Neurosurgical data bases

MARY A. FOULKES, PH.D.

Biometry and Field Studies Branch, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland

Prospectively collected data bases are discussed with particular reference to neurosurgical data. The objectives, design, and collection of data bases are detailed. Statistical issues that relate to the analysis of these data bases are reviewed, with examples from the National Institute of Neurological Disorders and Stroke-supported Traumatic Coma Data Bank. The utility and limitations of data base inferences are considered.

KEY WORDS • data collection • Traumatic Coma Data Bank • prospective study

When a patient’s course following neurological trauma or diagnosis is the focus of an investigation, a prospectively collected data base is often the only feasible research resource. A well-designed and carefully collected data base permits evaluation of the clinical and neuropsychological course. Achieving the goal of a sizable and usable neurosurgical data base with appropriately defined and collected data is not a serendipitous occurrence. This report outlines the requirements for and the limitations of prospective neurosurgical data bases.

The evolution of medical knowledge, the understanding of disease progression or of recovery, and the identification and measurement of critical indicators of outcome all occur at varying paces. Medical practice and biomedical theory must attain an advanced level before it is feasible to consider clinical or epidemiological experiments, such as a randomized clinical trial. When medical knowledge with regard to a particular neurological disorder is still in its infancy, an appropriate tool to advance its evolution is the prospectively collected data base, such as the recently recommended data base on surgery for intractable epilepsy. The clinical course of a specific disorder can be characterized in detail and the time course of events such as complications or recurrences can be examined. The outcome, particularly in the short term, can be documented by clinical, neurological, or neuropsychological measurement.

A data base may be an appropriate resource in the evolution of medical knowledge even when it is believed that the natural history, time course, and outcome are known. Before a clinical experiment is planned, a number of issues must be resolved. These include diagnostic consistency, accuracy, and classification; patient availability; and the technology, timing, feasibility, accuracy, and precision of measurements. The measurements of interest would be baseline characteristics, clinical course, clinical management effects, or outcome measurements. For example, is it feasible to record blood pressures or blood gas values during transit from the scene of an injury to a trauma center? Are those measurements of comparable accuracy across transit services or across geographic locations? Is uniform calibration possible? In the National Institute of Neurological Disorders and Stroke (NINDS)-supported Traumatic Coma Data Bank (TCDB), for example, blood gas measurements are missing in approximately 75% of the initial examinations. Serum osmolality data are missing for virtually all TCDB patients. Serious lack of consistency or data availability requires either decisive corrective action, some extra incentive for providing the data, or elimination of the data items (with their associated costs) from the data collection planned. A clinical experiment can later be planned once these issues have been demonstrably resolved. Details of efforts within the TCDB to achieve consistency in computerized tomography (CT) scans included training sessions, workshops, and multiple reader consistency checks.

Data Base and Design Collection Methods

For this discussion, a data base will be defined as described by Byar:3 “a prospective registry of computerized information concerning patients with one or more diseases. Such a data base is likely to contain baseline information collected when patients are first observed, information about the treatments received, and follow-up information concerning the eventual outcome.”

Designing a data base to coincide with a population base has numerous advantages, but requires extensive effort by all involved. Data from a population-based cohort can generate population-based statistics, such as the annual incidence of penetrating head injuries or the proportion of all severe head injuries that result in irreversible coma. Without a population base, there
are no denominators for calculation of incidence figures. For example, the data available by census tract or other geographic unit on population characteristics would allow incidence estimates specific to age, race, and sex. Population data relating to environmental exposures, unemployment rates, legal limits for driving speed and drinking age, and even tax rates could be used in conjunction with the individual patient data. Population data are necessary for accurate international comparisons, as well. Therefore, it is desirable that, whenever possible, a data base be population-based.

