GUIDELINES FOR EVALUATION OF FOLLOW-UP ARTICLES AND PREPARATION OF MORTALITY ABSTRACTS

RICHARD B. SINGER, MD
Consultant to ALIMDA

MICHAEL W. KITA, MD
VP and Associate Medical Director
UNUM Life Insurance Company
Portland, Maine

Introduction
Clinical medicine and insurance medicine both grow by accretion—by the gradual augmentation and improvement of what is known and measurable. There is an ongoing need to extend the results of previous follow-up (FU) studies, to update them, to collect data on impairments never before studied, and to apply all of these results in more meaningful ways. Articles with useful FU data not previously used in published mortality and morbidity abstracts provide a ready stockpile of sources of potential value to insurance medicine. Criteria by which “suitability” of FU studies might be judged were the subject of a recent article in the Journal of Insurance Medicine.1 In Part I of the present article we will develop a more extended discussion of the many factors that contribute to the suitability of non-insurance studies, or their unsuitability. Part II will provide an in-depth critique of methods that have been found to be useful in the transformation of observed FU data into tables of comparative mortality or morbidity and text for an abstract. We will not discuss methods and strategies for searching medical literature through the numerous on-line services that make use of the computerized index of the National Library of Medicine. However retrieved, once a FU article becomes available to a medical director, Part I may be helpful in the evaluation of the applicability of the data to the concepts of comparative mortality. Part II is intended to assist any volunteer interested in the creative process of preparing a formal abstract for submission to this Journal. However, the same methods may also be applied by the medical director who wishes to make a rapid and informal mortality estimate in a particular underwriting or claim case. We urge all medical directors to keep in mind this potential for personal application of life table methodology in their daily work, even if they do not aspire to preparing a formal mortality or morbidity abstract.

Part I-Evaluation of Articles

Condition Studied. The prevalence of the disease or condition under study in a FU article has an important bearing on the degree of classification needed and the usefulness of the results despite small size of the series. The reviewer should have some familiarity with the intercompany and other insurance studies, and with the mortality abstracts published, and should also have reference data on disease prevalence. Mortality and morbidity FU studies are made not only on chronic (and acute) diseases, but also on examination findings, such as blood pressure and weight in relation to height, test findings such as ECG or serum cholesterol, and nonspecific history such as chest pain or a single episode of febrile convulsion in childhood. To these we apply the term “conditions,” as opposed to diseases.

Classification. A mortality study of coronary heart disease (CHD) patients should be classified in much greater detail for subsequent use than a study of patients with amyloidosis. Those interested in comparative FU data in CHD patients will probably focus their attention on a particular aspect, such as myocardial infarction (MI), or bypass surgery, or the prognosis of acute MI patients with complicating ventricular arrhythmia. On the other hand, it is unlikely that anyone would wish to classify in great detail abstracts on a rare disease such as amyloidosis. The topic of disease classification will be dealt with in a JIM article now in preparation.2

Size of Series. A frequent question asked by reviewers of FU articles is, “what is the minimum size of a series (or minimum number of deaths), for the data to be worth the effort of preparing a mortality abstract?” Although there is no simple answer to this question, some criteria can be outlined. For example, a huge volume of cases, exposures and deaths is available in the 1979 and earlier Blood Pressure Studies. A clinical series of 2000 hypertension patients with over 100 deaths would add little to the results of these studies, unless the report included special aspects not found in the Blood Pressure Studies, such as malignant hypertension, now a relatively rare disease. For any rare disease, with no known previous study, a series with fewer than 50 patients and only 0-5 deaths might be of value for a mortality abstract. If minimum size standards can be developed in the future, it will have to be in relation to the prevalence of the disease, ranging from common to very rare, and to the existence of previous comparative mortality studies. With respect to comparative morbidity, such abstracts are extremely rare, although morbidity FU observations are being reported with increasing frequency in the medical literature.

Type of Follow-up Study. There are several different methods of designing and carrying out FU studies, some of which provide for “controls,” but more often than not a suitable “expected” mortality is not given in the article. Features of these types are given briefly below; details are more important to the analyst who prepares the tables than they are to the reviewer.
Prospective Study, planned in advance, to define a group with a common disease or condition, to determine all needed characteristics of each entrant, and to follow the entire group at stated intervals. Examples are the Framingham Study (entry at nearly the same starting date), and most Cancer Registries (entry over a period of years). The duration of entry period is important because many entrants will be withdrawn as survivors at the cutoff date, with differing durations of FU, because they have come to the end of follow-up. If all entrants have the same starting and ending dates and none are lost to FU, the only attrition is death, an important factor in rate calculations.

Historical Prospective Study, planned and carried out after the beginning of the entry period. Intercompany and individual company mortality studies are of this type, also studies utilizing the very complete record systems of the Mayo Clinic and the related Rochester (Minnesota) Epidemiological Project. Studies of this type outnumber purely prospective studies.

Randomized Clinical Trials are prospective studies in which patients from a large pool are assigned in randomized fashion to two or more groups, at least one of which serves as a placebo control, while a particular method of treatment is used for the other group(s). Randomization is often done on a double-blind basis, and such studies are often sponsored by one of the National Institutes of Health in the U.S., or public health agencies in other countries. Epidemiologists tend to consider these the "gold standard" for the efficacy of a drug or other treatment method, perhaps a term made more appropriate by their great expense. However, clinical trials share many of the problems of other FU studies, and some students of mortality studies believe that a good observational study, of the historical prospective type, may provide information as useful and as statistically conclusive as that derived from a clinical trial.

Case Control Studies involve identification of patients having a certain characteristic to be studied (such as adverse effect of a drug) from a large set of records, then identification from the same record pool of another series of patients lacking the study characteristic, but matched as closely as possible with respect to age, sex and other factors that are felt to be important to the outcome, either death or some morbid event.

