SOME STATISTICAL PROBLEMS INVOLVED IN A ONE-DAY CENSUS OF HOSPITAL PATIENTS

J. D. Colman and Nathan Morrison, Associated Hospital Service of New York

This is a report on work in progress, and presents some of the statistical problems that have arisen and the methods that are being used in studying them.

A brief description of the New York hospital census will be helpful as background. On May 10, 1961, all the short-term general care hospitals in the lower 17 counties of New York State, including New York City, Nassau, Suffolk, and Westchester, participated in the preparation of single-page case forms for each patient in the hospital at 11:59 p.m., and double-page forms for each patient who was discharged alive or who died on that day.

The single-page form (which was also the first page of the double-page form) contains items on the patient's demographic characteristics (age, sex, color, marital status, place of residence), admission diagnosis, date of admission, surgical procedures, if any, type of hospital accommodation, and method of payment. The second page of the double-page form contains items on the tests and treatments received by the patient, drugs administered, mursing care, and some summary figures on the hospital charges.

Each of the 230 hospitals in the 17 county area submitted forms for their cases as they were completed over a period of time (which began on May 11, 1961) without having to meet any deadline. We were more interested in accuracy and completeness than in speed. These forms have been carefully checked and edited and, when necessary, returned to the hospitals for correction or for additional information.

All types of hospitals, voluntary, proprietary and governmental, (including veterans' administration and public health service hospitals) in the area cooperated in this census. Hospitals which had both long-term and shortterm divisions were asked to report only on patients in the short-term divisions.

A total of 51,000 cases are included in the study, 46,000 involving patients who were in the hospital at 11:59 p.m. on May 10th and 5,000 persons who were discharged or who died on that day. As of December 28, 1961, about 50,000 of the case forms have been received from the 230 hospitals, and edited, coded, and keypunched. Complete data have been received from 220 hospitals. About a thousand more cases are expected to come in during the next three weeks. Most of these are cases sent back for correction. The cutoff has been fixed as January 19, 1962.

On the basis of pilot analysis of the records of a few selected hospitals, machine tabulation has already been started. The study has been a community enterprise from the beginning and could only have been carried out with the whole-hearted cooperation of all of the hospitals. The forms, procedures, tabulation plans, and statistical codes were developed in consultation with representatives of many institutions, both governmental and volumtary, experienced in working with such data. The study attempted to utilize and profit from the experience of the Professional Activities Survey, the Cleveland area study of 1956-57, the Michigan study (undertaken in 1958, and recently completed) and all past studies in this country and in Canada of a similar nature.

For all of the 51,000 hospital patients in the study, the single or first page form yields information on about 16 basic variables. For the 5,000 persons who were discharged or who died on May 10, 1961, the second page of the form yields additional information, some of which will need to be analyzed carefully on a qualitative rather than a quantitative basis. The data on the drugs administered offer many challenges, for example. However, it will be possible here to use the experience gained in two pilot studies in Michigan and in New Jersey.

In studying the data for the 51,000 cases, the approach will be to analyze the basic variables separately and in various combinations and to search systematically for meaningful interpretations.

Some of the basic variables are dichotomous (e.g., sex: male or female, and race: white or non-white.) Other variables have many more than two possible values (e.g., admission diagnosis: several hundred possible codes, and age: from the newborn to persons aged 100 years.)

It is planned to use a variety of statistical techniques in analysing the data ranging from frequency distributions, centering constants, to any appropriate multivariate analysis technique such as point biserial correlation coefficients for determining the degree of relation between a dichotomous variable and a contimuous variable, multiple and partial correlation, factor analysis, discriminant functions, and canonical correlation.

For the 5,000 persons who were discharged or who died on May 10, 1961, the date of admission - (the actual hour of admission and discharge is shown on the form) - is available and thus the length of stay distribution will be tabulated for these complete stays.

For the 46,000 persons who were still in the hospitals at 11:59 p.m. on May 10, the length of stay distribution will represent, of course, incomplete stays of various kinds.

A review of the published work and of some unpublished materials on hospital stay distributions reveals a number of promising approaches that can be taken in analyzing the two types of distributions arising out of the May 10th census.

