

In recent years, so-called record-anchored and record-linked studies have extended the potential utility of vital records for research. Two studies on the etiology of blindness, and survivorship among the blind, using the record-linked approach illustrate advantages and disadvantages of vital records for research in public health problems.

THE USE OF VITAL RECORDS FOR BLINDNESS RESEARCH

Irving D. Goldberg, M.P.H., F.A.P.H.A.; Hyman Goldstein, Ph.D., F.A.P.H.A.; Dana Quade, Ph.D.; and Eugene Rogot, M.A.

THE USE of vital records, particularly birth and death certificates, for health research has a long history. In recent years, the National Vital Statistics Division has expanded the research potential of vital records by what is referred to in this paper as vital record-anchored studies in which a universe is defined by a population of registered vital events. In such studies, a sample of vital events is usually taken and, for these events, attempts are made to secure supplemental information by survey methods utilizing mail inquiries, telephone contacts, and/or personal interviews.

The main types of vital record-anchored studies are three in number: (1) the retrospective or follow-back study that is illustrated by the "Pennsylvania Mortality Study" reported by Haenszel, Moriyama, and Sirken¹ designed to secure supplemental information from the physician, informant, and funeral director; (2) the cohort or follow-up study which prospectively collects information on two or more occasions for a group of persons with some common characteristics whose vital records have been selected for study, such as follow-up of the status of a sample of births by contact with the family, physician, hospital, etc.; (3) the record-matching study in which the person's

vital record is matched by other records pertaining to him, such as another vital record, census enumeration, case registration documents, hospital or clinic records, and so forth. This is illustrated by the ongoing study of Hauser and Kitagawa² in which death certificates are matched with census schedules to obtain a comprehensive picture of mortality differentials among the various areas and subgroupings of the population in the United States in relation to personal, social, economic, and housing characteristics.

There are, of course, numerous advantages in the vital record-anchored method. However, there are occasions where instead of a universe of vital events, one has a population of cases known to be afflicted with a specific disease or impairment, such as a register of reported cases or a particular hospital or clinic population. In this kind of situation, one possible objective of study may be the survivorship of the sample studied. Supplementary information may then be needed from vital records. Investigations of this type are designated in this paper as record-linked studies as distinct from vital record-anchored ones. In record-linked studies, vital records may be used in at least three ways: (1) for analysis of the information they con-

tain; (2) to secure information as to sources of additional data; and (3) as a device in the selection of controls.

As in the case of vital record-anchored studies, there are three main types of record-linked studies: (1) retrospective or follow-back study as exemplified by the retrospective study of cerebral palsy by Lilienfeld and Parkhurst³ where the birth certificate was used not only as a means for selecting the control group, but also for providing the basic information on factors of pregnancy and labor for analysis; and a series of studies by Pasamanick, Lilienfeld, and others⁴⁻⁸ where controls were selected from the universe of birth certificates by matching with the cases on specified variables, and where the principal data for analysis were secured by recourse to hospital case records; (2) cohort or follow-up study, where persons selected for study from the universe under consideration, as for example, cancer cases or blind persons, are followed up prospectively in order to secure information on survivorship status and causes of death by checking against death certificates; and (3) record-matching study where starting with a sample of nonvital records, attempts are made to match these against a set of vital records for supplementary information, as was done by Frazier and his associates⁹ in a study of cigarette smoking and prematurity. Of course, in many studies nonvital records are checked against other nonvital records (e.g., studies of association of diseases where a register of one type of disease is checked against a register of another type).

In this paper, an attempt is made to show how, starting from registers of blind persons,* two record-linked studies

* Starting with a community-wide register of cases, rather than an institutionalized population, may have the advantage that cases of all types of severity are likely to be sampled. This may not be true of cases receiving institutional care.

have been undertaken: one of these is retrospective and secures supplementary data from birth certificates in an attempt to determine the relationship between the occurrence of blindness in children and factors in pregnancy, labor, and delivery; the other is a cohort or prospective study and secures supplementary data from death certificates in an attempt to determine survivorship and causes of death as they relate to a blind population. These studies are still in progress.

Study of Prenatal Factors Associated with Blindness

In her report of a study on causes of blindness among school children, Kerby¹⁰ stated that throughout the entire series of cases studied “. . . it has become increasingly evident that factors affecting early development of the eyes constitute the most important problem in preventing blindness in children.” Her study included more than 4,400 school children from kindergarten through high school grades in 1954-1955. The cause of blindness was presumed to be due to unspecified prenatal causes in 42 per cent of the study group and to genetic origin in an additional 14 per cent. Excluded from these percentages are cases presumed to be due to specified prenatal infections (e.g., maternal rubella), and of course cases of retrolental fibroplasia caused by excess oxygen after birth.