Retrospective Studies involve investigation of the preceding course of a disease in a series of patients identified at some endpoint of the course, such as death and autopsy, or hospitalization. Rare exceptions are possible, but generally autopsy series cannot be utilized to prepare data for a mortality abstract. Almost all articles of this type would be rejected as unsuitable for abstract preparation.

Formation of the Series is a matter of direct concern for the representativeness of the experience. All kinds of bias may enter into the way in which the subjects or patients are selected for the study. A study of life insurance policyholders is biased to lesser degrees of severity in persons from a higher than average socioeconomic group. A study of a series of hospitalized patients is usually biased to higher degrees of severity of the disease studied. A proper sample of the population is needed to obtain all degrees of severity; this is the advantage of a registry such as the Rochester Epidemiological Project, which provides unusually complete case detection and follow-up of the residents of Olmsted County, Minnesota.

Demographic Data at Entry are important, because accurate expected mortality is dependent on accurate matching of the study group(s) followed by age and sex. Race, socioeconomic status and selection characteristics are also important. Race data are seldom provided, but usually mean age and percentage of females are given as a minimum. The problems generated in deriving expected mortality when more complete age/sex distribution is not provided are difficult for the abstract preparer to handle, but will not be detailed here.

Follow-up Notes. Salient characteristics are important in describing the FU achieved in the study, and usually these must be deduced from the data given in the Material and Methods section of the article. These include the entry period, the cutoff date, minimum, maximum and mean durations of FU, adequacy of methods used in FU, and percentage of patients lost to FU. If the authors rely on questionnaires for FU without supplementing data on non-respondents by other means, it will generally be evident that a large fraction of the entrants has been lost, and follow-up is so incomplete as to be unsatisfactory. One reason for non-response, of course, is death of the entrant, and the proportion of deaths in the lost patients or subjects may be higher than in those successfully followed. This part of the study should be evaluated with a critical eye. These and other aspects of FU and death ascertainment are very important for the analyst who prepares the abstract.

Classification of Patients or Subjects. In most FU articles the experience is given not only for the total series but also for subdivisions thereof. The total may be broken down by sex, by age, by diagnostic group, by features of the history, examination or test results at entry, or by a severity classification, such as the New York Heart Association functional class, or cancer stage, or some other method, including a scoring system. Division may be one factor at a time (univariate), combination of factors (cross-classification), or one of the statistical methods of multivariate analysis. Such classification information is important in describing results of the FU article.

Derivation and Presentation of Results appear in almost infinite variety in FU articles. Graphs of cumulative survival curves (P) are the most common method of presentation, sometimes with additional data, such as the numbers of survivors at various durations after entry. Actuarial life table methods are used to derive the curves, but observed life table data are given in only a minority of articles, and then in a variety of tabular formats, some omitting essential data. Expected mortality and survival are given in less than 20% of articles, although control rates of cumulative mortality or morbidity may be used for comparison with observed rates in clinical trials and case control studies. Terms such as risk ratio and relative risk are generally used instead of the more familiar term, mortality ratio. These are problems the abstract preparer must cope with, but the responsibility of the reviewer is to note with care the presentation of results. If expected rates are given, the source table should also be noted.
**Significance Tests.** Such tests are available in great profusion, and the methods used in any article are generally cited by name and reference. This information is of importance to the analyst who prepares the abstract.

**Article References.** These are an extremely important part of the article review, because they often point to other articles with good FU data, which may assist the task of abstract preparation. The authors often have done a good job of reviewing the literature, and references to this are apt to appear in the introduction or the discussion section of the article. The text and references of the article should therefore be read carefully with this point in mind. References that appear to be of potential value should be listed, with the page number of the text in which the citation appears, for future use.

**Potential Value of a FU Article.** With many factors to be weighed it is extremely difficult to devise a satisfactory method of grading the value of an article as a source for a mortality abstract. The method proposed here is a provisional one, little tested and subject to change with accumulation of experience in its use. A formal evaluation sheet of this sort is designed primarily for a cooperative review effort. If such an effort should be developed in the future it would probably be organized and supervised by the Mortality and Morbidity Committee and would involve a descriptive sheet for the article, of which this proposed evaluation system could be the final part. The guidelines in Part I can be used either independently or as part of an organized article review process. Five evaluation grades are defined:

- **A** - highest value (top priority)
- **B** - intermediate value
- **C** - relatively low value by itself
- **Q** - questionable value, article to be referred to other reviewer
- **U** - unsuitable for abstract

At present it is impractical to define a numerical grading score. The two chief factors to evaluate are the number of existing sources of comparative mortality or morbidity, and the size of the experience (number of observed deaths) reported. Total deaths are sometimes not reported if the data are in the form of survival curves; in this case estimate mean duration from maximum and minimum FU in years, estimate exposure as product of number of entrants and mean FU in years, and then expected deaths as product of E and q', using q' of 0.002 for a mean age over 60. The reviewer will have to rely on his own knowledge of the literature of mortality abstracts and insurance mortality studies to decide whether existing studies are abundant, limited, or virtually non-existent, but a bibliographic summary is in preparation to aid in classification of this factor. Unsuitable articles include those with case reports only, most autopsy series and retrospective studies, those with an approximate exposure yielding less than 1 expected death, and those with a serious defect in design, methods used, or the reporting of the FU data. If the reviewer cannot classify properly or is doubtful about the methodology, the article should be given a Q grading and referred to the most experienced reviewer for evaluation. If methodology appears to be satisfactory, most articles not considered unsuitable can be given a grading of A, B, or C in accordance with the following table:

**Size of Mortality Experience Reported**

<table>
<thead>
<tr>
<th>Existing Mortality Sources</th>
<th>Deaths &gt; 100</th>
<th>Deaths 26 - 100 &amp; exp. deaths &gt; 1*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost nil</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Limited</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Abundant</td>
<td>B (&gt;1000)</td>
<td>Q</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
* See text above for estimating d' in absence of report of total deaths† If there are several new articles in this class for a given condition, their value will be enhanced by combining them in a "composite" abstract

**Part II—Abstract Preparation**

The preparer of an abstract may utilize an article that he has personally reviewed or one from a stockpile, with a descriptive sheet. In either case, the preparation of an abstract is a creative process, one that requires a thorough knowledge of the standard format used, and an equally thorough knowledge of life table methodology (see Dr. Pokorski's full seminar text). The following guidelines are intended to be of further assistance in the preparation.

**Duration Questions:** Are results given by duration, or only for all durations combined? Is the maximum duration of follow-up short (e.g., 1-2 months), or long-term (about 4 years or more), or in between? This assumes a prospective or historical prospective study: a case-control, period, or retrospective study does not ordinarily involve duration as a factor. In some conditions, such as acute myocardial infarction (MI) or stroke, the early or short-term mortality may be extremely high, in which case the rate usually decreases quite rapidly over a period of days, weeks or months until it becomes more stable, a duration pattern that can be properly described on a yearly basis, or even by longer intervals. When early mortality is much higher than it is subsequently, the early experience should be excluded from long-term experience and presented in a separate table. It may not be necessary to put the early mortality in comparative terms. For example, in acute MI of hospitalized patients only three variables need be shown for each age, sex, severity or other category: the number of MI patients hospitalized, the number of deaths to the end of the early duration, such as 30 days after admission, and the early mortality rate, often given as a percentage, unlike the decimal units or deaths per 1000 exposed to risk, used for long-term intervals. When this separation is made, long-term mortality and survival have, as a starting-point, the end of the early observation period. If annual duration intervals are used, the first interval is truncated and is labeled in the duration column as "30d-1yr," or "1-12mo," or "HD-1yr," where HD stands for hospital discharge, when this is designated as the end of the early period. The first long-term interval is therefore shorter than the subsequent intervals of one full year, and one must be careful to make sure that exposure in **person-years** and **annual mortality** and survival rates are adjusted for the duration of less than 1 full year in the first interval. However, if
annual rates are shown, the first-year cumulative survival rate, $P_1$, must be derived from the internal survival rate, $p_1$, not the first-year annual survival rate, $1 - q$. Thereafter the interval and the annual survival rates are identical. If graphs of cumulative survival rates are the primary source of data in the article, the curve and the caption of the graph should be examined carefully to determine whether the early deaths have been excluded or not. If early MI deaths have not been excluded, the $P$ curve should show a steep initial dip reflecting the high mortality in the first month, and then should have much less of a slope in the remaining portion of the first year. The time scale or the caption should make it clear that 0 time is hospital admission. However, if early deaths have been excluded, there will be no initial steep dip, and $P = 1.00$ will correspond to 0 time of 1 month, or whatever the end of the early period is. Such a starting point for the rest of the time scale means that the first year is a full 12 months, not the truncated year that results from using hospital admission as the starting point for long-term as well as the early follow-up. Alternatively, the graph may show $P = 1.00$ at $t = 1$ month or $\frac{1}{2}$ year, in which case duration and the time scale are not based on hospital admission as the starting-point, and the first year of long-term follow-up is a truncated period. These details are important because, if they are not recognized, errors may result in calculation of the first-year mortality rate and in all values of $P$.

Another important aspect of duration is the statement in the article regarding the calendar years of entry, in relation to the calendar year(s) of cutting off follow-up. In most published studies patients have been entered in the study over a period of years, such as 1970-79, and followed to a year close to the last year of entry, such as a cutoff at the end of 1980. In this case the minimum duration of follow-up is 1+ year, and the maximum is 10+ years. A follow-up range of 1-10 years implies that patients must have been withdrawn alive due to end of follow-up in each year of observation after the first. Such a follow-up may be called a “double-decrement” study because patients are withdrawn either because of death (one decrement), or because of end of follow-up or becoming untraced while still alive at last observation (the second decrement). In contrast is the less common situation in which the study is designed so that all patients in effect have the same entry point and the same end point of follow-up. This situation constitutes a “single-decrement” study if there are no patients untraced. In each year of observation $w = 0$ and $E = I$. It is no longer mandatory to calculate $Q$ as $1 - P$ from annual values of $q$ and $p$. If there are 1000 entrants and 100 deaths over a uniform follow-up of 5 years, and no annual data are given, $Q$ can be calculated by the “ad hoc” method as $q = 100/1000 = 0.1$, and

$$P = 1 - Q = 0.900.$$  

Geometric mean annual rates of survival and mortality may then be derived. It can be shown mathematically that this value of $P$, the cumulative survival rate, will be derived regardless of the distribution of the 100 deaths over each of the five observation years. This uniform follow-up duration for all entrants presents a very different situation from the usual one of a wide range of follow-up durations and many patients withdrawn because of end of follow-up, even when there are very few or no patients lost to follow-up.

When the authors do not present annual life table data, including $w$, $E$, and $d$, it may still be possible to approximate the total $E$ and the mean annual mortality rate, if the mean follow-up duration, $\Delta_t$, is given with the total number of deaths. Suppose there are 200 entrants with a $\Delta_t$ of 5.0 years; then $E = (200)(5.0) = 1000$ person-years. If 50 deaths were observed, then the mean annual mortality rate, $q = 50/1000 = 0.050$, or 50 deaths per 1000 per year. If a valid matching of $q$ can be derived, so can $d^t$, the MR and EDR. This approximation is accurate, because the only way to calculate mean $\Delta_t$ is by the quotient, $E/I$: the authors must have calculated total $E$ even if they do not report it as such in the article.