The useful and interesting analogy between actuarial life tables and hospital stay tables has been mentioned in several forums and is explored in a 1953 article in the American Journal of Public Health by Robins and Sachs in which admissions, discharges, and patients remaining in hospitals are treated similarly to the births, deaths, of a living population in a life table.

Reference should also be made, in this connection, to a 1957 article in the American Journal of Public Health by Mortimer Spiegelman entitled "The Versatility of the Life Table."

An interesting approach is presented in a paper (not yet published) by Clifford A. Bachrach entitled "Estimation of Length of Hospital Stay from Discharge Data." One sen-tence may be quoted from this paper: "Given a distribution of length of stay of discharged patients, the stationary population model permits us not only to predict the mean stay-todate, but also permits us to estimate all other characteristics of the distribution of length of stay-to-date of patients remaining in the hospital." It will be possible, of course, to test this approach by using the length of stay distribution of the 5,000 discharged cases to predict the length of stay distribution of the 46,000 persons remaining in the hospital on May 10th, and comparing the prediction with the actual experience.

It will also be possible to apply the approach suggested in two 1959 papers by Balintfy and Flagle of Johns Hopkins. Balintfy's paper is entitled "A Stochastic Model for the Analysis and Prediction of Admissions and Discharges in Hospitals." His abstract, at the beginning of the paper, states:

> "Frequency distributions indicate that the length of stay of any particular patient in a hospital may be regarded as the joint effect of a large number of causes acting in ordered sequence during the time of recovery and defining the probability of discharge as a stochastic function of the length of treatment. Making use of this relation, statistics of expected figures of daily discharges and admittances have been developed as conditional values at given occupancy. The effect of chance and seasonal fluctuation, and the possibilities of smoothing is

demonstrated by the calculation of the method of forecasting expected occupancy levels.

"The paper reveals the basic statistical laws governing the admission, discharge, and length of stay distributions in hospitals. It proves that both admissions and discharges are compound Poisson processes following the negative binomial distribution and that length of stay follows the logorithmic normal distribution. This information is used to predict the daily census and indicate the direction of further research."

At the conclusion of his paper, Balintfy suggests the possibility of applying Markov processes in describing the variation in a hospital census over time. He also notes that the results concerning the forms of the distribution of arrivals and length of stay suggest the consideration of multichannel queuing models for analysis of hospital systems.

Some of these methods and models described by Balintfy and Flagle can be applied to the analysis of the May 10th hospital census data. The 51,000 cases form a large enough group to permit study of various sub-groups.

Another approach for examining the flow of patients through the hospital has been suggested by R. W. Revans of the University of Manchester in a 1959 article. Revans would study the major stages of a patient's experience in the hospital. It would be possible to test variations of this approach on a sample of the cases in the May 10th Census.

In addition to the explorations outlined above, the specific purpose of the study is to describe with reasonable precision the major elements of hospital service used by the population in the Greater New York area in 1961.

REFERENCES

- Clifford A. Bachrach "Estimation of Length of Hospital Stay from Discharge Data" - to be published (Journal of American Statistical Association.)
- Joseph L. Balintfy "A Stochastic Model for the Analysis and Prediction of Admissions and Discharges in Hospitals" -Proceedings of the Sixth Annual International Meeting, The Institute of Management Sciences, 1959.
- 3. Anthony F. Bartholomay "On the Linear Birth

and Death Processes of Biology as Markov Chains" - Bulletin of Mathematical Biophysics, Vol. 20, 1958.

- 4. Charles D. Flagle "The Problem of Organization for Hospital Impatient Care" - Proceedings of the Sixth Annual International Meeting, The Institute of Management Sciences, 1959.
- 5. P. H. Leslie "On the Use of Matrices in Certain Population Mathematics" - Biometrika, Vol. 33, Part III, November, 1945.
- A. G. McKendrick "Applications of Mathematics to Medical Problems" -Proceedings of the Edinburgh Mathematical Society, Vol. 44, 1925-26.

- 7. R. W. Revans "The Hospital as an Organism: A Study in Communications and Morale" - Proceedings of the Sixth Annual International Meeting, The Institute of Management Sciences, 1959.
- Morton Robins and Rose Sachs

 "An Application of Biometry to Hospital Statistics: Hospital Stay Tables" - American Journal of Public Health, February, 1953.
- 9. Mortimer Spiegelman "The Versatility of the Life Table" - American Journal of Public Health, March, 1957.