Because of the apparent importance of prenatal factors as a cause of blindness, and also because of the apparent lack of intensive studies to identify such factors, it was decided to initiate research to explore this subject.

One direct approach would have been to follow a group of pregnant women through pregnancy and observe their offspring over a period of years. However, such an approach was completely impractical in view of the long period of follow-up necessary, and the large num-

ber of pregnancies that would have to be studied because of the relatively low incidence of blindness. Therefore, the approach taken was a retrospective one similar in many respects to that employed by Lilienfeld and Parkhurst³ in their study of cerebral palsy in upstate New York.

The present study's objective is to determine whether mothers of blind children are characterized by an excess of selected disorders of pregnancy, labor, and delivery as compared to mothers of the total population of live births surviving the neonatal period.

Using a retrospective approach, this study starts with a group of affected children born in New York State during a specified period of years. The birth certificates of these children are searched for information on complications of pregnancy and labor, which appears on the confidential medical supplement of the certificate. The relative frequency of specific complications for the case group is then compared with that obtained for a control group. The latter is a stratified systematic sample of the certificates of all recorded live births in the state over the same time period.

It surely would have provided more complete and accurate information if the data on pregnancy and labor had been obtained from the actual hospital records,¹¹ but for reasons indicated below this was not considered feasible. However, an attempt will be made to review hospital records for a sample of cases and controls.

In designing this study, four concerns were paramount: (1) The cases should be classified blind according to a commonly accepted definition; for this study blindness was defined as visual acuity not greater than 20/200 in the better eye with best correction, or field of vision restricted to 20 degrees or less in the widest diameter. (2) Information as to the cause of blindness of the cases should be available, so that those with

well-understood etiologies might be excluded from the study group. (3) The study group should be sufficiently large for analysis yet, at the same time, reasonably homogeneous in that all cases were born within a fairly recent time period and within a limited geographic area. (4) The birth certificates for the study group should contain, in a reasonably uniform manner, the necessary information relating to the prenatal and delivery period.

These conditions for the study were satisfied by records available in the State of New York. The New York State Commission for the Blind has maintained a register of blind persons since 1913, and reporting of blind persons to the commission became mandatory by legislation in 1945. The legal definition of blindness in New York State is essentially that given above. All cases are coded by the commission according to the Standard Classification of the Causes of Blindness.* With the cooperation of the commission, records were found of some 800 registered children who were born during the 13-year study period from January 1, 1948, through December 31, 1960, and whose blindness was coded as due to unspecified prenatal influences, or of hereditary origin, or of unknown cause. The birth certificates of some 600 of these children were located in the files of the cooperating New York City and New York State Departments of Health. From these certificates, and their confidential medical supplements, items of information abstracted included date and place of birth, sex, parity, maternal age, birth weight, specific complications of pregnancy, labor and delivery, Rh factor, operative procedures, birth injury, and congenital malformations. Information about plural-born blind children has

* The classification, which was prepared by the Committee on Statistics of the Blind, National Society for the Prevention of Blindness, has been in use (with revisions) since 1940.

been collected as a matter of incidental interest, but the study group is limited to single births.

The control group is a stratified systematic sample of three out of every 1,000 live births recorded in New York State during the study period, selected as follows: The birth certificates in the New York City and New York State Departments of Health are filed in numerical sequence. Starting with each certificate whose number ends in the three digits 162 which was randomly selected, and proceeding sequentially, there were chosen the first two certificates on which a complication of pregnancy was recorded, and the first certificate on which no complication of pregnancy was recorded. However, certificates for multiple births were excluded, as were those which represented court cases where the original certificate had been replaced by an amended one which had no medical supplement. From the sample of births so selected the neonatal deaths were removed; this was readily accomplished for New York City births, where the fact of infant death is recorded directly on the birth certificate; but, for upstate New York this required a separate search of available punch cards for infant deaths (beginning with 1950) or of the death certificates themselves (prior to 1950). Thus, more certificates were sampled with recorded complications of pregnancy than without. This was done in order to obtain a sufficient number of some of the rarer complications among controls. From the number of certificates which had to be skipped over in order to find complications it is possible to estimate the frequency of complications in the general population of births. It may be noted that this set of controls can be used again for additional studies of this type in New York State where the condition of interest is not necessarily blindness.