It should be evident that analysis of these duration aspects is most important in the development of life table calculations for comparative mortality and in the design of the abstract tables. Without separation of the high early mortality experience, if it is present, first-year and average 5-year observed mortality rates will not be representative of the greater part of the long-term experience. There is no high excess acute mortality, of course, in the follow-up of chronic conditions which lack an initial period of high risk, or in which the entrants have survived this period (post-MI survivors, for example). On the other hand, a common type of follow-up reported is for patients treated with major surgery, and the early perioperative experience should be separated from the long-term experience. All too frequently this is not done in the cumulative survival curves reported. $P$ values based on $P = 1$ at operation can be adjusted easily by dividing by $p_0 = 1 - q_0$, where $q_0$ is the early, perioperative mortality rate.

Age/Sex and Other Patient Groupings: The Methods and Results sections, the tables and figures, and the other text of the article should be carefully examined to ascertain the scope of the age/sex groups, and the diagnostic, severity and other patient groupings. The calculation of expected deaths, mortality and survival rates will depend on the extent of detail in the age/sex entry data, and whether results by duration are given by age and sex. With several age/sex groups at entry it is possible to derive accurate values of $d^t$ and $q^t$ for the first year (calculate $d^t$ in each age/sex group as the product of first-year $E$ and $q^t$ taken from the table of expected rates, with the central age of the age group, then add the $d^t$ values, and divide this by the total $E$ to get the first-year $q^t$, all ages and both sexes combined). But if there is no follow-up by separate age/sex groups, the progression of $q^t$ by duration becomes a difficult problem (see section on expected mortality below). Bear in mind that mean age and age distribution will probably differ from one patient group to another, and $q^t$ will vary accordingly. If age/sex distribution data are requested from the authors, they should be requested for all patient groups for which mortality or survival results are given. The diagnostic and severity groups will determine many features of table format and size. Subdivision of the total series should not be attempted if numbers of deaths are too small in the individual data cells.

Need for Additional Data. It is important to decide on the need for additional data as early as possible in the evaluation of the article. Is an initial age/sex distribution given, and if not, how essential is it to have this information from the authors?
patient series with myocardial infarction, coronary heart disease with angiogram, coronary bypass surgery, aortic and mitral valvular heart disease, and some types of cancer, it is possible to use a mean age and obtain a good approximation of first-year mean $q'$ (see section on expected mortality). A less accurate approximation from the mean age is possible in other diagnoses if the age distribution is similar to the age distribution of one of the above. Age distributions reported in the Medical Risks volumes or in other reference sources should be consulted. Furthermore, it is of the utmost importance to obtain any earlier papers of the authors which may describe the age/sex distribution or other essential data pertaining to the series being reported in the current article. If the study is of special value, with large exposure and numbers of deaths, it is usually desirable to write to the authors requesting the age/sex distribution, if available, not only for the total series but for the diagnostic and severity groups also. A similar need may arise for exposure data, or for life table data supporting cumulative survival curves in graphs, or tabular $P$ values given only to two decimal places. Occasionally there is no reply, but often the requested data are made available. It should always be made clear that the data are desired for a mortality abstract that will be published, but with full credit to the original paper(s), and to the author(s) for the additional data received. Many mortality abstracts made in the past have been greatly enhanced by such additional data.

**Expected Mortality.** The choice of the "best" (most appropriate) mortality table to be used as a comparative standard for expected mortality is a most important one. Contemporaneous select and ultimate tables are the best for most studies of individual life insurance applicants and policyholders. Population life tables are appropriate for studies of patients or subjects that can reasonably be considered to be a random population sample. This assumption may be true for some clinically generated series, but a careful reading of the description of how the series was formed will frequently reveal that a considerable degree of selection has been used: age has been restricted, patients with certain severity or other characteristics have been excluded, patients with other high-risk conditions have been excluded, etc. If selection has been used it may be better to utilize a Group Life Insurance table or similar table, giving a lower expected mortality than would be derived from a population table. Race information is important, but seldom available in the article; it will be necessary to decide first on use of a total population or white population table from U.S. Tables. Because of the large sex difference in mortality at most ages it is better to separate the male and female age groups, even when the sex distribution is given only for the total series, not for the various age groups. Data are available that show the age variation and the male/female percentages in CHD and some other conditions (see the Medical Risks volumes). If early deaths are excluded, the age distribution should be that for the early survivors, not the initial entrants (early mortality is usually higher at the older ages). All of this involves careful review of the article, good judgment in selecting the expected tables for use, and in the derivation of the first-year rates matched by age/sex group to obtain an overall mean expected mortality, $q'$. Except for the age/sex distribution these derivation calculations seldom appear in the abstract. The work-sheets should always be made available to the reviewer of the abstract, and both text and all tables should define the expected tables used.