The designation of patients to be included and to be excluded from the data base is both a scientific and an economic decision. For example, while it may be desirable to collect data on both moderately and severely head-injured patients, it may be economically impossible. The eligibility criteria should reflect that group of patients about which inferences will be drawn. The eligibility criteria used by Jennett, et al., may be compared to those of the TCDB: the former permits inferences only to those patients who survive in coma for 6 hours postinjury, whereas the latter does not impose the 6-hour limit. Another example of eligibility criteria comes from the Centers for Disease Control (CDC) recommendations for trauma registries, which are based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes. The eligibility criteria should be simple to apply, and all data items should be recorded in sufficient detail to permit verification of eligibility within the data base. The eligibility criteria for a data base are generally less restrictive than those of other prospective studies, in particular, randomized controlled trials. The eligibility criteria will ordinarily preclude data collection on nondiseased individuals of the type included in many retrospective studies, such as case-control studies.

Patient consent is ethically required, as with any clinical investigation. In cases where the patient is unable to provide informed consent, a responsible adult relative or friend should be asked to do so as their surrogate. The request for consent provides the opportunity to explain the objectives of the project; to itemize any extra procedures, special testing, or family interviews that may be a part of the project; to indicate the planned follow-up duration; to convey the safeguards of patient confidentiality; and to state the patient's right to refuse participation. This is an important aspect in gaining the permission and the long-term cooperation of both the patient and family, and can substantially affect later success in minimizing unknown responses, nonresponse, and losses to follow-up review.

The duration of the data collection is a function of the resources and of the required sample size, if a fixed sample size has been specified. If major hypotheses can be formulated in detail prior to initiation of the data base, these will dictate the minimum sample size necessary to achieve a prespecified level of statistical power. Independent of the sample sizes required to address these hypotheses, there is the maximum sample size that can be attained given the resources available. A data base planned for the largest feasible number of participating institutions and the shortest feasible duration of patient accrual would be preferable, considering elapsed time of data collection versus sample size, plus the advantages of replication, minimizing the effects of technological and staff changes, and broadening the generalizability of results.

Definitions must be developed with careful attention. First, an unequivocal, practical diagnostic definition specifies which patients are included in the data base. Arriving at this definition is not a trivial task, since a data base of this type is usually undertaken in the context of a neurological disorder for which a modest amount of knowledge exists relating to clinical course, outcome, or optimum patient management. The minimum technological documentation of the diagnosis must be delineated. For example, if CT scans are required, then institutions without ready access to such equipment cannot participate.

A sine qua non for full institutional participation is admitting (or approaching for informed consent) all patients who satisfy the eligibility criteria. It is essential that all eligible patients within a given institution be entered into the data base (or at least approached for consent to enroll) to avoid biases in patient selection. A more desirable situation would be to capture all such cases within a defined geographic population, as discussed above, which often requires cooperation across institutions. The problems resulting from underreporting or selection were discussed by Agran, et al., in relation to pedestrian and bicyclist injuries.

While prospective data collection for a data base does not necessarily prescribe specific treatment regimens or management, and does not preclude simultaneous participation in a randomized controlled clinical trial, it is useful for the participating physicians to agree in principle on a treatment and management strategy. This has the effect of imposing some level of homogeneity upon patient management and reducing the necessary data collection of details of patient management. Within the TCDB, for example, the participating physicians agreed that all postresuscitation survivors would have intracranial pressure (ICP) monitoring for at least 72 hours.

Data-collection plans are critical to the utility of that data base for specific analyses. Details of patient history, diagnosis, treatment, hospitalization, and outcome must be collected. It is essential that each and every data item be thoroughly scrutinized. Each data item must be unequivocally defined, available for most patients (that is, not missing), pertinent for later analyses, consistent with all other data items collected, and recorded on the data collection instrument in a manner that will minimize errors. The temptation is to collect considerably more data items than can be adequately defined, reliably ascertained, accurately recorded, efficiently processed, correctly analyzed, and unambiguously interpreted. In general, after a certain point, col-
Neurosurgical data bases

lection of additional data offers diminishing returns in accuracy relative to information in the total data base. The point at which data collection should cease is dependent on the resources, the data base design, and the hypotheses under study. A waste of resources by excessive data collection should be avoided from the onset. Pollock and McClain\(^1\) presented the CDC-recommended core data items for trauma registries as a suggested list of minimum data items to require. As an aid in data collection, Kondziolka and colleagues\(^2\) proposed a neurotrauma assessment record, recognizing that “inconsistent terminology, scattered localization of information, and failing to record data are problems inherent in the written record.”

