THE NATIONAL HOSPITAL SURVEY OF DISEASE: THE FEASIBILITY STUDY
Frederic D. Weinfeld, The National Institutes of Health

Introduction

Valid national morbidity statistics are lacking for a majority of the neurological disorders, as well as for other diseases. Most current national estimates of health statistics are either based on informed guesses, or on precise studies conducted in a few limited communities, or based on censuses conducted in a few large geographic regions that have been chosen for the availibility of their data rather than for their representativeness. Therefore, the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) of the National Institutes of Health (NIH) undertook a number of studies to develop the survey methodologies needed for collecting national data from hospital records and from other medical providers. NINCDS then conducted a series of national surveys under contract which demonstrated the practicality of these methodologies for collecting clinical information and for deriving national estimates of the incidence and prevalence of a number of neurological disorders, including stroke, multiple sclerosis, and head and spinal cord injuries (1,2,3,4).

Since single-time surveys are expensive and consume years of effort to obtain estimates for only one point in time, we have applied the knowledge from our previous medical surveys to the design of a comprehensive data collection system. This system, called the National Hospital Survey of Disease, would collect data for the estimation of incidence and prevalence of a wide range of disorders and would also provide the needed trend data. The plans were developed in a collaborative effort with the National Center for Health Statistics (NCHS).

The major advantage of this continuing system is that it would be multi-purpose, thereby eliminating the need for separate surveys for each disorder and for each new time period. The overall savings in effort, money, and hospital respondent burden are so considerable as to prod us to endeavor to develop such a system. Perhaps of even greater import is that the system could be used by other Institutes in NIH as well.

The first major step in implementing this system has been taken with the completion of the Feasibility Study, which was a pilot for the national survey. The study investigated problems in the collection of data on diseases for which short-stay hospitals are an important or primary source of care. The study also examined the possibility of using the Hospital Discharge Survey (HDS), conducted by NCHS, as the vehicle to collect the disease data from hospital records. The HDS collects yearly data on a sample of approximately 224,000 discharges from 420 short-stay hospitals. Using the HDS sample of hospitals would obviate the need for continued recruitment of hospitals for every new disease survey. However, in order to collect the individual patient information needed to produce the required medical statistics, it would be necessary to supplement the information currently being collected in the HDS. For the HDS was designed to produce statistics based on the discharge event and not incidence statistics based on individual patients. The additional items of information required would depend on the clinical algorithm for each individual disease and the counting rules adopted in the study.

The Feasibility Study

The basic objective of the Feasibility Study was to determine whether it is practical to utilize an augmented Hospital Discharge Survey to obtain national data on the incidence and prevalence of a number of neurological and other disorders on a periodic basis from hospital records. More specifically, Its purpose was to find answers to a large number of methodological, operational, and medical questions so as to provide the basis for the design of the national survey and for the planning of other studies utilizing hospital records.

In general, the methodology consisted of selecting a sample of patient records classified to certain diagnostic rubrics in a sample of hospitals. For each sample patient, information concerning all other hospitalizations within a time period was abstracted. Using a clinical algorithm, a determination was made whether the case met the criteria for inclusion in the study. Questionable cases were reviewed by physicians who also conducted a quality check on a sub-sample of records. The determination of whether the case was to be included in the incidence count was made using the counting rules.

The study was conducted in 27 short-term hospitals in the eastern United States; half of these hospitals were participants in the Hospital Discharge Survey. A sample of

approximately 2,200 patient records covering the 15 selected diseases included in the study were abstracted from the hospital records. The records were identified from the diagnostic index in each hospital and also by an alternative procedure in the HDS hospitals which used the diagnoses coded on the HDS abstract form.

Since the Feasibility Study was conducted on a large scale, it was possible to test many operational procedures and alternative methodologies which other more limited studies are unable to do. The study investigated the feasibility of various design alternatives and compared the cost effect, as well as the sampling and non-sampling effects of several design factors. It focused on augmenting the HDS so that it would be capable of producing incidence, prevalence, and other morbidity statistics for selected diseases without jeopardizing the discharge statistics currently being compiled by NCHS. Topics investigated included:

1) Methods of soliciting the cooperation of hospitals,
2) Data collection procedures for abstracting information from hospital records,
3) Disease algorithms to determine the cases which meet the study criteria,
4) Counting rule algorithms to unduplicate the case count, and
5) Methodology for producing selected morbidity statistics, including estimates of hospital disease incidence, prevalence, and care.

Findings

Some major areas of investigation deserve comment:

1) Hospital stays and search codes.
There are many considerations involved in a study using hospital records. There is usually more than one diagnosis recorded for each patient on the discharge record. Hospitals maintain listings of patient discharges by diagnosis and patients are listed under each of their recorded diagnoses. The primary diagnosis is not notated and so some of the diagnoses may be secondary conditions while others may be old conditions unrelated to the current hospital stay. Records of hospitalizations are usually identified by serial code numbers and the identification of a patient may be by a serial number, a unique number, or by some other combination. Cross-referencing of different listings is therefore necessary to locate all hospitalizations of a patient in a hospital.