It is well known that, regardless of the table of expected mortality used, the rate increases geometrically by a nearly constant factor of $1.1$ per year between the ages of 40 and 80 years. From age 2 to age 36 years annual mortality rates in the U.S. white population are very low, under 1 per 1000 in females, and under 2 per 1000 in males. Starting under age 80 in males and over age 80 in females the geometric factor commences a gradual decrease, and is less than 1.02 at age 110. In chronic diseases, such as coronary heart disease (CHD), the patient sample usually has a wide age range, even if older patients are excluded, as they often are in clinical trials. The mean age ($\bar{x}$) is often of the order of 50 to 65 years, with a 95% range about ±20 years around the mean. Given the geometric increase in $q'$ in the life table and such a mean age and age distribution about the mean, it can be easily shown that $q'$ values at ages older than $x$ will contribute more to the mean $q'$ than $q'$ values at the younger ages. In other words, the mean $q'$ in a male or female cohort will always exceed the tabular $q'$ corresponding to the mean age, $\bar{x}$ (i.e., mean $q'$ exceeds the $q'$ of the mean age). If only the mean age is given in an article, with a range or S.D., one can approximate the mean $q'$ by entering the male or female table with an adjusted age, e.g. ($\bar{x}$+3). For example, a group of men surviving a coronary bypass operation (CBPS) might have a typical age distribution that yielded a mean age, $\bar{x}$, of 51 years, and a mean $q'$ of 0.0106. The tabular $q'$ for white men age 51 is only 0.0078, but the $q'$ for age (51 + 3) or 54 years is 0.0103, a much better approximation of the mean $q'$. Such an approximation has been confirmed empirically in a large number of cohorts of patients or subjects when the mean age is of the order of 50 to 65 years and the range is of the order of ±20 years. If the range is only 10 years, as in a group of patients 40-50 years old, the mean age, such as 45 years, can be used to enter the table to obtain an accurate mean $q'$, unless the age distribution is extremely skewed. Both mean age and range of age are therefore of importance in any use of the empirical factor of +3 years to add to the mean age. When age/sex distribution is not given in the article, authors usually report the percentages or numbers of men and women. An overall mean $q'$ for this sex distribution may be approximated by obtaining the adjusted $q'$ values separately from the male and female expected tables and weighting them according to the respective male and female percentages. Data on the derivation of an accurate mean $q'$ from mean age and range of age are given in Table 642D, for post-CBPS patients. Unpublished data also exist (RBS files) for patients surviving an acute myocardial infarction (MI), patients with aortic or mitral valve disease, treated surgically, and patients with cancer of the colon or thyroid.

Unfortunately, an accurate first-year mean $q'$, all ages combined, does not solve another problem in the progression of mean $q'$ with duration, so that accurate values of $q'$ will be available to match annual or mean annual values of observed $q$. Despite the fact that each annual survivor in the cohort is, indeed, one year older than he or she was in the preceding year, the mean age of the cohort very often advances at a variable fraction of a year with each year of duration, and may even
decrease temporarily after the first year. This is due primarily to the effect of a higher observed mortality at the older ages, which results in a significant shift in age distribution to a younger mean age, despite the fact that all survivors are a year older in the next successive year of follow-up observation. This observed mortality effect is abetted by a wide initial age range and a high rate of withdrawal due to end of follow-up. It is difficult to make any valid recommendation as to decreasing the geometric increase factor of 1.1 by a given amount. Only about one half of the series of CBPS patients tested from follow-up of individual age groups showed a factor much below 1.1, averaged over at least 5 years (there may be considerable random variation from year to year). The annual factor in the CBPS cohorts varied from 1.05 to 1.08 when it was less than 1.09. In the series of patients surviving an acute MI, with higher mean ages and wide ranges of age, the increase factor ranged from 1.03 to 1.08. In cancer of the colon, the annual increase factor may range from 0.93 to 1.08 in localized and regional extension stages, with a mean of about 1.02 at 1-5 years, and 1.04± at 5-10 years. The decrease in mean q' at 1-2 years was even more striking in metastatic cancer of the colon, with a factor of 1.00 averaged over 1-5 years and about 1.05 over 5-10 years. In thyroid cancer, which is characterized by an unusually wide range of age, the early decrease in mean age and mean q' is very prominent and persistent, giving a mean factor of only 0.68 in duration years 1-5! There is a need for similar testing of other patient cohorts, not only in cancer, but in a wide range of other diseases, to establish some sort of pattern of progression of q' in annual survivors of such cohorts. It is important to remember that editors or reviewers of any formal mortality abstract will be desirous of seeing all calculations of expected mortality, and they should be provided with any worksheets (legible and labeled!) that will permit them to follow your derivations. For many abstracts, a new section, Expected Mortality, would be proper to add to the standard abstract format, as illustrated in Table A of the Mortality Abstract on the Aspirin Component of the Physicians' Health Study [JIM 22 (4): 281 (1990).

Tables Based on Life Table Data

1. Early Mortality Results. Generally the observed mortality rate is given as q_o = d_o/l_o, this subscript indicating the early mortality period of 4 weeks, hospital admission to discharge, or however it is defined. These three variables are apt to be the only ones displayed (expected deaths or rates and comparative indices are seldom used), and the acute mortality rate is often given as a percentage rather than a decimal or rate per 1000. The table design may vary widely according to the type of data displayed. Many examples can be found in the first Medical Risks volume, and additional ones in the tables of Abstract 609, or Tables 613A-B, 618A-B, and 620A in the new volume.