The participating physicians must collaborate with the designer of data-collection forms and with the computer programmer responsible for data editing to ensure that the data collection will facilitate rather than impede the project. Programmers can assure that no records indicating pregnant males enter the data base, but more subtle and complex interrelationships between data items must be defined by the participating physicians. “Data quality is everyone’s responsibility.”\(^10\) It cannot be too strongly stressed that these issues must be resolved \textit{a priori} to avoid collecting massive amounts of meaningless data. The data collectors should not be required to do tasks to which a computer is better suited, such as to calculate body mass or sum component scores.

After the data forms have been designed, the physicians have defined all possible inter-item relationships, and the programmers have put all edit modes in place, it is time to test the feasibility of the system. This test should involve every participating physician, and each should complete data collection on a small number of patients. Once these test data are keyed and edited, all data management aspects should be thoroughly reviewed and critiqued, including the definitions of diagnostic and data items; design of forms; inter-item relationships as defined by the edit modes; the timing of data collection through the hospitalization course; and the timeliness of data collection and processing, and much more. Reliability between raters should also be assessed at this time. Numerous modifications will result from this process, and may require revisions and additional tests. Once the data collection process is finalized, all test data should be eliminated (not used for analysis).

\textbf{Conduct of the Study}

With the definitions, forms, and procedures finalized, the prospective data collection begins in earnest. Goldberg, et al.\(^12\) have stated that “There is a general tendency to underestimate the time and resources required to develop a functional registry.” The design of a data base should not be so ambitious as to tax the abilities and resources of the participants, either medical staff or patients, particularly with regard to long-term follow-up monitoring. Unless such follow-up review has routinely been recorded, or specific resources and commitments have been allocated to this task, it should not be attempted. It is a waste of effort to attempt a long-term follow-up review when less than 30% to 40% of patient follow-up data will be captured. The problems of inference based on selective follow-up monitoring on a small proportion of the patients have been discussed by Levin, et al.\(^24\)

The eligibility criteria for patients have been discussed, but there should also be eligibility criteria for the participants in the data base. There are certain institutional factors that are essential to successful participation in any collaborative venture and certain desirable characteristics of a principal investigator. The institute should foster the successful design, testing, data collection, and analysis by recognizing the long-term collaborative nature of such an endeavor, and in affording participants the time to complete careful data collection and responsible analyses. As a part of this proposed education in outcomes management, Langfit\(^23\) recommended that students “learn how to enter clinical information into the data base” in their 1st year in medical school. The institute should value the individual investigator’s contribution to the neurosurgical data base without demanding premature presentation and publication of results. The institute must also provide the technical staff and resources, from CT scanners to microcomputers. Many institutions have highly competent nurses, medical technologists, and clerks, but none of them may be capable of functioning as data entry personnel. This function may be new to the participating institution, which may require assistance in acquiring appropriately trained staff or in retraining existing staff.\(^14\)

The individual investigators must be willing to devote the time and energy necessary to lay the groundwork. As mentioned earlier, the investigators must be actively involved in defining the data to be collected, in defining the logical consistencies between data items, and in testing the procedures as outlined. The investigators must be diligent throughout the conduct of the data collection to insure consistency, timeliness, and accuracy, and must maintain that level of diligence throughout the analysis of these data. Since these data usually will be collected in the absence of a rigorous experimental design, the results may be open to a variety of interpretations, and the investigators must be open to multiple interpretations of the data. In addition, the group of participating investigators should be sufficiently heterogeneous in their medical philosophy and training to promote the introduction of a variety of interpretations of the results.