One problem encountered was a great

difficulty in deciding on the spot, for many certificates, whether or not a complication of pregnancy had in fact been recorded. There were several reasons for this difficulty. The question on complications is different on New York City certificates from upstate certificates, and both versions of the question have been reworded from time to time. Also, many conditions are recorded which cannot be regarded as true complications; and different terminologies are used by physicians to indicate essentially the same complications. The clerical help was therefore instructed to regard as a complication any condition mentioned and to copy it exactly as recorded, unless it fell into a small list of specified items which could safely be ignored. Editing of the study forms will thus increase the proportion of uncomplicated births in the final control sample to somewhat greater than the planned one in three.

As was noted earlier, more reliance could have been placed on the prenatal and delivery data required for study if such data had been obtained directly from hospital records instead of from birth certificates; and, a matched set of controls might have been used as in the previously mentioned series of studies by Pasamanick and Lilienfeld. However, use of hospital records for the blindness study would have presented serious difficulties. These problems include the many hospitals which would have been involved in the record search; the unavailability or inaccessibility of appropriate hospital records covering the entire 13 years of study; the question of comparability of hospital records; and the excessive time and expense which would have been involved in the task of abstracting records from so many hospitals in all parts of the state. In contrast, the completeness of birth certificate reporting, the availability, accessibility, and centralized location of the necessary vital records, the ease in selecting a con-

trol sample, and the ease in abstracting the necessary data made it more feasible to use the birth certificate as the source of data for analysis.

Nevertheless, it is clear that birth certificates have at least two major limitations in a study of this kind. To begin with, only those items which appear in the birth certificate format can be studied, and these items vary over time (due to periodic revisions) as well as between states or other jurisdictions. Second, a question which is asked on the certificates will not necessarily be answered, or it may be answered inaccurately. The study by Oppenheimer and her associates¹¹ has shown that some items, such as birth weight, tend to be filled out accurately on nearly all birth certificates; whereas other items, including complications of pregnancy, are far less adequately reported. The effect of underreporting of factors associated with blindness would generally be to reduce the chance of finding the association, and hence to increase the credibility of any associations found; but the possible effect of misreporting is variable and, probably, generally harmful.^{12,13}

In view of the serious limitations in birth certificate data it is clear that only a relatively crude analysis can be justified. It is intended to compare the relative frequency of various factors of pregnancy, labor, and delivery in the case group with their relative frequency in the controls. This may be done separately for different categories of blindness. However, any significant associations found must be viewed as leads only, with further study needed to establish cause-and-effect relationships. Thus, the project may be viewed as a pilot study, looking for clues toward the explanation of the occurrence of blindness in children.

Study of Survivorship Among the Blind

In another study, survivorship patterns and causes of death among the

blind are being investigated. This particular subject appears to have been little explored. The only previous study known to the authors was conducted by the Metropolitan Life Insurance Company during the ten-year period, 1923-1933.^{14,15} It relates to the experience of 11,716 industrial policyholders who were granted disability allowances because of total blindness incurred subsequent to the issue of their policies.

As a group, it was found that the death rate of these blind policyholders was two and one-half times that of all industrial policyholders. For blind children, a very marked excess in mortality was noted with an extremely high death rate from cancer; for all ages, the causes of death which showed the highest mortalities as compared with normal (i.e., all industrial policyholders) were syphilis, diabetes, diseases of the arteries, and chronic nephritis. The eye condition most frequently reported as the cause of blindness was cataract. Next in importance was glaucoma and third, atrophy of the optic nerve.

Since the Metropolitan Life Insurance Company study was conducted approximately 30 years ago with the findings limited to totally blind policyholders, and since there is a continuing interest in this problem, further investigation was indicated.

The present study utilizes the statewide register of the blind maintained by the Massachusetts Division of the Blind in Boston. This register has been required by law since 1935, and reporting of the blind to the division became legally mandatory in 1943. The legal definition of blindness in effect in Massachusetts since the mid-thirties is, with minor exception, the same as that used for the New York State study.

With this definition in mind, records for each new blindness registration during the 20-year period from January 1, 1940, through December 31, 1959, are traced to the study closing date of December 31, 1961, to determine survival status as well as blindness status and to

obtain selected demographic and medical information. The theoretical end-points to follow-up are: (1) still blind on the study closing date; (2) sight restored before the study closing date; and (3) died before the study closing date. A fourth group is virtually inevitable in practice, viz., those lost to follow-up before the study closing date.

For each person not known to be alive a search for the death certificate is instituted at the Massachusetts Office of Vital Statistics in Boston and, where found, cause of death information is abstracted from the certificate.