2. Extensive Life Table Data. The essential observed data are l, E, and d, from which are derived the mortality and survival rates; E itself is derived from l and w, but it has become the practice in mortality abstracts of the past 10 years to omit the column of w data to save space. If l values are available they probably should be included, even though it was not the custom to report them in the tables of the 1983 Impairment Study. Life table data in an article are considered extensive when they include results by age and duration, or a combination of age and duration. Prototypes of detailed life table results of this sort may be found in Tables 625A and 650C-D. Subheadings within the table may be used to distinguish different severity, diagnostic or other categories of the data. As shown in Table 625A, 9 columns are needed to present what are generally considered to be the minimum essential results for observed and expected data, and comparative mortality: duration, l, E, d, d', MR, q, q', and EDR. With omission of all survival data, survival ratio and cumulative mortality ratio, this compresses into one table what used to be shown in two companion tables in Medical Risks. In the new volume a strenuous effort was made to reduce the number of data columns and thus "simplify" the tables. One of us (RBS) had to fight to retain data considered essential in some of the important abstract drafts, and longer tables did not always prevail. Even with 9 columns the detailed tables still omit all survival results, and they do not contain confidence limits, which the reader has to take out of a reference table of Poisson limits. Formerly q and q' were given as 3-place decimals, but the tendency now is to give them as deaths per 1000 per year, to match EDR. When the life table data are given by a combination of age and annual duration it is generally necessary to save vertical space by combining individual years into longer intervals, such as 2-5, 5-10, etc. If the results are for all ages combined, separate annual data may be presented up to 5 or even 10 years. The pattern of mortality with duration determines at what follow-up year it is possible to combine results without sacrificing a trend that may be important for the reader. Age groupings are most commonly decennial, but there may be even fewer groups, such as <55 and 55 up. Detailed life tables of this sort are the easiest ones in which to develop q' and d', and hence the indices of comparative mortality. Unfortunately, such detailed results are not often available in the published article. If the authors are cooperative, the data they supply sometimes include quite detailed life table results.

3. Abridged Life Table Results. This term is applied to a variety of life table formats that are less detailed than those containing data by age and duration. The results are based on E, d, and d', with derived q, q', MR and EDR, but not infrequently one or more of these variables may be lacking. If both E and EDR are missing, the life table is incomplete as well as abridged. An example of a fairly complete life table is Table 324A, in which mortality ratios are given for widowed and married persons by sex and age in the same table. The two groups are matched for age and sex, so that l and E values are identical for them. Two sets of data are given for d, d' and MR. EDRs can be derived from the E, d, and d' data, but have been omitted because the table already contains 10 columns of data. Table 623A contains only d, d' and MR data for the CHD experience of the Prudential Insurance Company of London; with E data omitted it is impossible to calculate EDR. It is a disservice to the potential users of the tabular results to omit essential exposure data, so that EDR cannot be derived: when all ages are combined, EDR is a more reliable estimate of excess mortality than MR, because it is less sensitive to age differ-
ence between groups. In the 1976 Medical Risks volume, tables at the end of each major disease section present combined limited results from various sources. These do contain E, d, d', MR and EDR data, generally for all ages and durations combined from one source; such tables may be regarded as collections of abridged life table data.

Occasionally "oddball" presentations of data are encountered. An abridged life table containing P by duration may give the P value for the start of the interval, instead of the end, as is the usual custom. In a follow-up study of asymptomatic carotid bruit, Thompson et al.5 reported, not life table data, but follow-up data by annual duration consisting of three columns: "No. of Long-term Survivors," "No. of Long-term Deaths," and "Total No. of Patients." In one table the grand total at the bottom of the third column was 132 patients, with 89 total survivors and 43 total deaths in 16 years of follow-up. Since the authors state that no patient was lost to follow-up, the distribution of survivors in the first column represents annual values of w, patients withdrawn alive due to end of follow-up. The annual distribution of deaths in the second column is clear enough, but it may not be easy to recognize the significance of the data in the first column as actual values of w. They are not so labeled—the mortality abstract preparer must figure this out for himself, and how to take the data and construct the actual observed life table data, with columns of l, w, E, d, q, and, if desired, p, P, and Q. Both the reviewer of articles and the preparer of tables must be prepared to cope with such special situations.

4. Ad Hoc Life Table Data. When follow-up is complete to the end of the period of observation, w = 0 for each year of observation, l = E, and Q may be calculated directly as 1 - l. Or, if I is given annually, mean annual 4 = 10d/S. An average value of q over several years may be annualized as an aggregate or a geometric mean, depending on the available data. Examples of these may be found in two different post-MI series, Table 613C and 614E.

5. Life Table Data by Cause of Death. In all tables of this sort there is a common value of E for each cause of death and for the total. An example is Table 624P, the experience by cause of death in the Japanese Declined Lives Study of applicants with CHD. E is given in the total row at the bottom of a separate E column, but it could have been given in the Table heading. In addition to the cause of death column, other data in this table were d, d', MR, aggregate mean annual q, and EDR. Mathematically, the EDR values for each cause of death are additive to the total EDR. There is, however, no simple relationship between the MRs for individual causes of death and the total MR.

Tables Based on Cumulative Survival Curve Data

1. Tabular P Data. These are more accurate than measurements from a survival curve graph provided the results are given to three places (72.5% or 0.725, rather than 72% or 0.72). Interval survival rates are calculated with the needed accuracy as the quotient of the P values at the end and the beginning of the desired interval. The mean annual 4 is then derived as 1 - 4, where 4 is the geometric mean of the interval survival rate. If the P data are reported faithfully, values to only 2 decimal places produce less accurate values of mean annual 4 than if they are given to 3 decimal places. Three-place P data are given in Table 628A at 5, 10 and 15 years in various groups of post-MI men followed in Dublin, Ireland. The data in this table include P, P', the survival ratio, and mean annual d, d', EDR and MR as 100 d'/d. More often the data for survival ratio may be omitted, and a column of interval pi data is provided, giving the intermediate step in the derivation of geometric mean annual 4. Remember, if early deaths have not been excluded, although they should have been, it is necessary to adjust each P value by dividing by the po value, and to adjust the first interval to start at the end of the early period.