Incoming data must be carefully reviewed on a continuous basis during the collection period. The timing of data submission for key entry and processing is an important indicator of the participation. If data from a particular physician or institution are delinquent (the definition of delinquent being previously established),
that may be an indication of lack of resources, a decline of interest in participation, or infeasible data collection requirements. In any case, such delinquency requires immediate action, and can only be detected in a timely fashion by consistent monitoring of performance. Identifying the technician, physician, or institute contributing each data item affords the analyst the option to stratify by contributing physicians or institutes. For some analyses, the data from a single contributing physician, technician, or institute may be excluded but, for many analyses which rely on the population base, this could bias the results when the individual or institute in question managed a specific subset of patients. Another indication that data from a particular individual should be excluded from analyses would be when there is evidence of falsification of data. Schey and Davis\(^3\) provided a mechanism for detecting such occurrences in multiple readings by examining frequency distributions.

Prior to initiating data collection, plans should be outlined for the completion of the data collection and the closure and archiving of the data base.\(^6\) The optimum method for ending the patient follow-up period has been debated in the arena of clinical trials: whether to follow each patient for a fixed duration or to follow all patients to a fixed last date. In the context of a data base, the issue is not one of experimental design, but rather one of fiscal limitations and the research questions addressed. Therefore, the practical stopping rule would be to end patient accrual on a specific date, and to similarly end patient follow-up review on a specific date (although not necessarily the same date). A data base should have a flexible plan to accommodate changes in technology over time and to expand the data items recorded. If the initial data base design is static, changes will be difficult if not impossible.

Lastly, the analyses planned should be outlined in detail at the onset of the data base. This does not mean that all such analyses will be known or anticipated at the onset, but that the design of the data base, the data item definitions, the edit modes, the data files, and the software development or acquisition should take into account the planned analyses. An essential participant is a statistician who is familiar with the analysis of observational studies, is willing to modify or purge any inappropriate or impossible analytic plans, assists in defining edit modes, and collaborates with the physicians on analyses and interpretations that do not exceed the scope of the data. A clear publication policy should be defined and agreed upon by all participants, including authorship, writing responsibilities, and internal review procedures. The policy for data access by non-participants should also be defined in advance.

### Statistical Issues

There are a variety of statistical issues that bear on the analyses of these data; only a few such issues will be briefly discussed here. In the analysis of observational studies or of experimental studies, these issues are neither original nor insoluble. A major problem in the analysis of observational data derived from data bases arises in recognizing these statistical considerations where appropriate, which can be difficult even for an experienced analyst.

A recurring concern for the analyst dealing with a neurosurgical data base is the issue of treatment evaluation. As mentioned earlier, the prospective data collection for a data base does not prescribe specific treatment regimens or management; however, some details of the patient management are usually recorded. Given the data collection, it is possible to classify patients into those treated and those not treated with a particular intervention. The analyst must be continuously vigilant to avoid comparisons of such patient groups.\(^5,7\) Patient groups defined by an intervention should be similar in every respect, save that intervention, in order that outcome effects can be attributed to that intervention. Unfortunately, in any nonrandomized comparison, this similarity can never be demonstrated since the patient groups can always differ in some unmeasured characteristic. Significant treatment effects may be due either to the intervention or to those unmeasured factors. Anderson, \textit{et al.},\(^2\) stated that "Without randomization, there are often many alternative explanations of observed results." The same problem occurs when those patients treated by the new method are compared to those treated by the old method (historical controls). Numerous investigations\(^26,34\) have shown that this design results in a biased conclusion in favor of the new therapy.

Biases can also be introduced by variation in observers (clinical staff, evaluators, data collectors), across institutions, over time, within institutions, and in patients. Observer bias can be the cause of a perceived change in a patient's status over time, when no objective change in the patient's status took place. Knowledge of the patient's risk group, medical history, or treatment can also influence the observers and introduce bias.\(^13\) Such bias can be minimized by utilizing blinded evaluations whenever possible, by maintaining stable, well-trained staff with coordination across institutions, and by minimizing the duration of data collection. Selection bias could affect who is referred to the participating institution, who enrolls, and who receives a particular intervention.\(^5,7\) For example, Lou and Jankovic\(^27\) discussed the potential for selective referral to their movement-disorder clinic and thus to their data base on essential tremor. Designing a population-based data base and renouncing inappropriate treatment comparisons will help to avoid these selection biases. Ascertainment bias can also be affected by interventions that may result in differences in the timing and intensity of monitoring patients. In this case, whether the response is observable is not independent of the intervention chosen; examples of this include multiple early determinations of Glasgow Coma Scale\(^36\) scores for patients undergoing immediate surgery, long-term ICP measurements for pa-
Neurosurgical data bases