The specific study objectives are:

1. to determine survival rates for blind persons by age, sex, and cause of blindness and compare with corresponding rates observed in the general population;

2. to determine sight recovery rates for blind persons by age, sex, and cause of blindness; and

3. to determine the distribution of causes of death among blind persons and compare with corresponding distributions observed in the general population.

The anticipated statistical analysis will be modeled after Chiang's¹⁶ work on the follow-up study which takes into account competing risks. Deaths will be subdivided by cause, into several mutually exclusive categories, each of which is considered a competing risk. Similarly, the "restored sight" group will be treated as a competing risk.

In the analysis, the total study population of about 12,000 white persons will be divided into broad groups according to age at registration, sex, and cause of blindness. The same analysis will be carried out for each of these groups. The results can be presented for each year of follow-up and can include various probabilities of interest (see Appendix).

To compare survival experience and causes of death among the blind with the general population of Massachusetts, use will be made of the x-year survival rate and the partial crude probabilities

of death by cause. Comparable values for the white population of Massachusetts can be derived from the 1949-1951 Life Tables and from proportionate mortality data available in the annual mortality volumes of the National Vital Statistics Division.

The nature of the study will permit various additional evaluations of interest to be made relating to the cause information on the death certificate. For example: Of those decedents known to have been blinded by diabetic retinopathy, how often was diabetes actually mentioned on the death certificate, at least as a significant condition present at death? Also of passing interest, how often was blindness mentioned as a condition present at time of death? From special checks being made on various demographic items appearing on both the death certificate and blindness agency records, a fuller knowledge of reporting errors should be realized and the statistical treatment of the data collected guided accordingly. The key role played by the death certificate in this study is obvious, and here it is really indispensable. The certificate enables one to: (a) certify as to death; (b) obtain an accurate date of death; (c) obtain the causes of death and the coded underlying cause of death;* and (d) obtain demographic information which may not always be available in the blindness records.

Of course there are also serious inherent limitations of the death certificate. Foremost among these is the question of accuracy and comparability of the causes of death as recorded over the entire study period (from 1940 through 1961). During this period, one must cope with causes of death classified according to three revisions of the International Statistical Classification of Causes of Death—the 5th, 6th, and 7th

* Coding was done by the Massachusetts Health Department and the codes have already been written in on the certificate.

Revisions. There were also several changes in format of the Massachusetts certificate itself.

Perhaps one further aspect of this study deserves some mention. Earlier, it was indicated that the index information for each case is first obtained from the blindness agency's register. In some instances, the fact of death may be known but the exact date or place of death may not be known. Finding the correct death certificate may become a very difficult task. Other information, such as age, sex, race, place of birth, marital status, as well as alternate or Americanized spellings of the name, has been found to help in the matching

process. Nevertheless, a number of instances will remain in which the death certificate is either not found or not definitely matched with the decedent's known characteristics.

Summary

The research potential of vital records has been expanded in recent years in the form of record-anchored and record-linked studies. Application of the record-linked approach in two studies on the etiology of blindness and survivorship among the blind has served to illustrate some advantages and disadvantages in the use of vital records for public health research.

APPENDIX

Some probabilities of interest in Study of Survivorship Among the Blind

Probability	Definition
Probability of retention	Probability that a person who is on the register at time x will be retained on the register until time $x+1$.
Probability of removal	Probability that a person who is on the register at time x will be removed from the register during the interval $(x, x+1)$.
Crude probability of death	Probability that a person who is on the register at time x will die in the interval $(x, x+1)$, in the presence of all other competing causes for removal.
Crude probability of death from Cause 1	Probability that a person who is on the register at time x will die from Cause 1 in the interval $(x, x+1)$, in the presence of all other competing causes for removal.
Crude probability of death from Cause k	Probability that a person who is on the register at time x will die from Cause k in the interval $(x, x+1)$, in the presence of all other competing causes for removal.
Crude probability of restored sight	Probability that a person who is on the register at time x will recover vision in the interval $(x, x+1)$, in the presence of all other competing causes for removal.
Net probability of death	Probability that a person who is on the register at time x will die in the interval $(x, x+1)$, if all other competing causes for removal are eliminated.
Partial crude probability of death from Cause 1	Probability that a person who is on the register at time x will die from Cause 1 in the interval $(x, x+1)$, in the presence of other competing causes of death but with all other competing causes for removal eliminated.