2. Graphic Survival Curve Data. The estimation of P from measurement of points on the curve of a graph is beset with error because of (1) the generally small scale of the graph, (2) inaccuracy in plotting the curve, and (3) optical distortion in photographic reproduction of the graph. The survival scale may be truncated, e.g., 100% down to below 50%, interruption in scale, then 10% to 0%. In this case measurements must be made of Q rather than P; downward from a horizontal line at P = 100%, parallel to the time scale at the bottom. The length of such a vertical scale must be derived from the measurement of the portion that is not truncated, in this case twice the measurement from P = 50% to 100%. With a millimeter rule and a magnifying glass our experience has been that the best one can do is measure to the nearest 0.2 millimeter, and we recommend that you measure the scale and each point several times and use the average for each point. As a decimal, P is then calculated as the ratio y(point)/y(scale), where y is the vertical measurement in mm. Occasionally the graph gives a curve for Q, not P. The measurement process is the same, and then P = 1 - Q. You must observe the same precaution in adjusting the P values from the graph to exclude the early mortality experience, if this is indicated and has not been done in construction of the graph. For the duration intervals chosen, such as 5 years, interval survival rates, Pk, may be derived as the quotient of the appropriate P values, and then the geometric mean annual survival rate calculated as follows: 4 = 1 - 4k/2, if the interval is 5 years (k = 0.2). Finally, q = 1 - p. The mean annual mortality rates are preferred to interval mortality rates, because they are more closely related to the aggregate means that are always used when E and d are available. Mortality ratios based on cumulative or interval mortality rates are always lower than annualized mortality ratios. Furthermore, interval mortality rates cannot be used for the direct calculation of EDR.

Derived rates are given in the table to 3 decimal places, for example the final geometric mean annual 4 might be 0.026 or 26 per 1000. However, it should be recognized that the error of this derived rate could be ±0.003, or more than 10%, because of the error of measurement of P. Another example of a table based on survival rates is Table 671B, giving results on mitral valve replacement by duration and type of prosthetic valve, all ages combined. After the duration column the variable are P, Pk, 4, 4k, EDR as (4 - 4k), and MR. Readers unfamiliar with this type of table may be baffled by the derivation of 4; the relationship can be given in a footnote, for the sake of completeness. It is always desirable to have tables (and figures)
self-explanatory, without forcing the reader to go to the text.

3. Graphic Survival Curve with Values of \( l \) by Duration. In this situation you should remember that \( 1 - b = d_1 + w_1 \). It may sometimes be possible to reconstruct a life table by a method of successive approximations, given the various relations, 

\[
E = l - 0.5w, q_1 = 1 - (P_2 / P_1), q_1 = d_1 / E_1,
\]

and the fact that the sum of the individual \( d \) values must equal the total deaths as given in the article. If the total is not stated, a valuable check of the allocation of \( d \) by duration is lost. The calculation of an \textit{annual} \( q \) from successive annual \( P \) values is generally inaccurate. If annual deaths tend to be relatively small and \( w \) values large, this approximation method would be impractical, and it would be necessary to rely on the usual development of the survival rates, without any use of the \( l \) values. Despite the uncertainties, a successful life table approximation by this method will probably yield EDR and MR results that are more accurate than those derived from the survival curve alone. However, in the text or in a table footnote or both it should be emphasized that \( E \) and \( d \) are not exact observed data, but approximations.

Other Tables

1. Clinical Trials. A randomized double-blind clinical trial is generally considered to be the acme of statistical testing of the effectiveness of a type of therapy by comparing the test group mortality or morbidity against an age/sex-matched control group, with similar characteristics. Despite the matching process, both groups should have expected rates of deaths calculated from the appropriate standard tables, just as in observational studies. Failure to do so may obscure features of the selection process that are important: for example, in the clinical trial conducted under the auspices of the University Group Diabetes Program, controversy arose about the validity of the finding of no reduction in mortality with use of an oral drug. The clinical investigators and those on both sides of the argument never tested the mortality of the control group, which one of us (RBS) found to be almost the same as that of the general population, indicating selection of a very mild set of diabetics for both control and test groups. Comparative mortality should therefore be shown for both the placebo and the treatment groups. An example of a table based on a clinical trial is 639A, giving mortality by age and sex in a randomized trial of propranolol in early survivors of acute MI. The first two columns contain the age group and proportion of females. Subsequent columns give the data for \( l, d, d', MR, \bar{q}, \bar{q}', \), and EDR. These results were derived from one of the reports of the “BHAT” Study, a large multicenter study sponsored by the National Institutes of Health. In this study, as in many clinical trials, older MI patients were excluded (age 70 and up in this instance). The significant reduction in mortality from that in the placebo group is reflected in the lower EDR and MR values found in the group treated with the chosen beta-blocker. Numbers of deaths were very small under age 40, but were substantial at older ages, and excess mortality increased with advancing age in a pattern typical of post-MI patients.

2. Intra-Series Comparison. In some large-scale prospective studies such as the Framingham Study or that of the Pooling Project, the total cohort or a large fraction thereof is used to provide the “expected” mortality experience. The total series is subdivided into parts, for example quintiles, according to levels of systolic or diastolic blood pressure, and comparative mortality is derived. In Table 601B this has been done for results of the Pooling Project in this manner. The table shows the rank of the quintile in the left-hand column, and comparative mortality in two sets of columns, systolic pressure on the left and diastolic pressure on the right. The data reported are \( d, d', MR \) and EDR. In this table format \( E \) has been omitted to conserve space, but the presence of positive EDR values indicates that \( E \) was observed. With this type of expected mortality, in the Pooling Project, all subjects without ECG abnormalities, some of the group divisions will have lower mortality and some higher than the average for the total. As a consequence some MRs will be under 100% and some EDRs will be negative (Ed Lew chose to omit the negative EDR values from the tabular results of this study).