tients monitored for only a short time, and Galveston Orientation and Amnesia Test results for patients discharged in coma. These biases are important to consider, since they limit inferences that can be drawn from these data.

Since a data base is often initiated without a primary hypothesis, there is no single overriding response or outcome measure. The data base then contains a number of variables which in any given analysis could serve as an outcome measure. Multiple responses recorded on a single patient are not independent, and must be treated as such in any analysis. The analyst must rely on multivariate analyses or repeated-measures analyses, ensuring that the correlation structure assumed in the analyses is appropriate in each case. The results of analysis will be at best inefficient and at worst incorrect if the correlated nature of these data is ignored. An example would be to incorrectly analyze multiple ICP measurements for the same patient as independent observations, rather than correctly analyze these as a nonstationary time series. Longitudinal data can be either dependent responses with repeated measurements, or independent factors that should be analyzed as time-dependent covariates.

A major difficulty in analyzing these data is the factor of missing data. Missing data can be ignorable (ignoring completely at random) or not ignorable.19 Ignorable missing data can pose a difficult problem in relation to the power of hypotheses tests by reduction of sample size. Most missing data within a data base are not ignorable. In the case of nonignorable data, when the probability of a non-missing response depends on the value or level of the response, analysis that ignores missing data will yield biased estimates. This problem is especially difficult as repeated responses over time approach a nadir or zenith and the probability of a non-missing response declines.20 An example of this occurs when head-injured patients refuse follow-up review due to resumption of employment. In general, imputation of missing values based on any of the variety of imputation methods is not a viable solution. The best solution is a preventive one that assures a minimum of missing data. These missing data problems must be carefully considered, but must also be distinguished from the problem of structural zeros within the data where one would not expect a valid response. Structural zeros should be those situations that data editing should prohibit, but might require special handling in analyses; for example, data on therapeutic side effects showing patients recorded with fatal alopecia violate a structural zero.

Pre-existing conditions and comorbidity that may relate to outcome should be recorded in detail; however, it is not always possible to anticipate at baseline the conditions that should be of concern. If pre-existing conditions or comorbidity are recorded at baseline (which may be defined in a given data base as date of onset, admission, diagnosis, or injury), they can distinguish patients who should or should not be included in cohort incidence analyses. When the pre-existing conditions or comorbidity in question was not definitely established at baseline, post-baseline analyses cannot distinguish between pre- and post-baseline cases. Determining prior seizure status for severely head-injured cases is an example. Psychiatric, developmental, or neurodegenerative disorders have been suggested21 as comorbid conditions to be recorded in a data base covering patients with surgery for intractable epilepsy.

When eligibility conditions, for either entering the data base or defining an analytical subset of patients, relate to an extreme value cutoff point on some response, then that response would tend to be less extreme if remeasured later. This phenomenon, known as "regression to the mean," was described by Sir Francis Galton22 regarding his genetic experiments. If only patients with extremely low initial pO2 were selected for a particular analysis, regression to the mean would occur. However, if the average of two or more measurements per individual is used, the regression to the mean will be reduced. Efron and Morris23 illustrated an estimator for prediction of future observations which could be used to define eligibility criteria, and which does not suffer from regression to the mean. The analyst must also be wary of regression to the mean in comparisons of change in a response to its initial value.