Probability	Definition
Partial crude probability of death from Cause k	Probability that a person who is on the register at time x will die from Cause k in the interval $(x, x+1)$, in the presence of other competing causes of death but with all other competing causes for removal eliminated.
Net probability of survival	Probability that a person who is on the register at time x will survive until time $x+1$, if death is the only acting cause for removal.
x-year survival rate	Probability that a person who is on the register at time 0 will survive until time x if death is the only acting cause for removal.
Net probability of restored sight	Probability that a person who is on the register at time x will recover vision in the interval $(x, x+1)$, if all other competing causes for removal are eliminated.

REFERENCES

- Haenszel, W.; Moriyama, I. M.; and Sirken, M. G. A Proposed Study for Extending the Scope and Improving the Quality of Mortality Data: Improving the Quality of Statistical Surveys. Washington, D. C.: American Statistical Association, 1956.
- Hauser, P. M., and Kitagawa, E. M. 1960 National Mortality Study (in progress). Population Research and Training Center, University of Chicago (Ill.).
- Lilienfeld, A. M., and Parkhurst, E. A Study of the Association of Factors of Pregnancy and Parturition with the Development of Cerebral Palsy: A Preliminary Report. *Am. J. Hyg.* 53:262-282 (May), 1951.
- Lilienfeld, A. M., and Pasamanick, B. Association of Maternal and Fetal Factors with the Development of Epilepsy: 1. Abnormalities in the Prenatal and Paranal Periods. *J.A.M.A.* 155:719-724 (June 19), 1954.
- Pasamanick, B., and Lilienfeld, A. M. Association of Maternal and Fetal Factors with Development of Mental Deficiency: 1. Abnormalities in the Prenatal and Paranal Periods. *Ibid.* 159:155-160 (Sept. 17), 1955.
- Pasamanick, B.; Constantinou, F. K.; and Lilienfeld, A. M. Pregnancy Experience and the Development of Childhood Speech Disorders: An Epidemiologic Study of the Association with Maternal and Fetal Factors. *A.M.A. Am. J. Dis. Child.* 91:113-118 (Feb.), 1956.
- Pasamanick, B., and Kawi, A. A Study of the Association of Prenatal and Paranal Factors with the Development of Tics in Children: A Preliminary Investigation. *J. Pediat.* 48:596-601 (May), 1956.
- Kawi, A. A., and Pasamanick, B. Association of Factors of Pregnancy with Reading Disorders in Childhood. *J.A.M.A.* 166:1420-1423 (Mar. 22), 1958.
- Frazier, T. M.; Davis, G. H.; Goldstein, H.; and Goldberg, I. D. Cigarette Smoking and Prematurity: A Prospective Study. *Am. J. Obst. & Gynec.* 81:988-996 (May), 1961.
- Kerby, C. E. Causes of Blindness in Children of School Age. *Sight-Saving Rev.* 28:10-21, 1958.
- Oppenheimer, E.; Schwartz, S.; Russell, A. L.; Taysback, M.; and Hunt, E. P. Evaluation of Obstetric and Related Data Recorded on Vital Records and Hospital Records: District of Columbia, 1952. Washington, D. C.: Department of Health, Education, and Welfare, Public Health Service, National Office of Vital Statistics, Vital Statistics—Special Rep. 45:359-416 (Nov. 20), 1957.
- Rogot, E. A Note on Measurement Errors and Detecting Real Differences. *J. Am. Statist. A.* 56:314-319 (June), 1961.
- Diamond, E. L., and Lilienfeld, A. M. Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies. *A.J.P.H.* 52:1137-1144 (July), 1962.
- Metropolitan Life Insurance Company. Mortality of the Blind. *Statist. Bull. Metrop. Life Insur. Co.* 16:1-3 (Sept.), 1935.
- Dublin, L. I.; Lotka, A. J.; and Spiegelman, M. Length of Life. (Rev. Ed.) New York, N. Y.: Ronald Press, 1949, p. 209.
- Chiang, C. L. A Stochastic Study of the Life Table and its Applications: III The Follow-up Study with the Consideration of Competing Risks. *Biometrics* 17:57-78 (Mar.), 1961.

Mr. Goldberg is assistant chief, Dr. Goldstein is chief, and Mr. Rogot is statistician, Biometrics Branch, National Institute of Neurological Diseases and Blindness, National Institutes of Health, Bethesda, Md. Dr. Quade, formerly in the Biometrics Branch, is now in the Department of Biostatistics, University of North Carolina School of Public Health, Chapel Hill, N. C.

This paper was presented before the Statistics Section of the American Public Health Association at the Ninetieth Annual Meeting in Miami Beach, Fla., October 16, 1962.