3. Abridged Risk Ratio Data. The authors of some follow-up studies have chosen to present only the final comparative result in terms of what they call a “risk ratio,” given as a decimal to one place, or sometimes an “odds ratio,” if a morbid event is under consideration. These are age/sex-adjusted, and are equivalent to a mortality ratio or a morbidity ratio, but they are not expressed as a percentage. If an abstract is prepared from such an article, only the MR results can be given. EDR cannot be calculated, and the authors generally do not give the observed data from which the risk ratio has been calculated. This form of data presentation is often used in occupational risk studies. Special significance may be given by the statistician to terms such as odds ratio, if a case control study is involved (see below). “Relative risk” is another term that is sometimes used.

4. Case Control Studies. Usually regarded as retrospective studies rather than prospective, the design involves identification of a group of patients or subjects from past records with a particular type of treatment modality, and from current follow-up the mortality or occurrence of a morbid event, such as development of a type of cancer. The database must be such as to permit identification of a control group, matched by age, sex, and perhaps confounding risk factors, who did not have the treatment under study. A large database of nurses has been used to produce various case control studies on questions such as the possible relation of progesterone to breast cancer. Such studies have generally not been used for the preparation of mortality or morbidity abstracts, but they might be used if their limitations are recognized. See Feinstein and Horwitz’s excellent review.?

5. Attained Age Cohort. In chronic diseases that are severe or conspicuous enough to result in most cases in the population coming to medical attention, it may be justified to create a cohort of a series of such patients and analyze comparative mortality by attained age from an early age as a starting point. An example of this is shown in Table 663A, results of 257 patients with Marfan’s syndrome, diagnosed and followed at a Johns Hopkins Clinic over a period of many years. Survival curves from age 10 to age 60 were used to construct compar-
ative mortality results by decade of attained age. After the age column the data shown in order are $P$, interval $L$, $Q$, $Q'$, $MR$ (interval $R$), annualized $q$, $q'$, $EDR$, and $MR$ from the annual mortality rates. Another example is based on data from a Danish registry of patients with Tetralogy of Fallot alive on January 1, 1950. Incidence rates of this severe congenital heart disorder and numbers of live births back to 1890 were used to estimate cases and deaths by attained age. Such reconstruction of life tables from observations of this sort constitute a real challenge to the student of life table methodology. Very careful consideration must be given to the circumstances of series of this sort. Generally it is not permissible to use retrospective exposure from date of first contact with a patient back to date of onset, because deaths prior to first contact have been automatically eliminated. Congenital conditions present a special situation.

6. Organizing Data from Several Articles into a "Composite Abstract." This task arises either from the need to increase the aggregate number of deaths by giving results from several small series of patients, or from the desirability of showing together the results of several series of patients with a particular feature of their generic diagnosis, for example results of several series of patients with single-vessel disease after coronary bypass surgery (CBPS). Quite a few examples of this sort may be found in the new reference volume. One example is Abstract 648, in which post-CBPS patients are analyzed with respect to age and sex. In the long-term results the usual detailed life table format is used: age, reference to series, $I$ (at 30 days after operation), $E$, $d$, $d'$, $MR$, $q$, $q'$, and $EDR$. Results in this case are sufficient to fill 6 pages, even though the basic information on the various series had been gathered previously. It is very difficult to attempt general instructions for preparing composite abstracts; everything depends on the nature of the data and the desired objectives. In the new reference volume several composite abstracts were prepared for acute MI and for CBPS patients. Miscellaneous limited data were also combined in the 1976 Medical Risks volume at the end of each major disease section. These may be studied as examples of what may be achieved with such a presentation.

Conclusion

There appears to be a strong consensus within ALIMDA membership, including the Executive Council and the Mortality & Morbidity Committee, that mortality abstracts are a useful supplement to the intercompany Impairment Studies, which provide the data essential to medical underwriting and risk classification. The Board of Insurance Medicine now sponsors the Mortality Methodology Courses as one requirement for Board certification. Some members of ALIMDA worked hard with members of the Society of Actuaries in the formidable task of preparing the two Medical Risks reference monographs. Our Journal of Insurance Medicine, under the editorial leadership of John Elder, now offers high-quality mortality abstracts in each issue. The ALIMDA Research Center was started two years ago, partly with the idea of facilitating systematic search for and retrieval of FU articles from the medical literature. Despite all of this support in principle, the production of mortality abstracts for publication in the JIM has remained modest. Efforts to organize greater production and encourage a wider network of contributors has not met with much success, but there has been a stirring of professional interest.

Last May in Portsmouth, NH, a two-day workshop sponsored by the Board of Insurance Medicine provided a lively "hands-on" opportunity for ALIMDA members to intensively study methods of Life Table analysis and to formally abstract an article from a medical journal for themselves. This present article is, in part, an outgrowth of that workshop. Another such workshop is scheduled for this May in Portland, ME. It is our sincere hope that the practical suggestions, hints, and empirical rules offered in this article and in these workshops will encourage more than just a few medical directors to try their hand at the rewarding task of transforming the results of a medical FU study into a high-quality abstract.

Among the things that distinguish insurance medicine as a discipline are its quantitative approach to matters of prognosis and outcome and its application of actuarial methods to risk classification and assessment. Formal mortality abstraction is only one part of the universe of quantitative reasoning that medical directors engage in, but it is still an important part. Not merely a relic of our history and heritage, it clearly involves skills and perspectives that are also part of our future as a profession. Making these methods more generally accessible, identifying and overcoming obstacles to their use, and supporting a spectrum of approaches from the soundly reasoned, curb side estimate to the detailed, published abstract are goals worthy of pursuit!

This article is by no means exhaustive in its description of topics. Nothing has been said about articles involving survival models, meta-analysis, life expectancy, the intricacies of statistical analysis and the like, but there is much that each of us can do by simply applying basic methodologies to the abundantly available medical literature and generating valuable mortality abstracts. Let's do it!

REFERENCES