Misclassification or measurement error can plague data bases just as any other data-collection process. The results of nondifferential misclassification in both factors in a hypothesis-testing situation will always be in one direction; that is, toward the null hypothesis.24,25 This occurs either by reducing observed strengths of association or by reducing observed group differences and consequently increasing the chances of a Type II error. Thus, misclassification should be avoided from the onset by relying on consistently applied, clearly understood, easily implemented, and routinely monitored data definitions and procedures. Continual monitoring of incoming data should focus on possible misclassifications or secular drifts in the application of data definitions. Even with these precautions, publication of results should indicate the possibility and effects of misclassification. Where appropriate, models that explicitly recognize the presence of measurement error in the independent variables should be used.

The data available for analysis can be censored, either left- or right-censored. Censored data occur when the data beyond a specified point in time are not observed; for example, in the TCDB, the duration of recording ICP data was truncated (left-censored) at a maximum duration of 240 hours even though ICP monitoring may have continued beyond 240 hours. Most data are recorded such that a unique censoring time is known for each patient; that is, the data are multiply-censored. The incidence of singly-censored data, with the same censoring point, is infrequent but could occur due to administrative censoring, such as ICP monitoring beyond 10 days within the TCDB. The statistical literature suggests analyses of censored data beyond the usual
censored survival time analyses that include modeling a risk function for censoring, sensitivity analyses, and competing risks analysis.39

Numerous additional statistical issues face the analyst in appropriately interpreting the information provided by a database. Some of these include the problems of multiple hypothesis tests, multicollinearity, or misspecification in model building. The nature and complexity of these issues preclude a detailed discussion here, but the variety of statistical issues affecting analyses and interpretation indicate the need for experienced analysts who can anticipate problems and deal appropriately with these data.

Purposes and Limitations of the Data Base

A neurosurgical data base is a valuable resource for generating hypotheses to be investigated more fully with specifically designed independent studies. Since the analytical results may support a variety of interpretations, some of the hypotheses generated may relate to distinguishing between these various interpretations. The data base may generate hypotheses that are perused at great expense and effort but are merely constructs of the data values recorded and the manner in which data were collected.

A data base can also provide valuable information for planning future studies, beyond the data definition and feasibility stages. For the institutions participating, precise estimates of future patient accrual rates for specific eligibility criteria can be provided. This is particularly important, for example, in demonstrating the availability of racial/ethnic minorities or the availability of patients with rare exposures. Sample size estimates and accrual rates can be calculated with greater reliability given the information from the data base.

A neurosurgical data base, if properly planned, can permit verification or refutation of historically held views on diagnosis or natural history. Note that this is not the historical controls issue. A conjecture put forth by a neurologist or neurosurgeon in the years or decades prior to the initiation of the data base can be tested; such a study might be the role of alcohol abuse in exacerbating the consequences of severe head injury.33

As stated earlier, a neurosurgical data base such as that discussed here cannot serve the function of a randomized controlled clinical trial. It cannot be used to address a number of questions that require rigorous experimental conditions mandating treatment or management to provide answers, such as the optimum duration of ICP monitoring. However, a neurosurgical data base can provide a substantial amount of the groundwork necessary to adequately and economically plan for such experiments. A neurosurgical data base, when appropriately collected and used, can also provide the investigator with a tool to organize clinical impressions, to review series of patients in a consistent manner, and to generate hypotheses that form the basis for the next level of scientific investigation.

Acknowledgments

The author acknowledges the extensive insight and experience shared by the NINDS TCDB investigators: Howard M. Eisenberg, M.D., University of Texas Medical Branch; Lawrence F. Marshall, M.D., University of California at San Diego; John A. Jane, M.D., Ph.D., University of Virginia; Harold Young, M.D., and Anthony Marmarou, Ph.D., The Medical College of Virginia. The author appreciates the numerous helpful discussions with Jonas H. Ellenberg, Ph.D., and James M. Dambrosia, Ph.D., of the NINDS.

References

Neurosurgical data bases


Address reprint requests to: Mary A. Foulkes, Ph.D., Biometry and Field Studies Branch, National Institute of Neurological Disorders and Stroke, Federal Building Room 7A08, Bethesda, Maryland 20892.