For most diseases, a case may be classified into one of many diagnostic categories according to the specific manifestations of the disease. Therefore, a disease may require a search of many categories to retrieve all the cases. However, the yield of acceptable cases found within the different categories varies considerably, as does the number of cases. Since it is quite costly to abstract a case, the number of unacceptable cases was minimized by assigning each of the diagnostic search codes to different sampling strata by according to its estimated size, yield of cases, and effect on the incidence estimate. In addition, hospitals use several different classification systems so that the different diagnostic rubrics must be matched. (The standard classification system is the International Classification of Diseases - Adapted.)

The Feasibility Study developed and tested operational procedures to locate all patients indexed to a certain medical condition, to sample these patients, to locate the previous and subsequent hospitalizations for the sample cases and to retrieve the records for abstracting and copying. The study also determined the yields for specific search codes for diseases in the study. This information will be most valuable in the future for determining the sampling ratios to be used in a national study.

2) Clinical algorithms.
Reliance could not be put on the diagnostic listing for determining acceptable cases of a disease because the listed code might not be the primary diagnosis or the case might not conform to our medical criteria, in addition to the strong possibility of coding errors. Therefore, detailed clinical algorithms were developed for each disease to determine a valid case. The algorithms for the neurological disorders were developed by a panel of neurologists and the other disease algorithms were developed in cooperation with participating NIH Institutes. These algorithms specified the medical symptoms and conditions necessary for classifying a case into one of several certainty levels, ranging from "Definite" to "Undocumented'. As might be expected, the study pointed out revisions needed for some of the algorithms and coding instructions.

3) Unique and multiple counting rules.
A decision was needed on what counting rule should be used in the main survey. Such a rule is needed to deal with the likelihood of several separate discharges being recorded for a patient during the study period, all of which constitute the same medical event. For

example, cases may be transferred from one hospital to another for special tertiary care or cases may be discharged from more than one hospital for the same condition.

A counting rule specifies the conditions under which an acceptable case is eligible to be counted. One can adopt a unique rule that counts the case only when one uniquely identifiable discharge falls in the sample, so that no person is eligible to be counted more than once. Or one can adopt a multiple, or multiplicity, counting rule that apportions the count among all the separate discharges of the same event and specifies the appropriate fraction be added to the total count. Thus, if there are three discharges for one patient that could be encountered in a survey with equal probability, all of which constitute the same event, then, if one of the three discharges falls into the sample, the multiplicity rule specifies that it shall contribute only one-third to the accumulating total.

Both the unique and multiplicity counting rules can be designed to yield unbiased estimates but the two methods are subject to different sampling variability and costs. A multiplicity counting rule may have advantages when the event to be counted is rare - hospital stays for several of the diseases in the Feasibility Study occur only about once in 10,000 hospital stays. Another possible advantage of this rule is that it may pick up more cases of hospitalized incidence or prevalence per dollar of direct data collection costs than a unique rule. However, even though the error variances using the multiplicity rule may be smaller, this advantage is outweighed by the fact that the estimate of the multiplicity factor is frought with error since it is difficult to ascertain complete information about other hospitalizations from the hospital records. The unique counting rule is operationally simpler since it does not require abstracting any stays occurring after the sample stay in the same year, while the multiplicity rule requires abstracting all the stays in the year. Therefore, it was decided to compare the efficiency of the two rules.

An eligibility rule was also used. It specified that a case which was transferred in from another hospital was included in the study; whereas, a case which was transferred out of the hospital to another hospital was excluded. Also, a case which was admitted and discharged in different years (during the December-January time span), was counted in the year of discharge.

The results from the study show that the estimates of incidence and prevalence and the cost-effectiveness ratios from the unique and multiplicity rules differ very little from one another and depend upon the pattern of hospitalization for the disease under study.

4) Determination of incidence and prevalence.

The purpose of the data system is to produce estimates of hospitalized incidence and prevalence. Hospitalized incidence was defined as the number of different persons who, during a particular calendar year were discharged at the completion of their first stay during which an acceptable diagnosis for the particular disease was recorded. Therefore, it is required to know that a stay is indeed the first "diagnostic stay", not only in the sample hospital, but also in other hospitals not in the survey.

The important unresolved issue which needs further investigation is whether the information in the medical records on previous stays in other hospitals is complete and accurate. The present study documented what was recorded in the record on stays in other hospitals but there was no check on its completeness. This is a critical aspect for either counting rule because if prior stays in other hospitals are omitted, it would introduce uncertainty into the estimates.

Conclusion

The findings from the Feasibility Study have been encouraging. It is possible to collect medical information from hospital records and produce valid morbidity statistics using HDS hospitals and diagnostic listings. Also, the collection of data on neurological disorders can be successfully merged with the interests of the Cancer and Heart Institutes. For almost all of the major questions asked, the study yielded fairly conclusive evidence. The solution to a few problems were not completely resolved and an ancillary study might be needed before a national survey is undertaken.

REFERENCES

1. Anderson DW and McLaurin RL (Editors): Report on the National Head and Spinal Cord Injury Survey. Journal of Neurosurgery 53 (suppl): S1-S43, November 1980.

2. Baum HM and Rothschild BB: The Incidence and Prevalence of Reported Multiple Sclerosis. Annals of Neurology 10: 420-428, 1981.

3. Weinfeld FD (Editor): The National Survey of Stroke. Stroke 12 (suppl): I1-I91, March/April 1981.

4. Weiss W and Weinfeld FD: Large Scale Population Surveys. In Schoenberg BS (ed). Advances in Neurology. New York, Raven Press, pp. 83-92, 1978.