. These issues are not to be taken lightly. Asking the right question(s) is important, but getting patient and medical support is mandatory.

No matter how sophisticated the data elements and data collection systems, human factors make or break any good research effort. Prestudy planning is a large undertaking. Each person involved in each activity is part of the ongoing effort to prevent errors in data capture and transmission. The importance of these processes cannot be overstated. The initial time investment to make the best decisions possible and to consult with colleagues and experts is time well spent.

References

1. Charlson M, Horwitz R. Applying results of randomised trials to clinical practice: Impact of losses before randomisation. *BMJ* 1984;289:1282-4.
2. Deyo R, Cherkin D, Weinstein J, Howe J, et al. Involving patients in clinical decisions: impact of an interactive video program on use of back surgery. *Med Care* 2000;38:959-69.
3. Phelan E, Deyo R, Cherkin D, et al. Helping patients decide about back surgery: a randomized trial of an interactive video program. *Spine* 2000;26.
4. Weinstein J, Principal Investigator. *Spine Patient Outcomes Research Trial*. Hanover, NH: Trustees of Dartmouth College; 1999.
5. Weinstein JN. The missing piece: embracing shared decision making to reform health care. *Spine* 2000;25:1-4.
6. Weinstein JN, Brown PW, Hanscom B, et al. Designing an ambulatory clinical practice for outcomes improvement: from vision to reality. *The Spine Center at Dartmouth-Hitchcock, Year One. Qual Manage Health Care* 2000;8:1-20.

Address reprint requests to

James N. Weinstein, DO, MS
7251 Strassenburgh
Dartmouth Medical School
Hanover, NH 03755
E-mail: james.n.weinstein@dartmouth.edu