Minding Your P's and Q's

GETTING THE MOST FROM INCOMPLETE DATA:
SOME THINGS ARE SOLVD, OTHERS ARE NOT

Michael W. Kita, MD

This issue marks the second anniversary of P's and Q's, which has become something of a regular feature of the *Journal of Insurance Medicine*. Intended primarily as a commentary on methodologic aspects of life table analysis, it also from time to time addresses other quantitative issues in medical actuarial studies, and offers occasional notes of morbidity or mortality interest. Each "P's and Q's" commentary typically begins with a discussion of the abstracts that are found in the same issue. Many of the observations are self-explanatory. Others can be understood quite easily if the reader has a copy of the particular abstract — or at least its key tables — close at hand. Sometimes the reader would also benefit from having a copy of the original article (the one being abstracted) handy as well. That way one can most readily see what materials the abstractor had available to work from, and what kinds of original data (graphs, tables, text) lend themselves to life table representation.

This issue of the *Journal* contains two abstracts. The first is Singer's abstract on congestive heart failure (CHF). It is a composite abstract, synthesizing the results of the separately reported "Treatment" and "Prevention" arms of the SOLVD investigation (Studies of Left Ventricular Dysfunction). It is also a combined mortality/morbidity abstract, focusing chiefly on mortality experience, but also reporting (Table 370M1-6) various event-rates of interest. Singer's abstract presents four-year follow-up (FU) data on nearly 7000 patients, with FU 100% complete on the Treatment ("ever-symptomatic") group, and 99.8% complete on participants in the Prevention ("incipient" or never-previously-symptomatic) group. It represents a randomized, double-blind, placebo-controlled, multinational, multicenter clinical trial, compiling over 20,000 person-years of exposure.

The second is a morbidity/mortality abstract on colorectal cancer screening using rehydrated stool guaiac tests. It is a unizonal, multiserial, multisecular[1] randomized clinical trial providing 13 year follow-up of over 40,000 subjects with follow-up 100% complete. This represents a combined experience of nearly 550,000 person years of exposure.

Vice President & Chief Medical Director, UNUM life Insurance Company of America, Portland, Maine.

The abstracted articles supplied information deemed to be of grade A value ("Guidelines"[2]) when judged by size of the reported mortality experience, and extent of existing information on the condition studied. They also satisfied the minimum data requirements of Checklist A ("Finding Suitable Articles"[3]). However, the fact that the source-articles did not meet the checklist's *optimal* data requirements forewarns one to expect some methodologic difficulties in preparing life tables. Two basic challenges arise in these circumstances: 1) how to make the most of the *observed* data, and 2) how to make some estimate of *comparative* mortality. For each of these the abstractor must choose whether to use only what he is given, or to develop what he needs.

Making the Most of Incomplete Data

The ardent abstractor is always on the lookout for information that can be transposed into life table format. He or she is driven by two ambitions: 1) not to settle for summary (total) results when interval results can be deduced, and 2) not to settle for geometric (annual) rates when aggregate rates can be determined.

There are times when only summary data are available. They may be worthwhile, of course, but a one-line life-table is a bit pitiful to look at (see for example Table A in "Mortality Among Workers Exposed to Ethylene Oxide,"[4] which was called a Mortality "Extract" instead of a Mortality Abstract for this very reason). Before conceding that summary data must be settled for, the abstractor might attempt to locate other published reports. Sometimes supplementary data are reported elsewhere, or earlier (partial) results may have been reported previously. Any collateral reports of this nature could permit some tabulation of experience by duration. Best of all, of course, is if the original authors can directly provide the needed information that was not included in their published report, and many are responsive to such requests.

Sometimes interval information is provided in indirect fashion — via cumulative survival or mortality data, interval death tabulations, etc. But if there are no accompanying exposure data, or if exposures cannot be recreated from available information, then aggregate annual rates (q) will not be determinable. Lacking exposures,

one must settle for quotients (eg. d_i/ℓ_0 ratios) or geometric annualized rates. Why does it matter? Aggregate rates are more informative inasmuch as they preserve information about the *pattern* of mortality (temporal distribution of deaths, early and late). By definition, such rates are exposure-weighted and require exposure data in order to be calculated.

If the study being abstracted is a single-decrement study, then exposure is readily derived from ℓ_0 and interval d data (for example, Table A, "Mortality in Asymptomatic Patients with Carotid Bruit"[5]). A single-decrement study is one in which the only terminations to follow-up are by death (d). When study subjects are lost to follow-up or withdrawn from the study, there are now two ways (w and d) that the number of subjects remaining alive and under study can be decreased. The result is a double-decrement study, which is the typical clinical situation, and the circumstance found in both the SOLVD and stool guaiac studies abstracted in this issue. When one sees the comment that "FU was 100% complete," it does not mean that there were no withdrawals (w). Rather, it means vital status was known on all participants through end of follow-up. No one may have been *lost* to FU, but significant numbers could have been "withdrawn alive" during the course of the study. This is especially likely if rolling-entry periods were used. Both abstracts in this issue had to deal with missing information related to such withdrawals.

Getting SOLVD

What did Singer have to work with? He had some interval and summary data, but no explicit exposure information. The two SOLVD articles provided him with cumulative mortality graphs (Q) showing results by year, along with "ℓ" data at 6-month intervals. No information on interval withdrawals was given. But the available data was adequate to permit estimation of exposure by the life table reconstruction method.[6] In the case of the Treatment group, information contained in the author's table 3 supplied "d" for each annual interval and permitted the inference — by matching to information from the cumulative mortality graph — that no withdrawals (w) occurred until the end of year two. For the Prevention group, information on the range (minimum and maximum) and average duration of FU permitted an estimation of the temporal pattern of withdrawals. The number of withdrawals was inferred from the number of deaths (derived from cumulative Q information) and numbers alive at start of successive intervals (e.g. $\ell_{i+1}-\ell_i=d+w$). Singer next derived the interval exposures $(E_i = \ell_i -\frac{w_i}{2})$ and calculated exact q's (as $\frac{d}{E}$) for each annual interval. He also used survival data to derive approximate interval q's $(1-\frac{P_{i+1}}{P_i})$ to cross-check his reconstructed results.

Singer's next challenge was to come up with a reasonable "expected mortality" to use for comparison with his observed mortality. The process involved deciding what tables to use, and then how best to enter them and advance by duration. He needed to formulate a mean q' for the first interval, and then annual q''s for succeeding durations. However, the authors supplied only an age range, mean age, and male/female percentage — not a detailed age/sex composition of their various groups. Should he work from mean age, or could he do better?

Appreciating that the majority of study subjects had an MI history, Singer turned to the *1990 Medical Risks*[7] and found in Abstract #607 (Table 607A, page 6-86) several representative age and sex distributions for MI patients. Noting that the SOLVD subjects ranged in age from 21 to 80 years (mean age around 60) and 14.5% were female, he created an age and sex distribution that could be presumed for the SOLVD study. He adjusted the composition, reducing the female proportion of Abstract #607 and augmenting the percentages of younger age individuals until the known M/F percentage, mean age, and age range for the SOLVD group were duplicated. This furnished slightly different distributions for the Treatment and Prevention groups.

Singer then chose as the basis of expected mortality the 1975-1980 Basic Select tables. Why the select tables — why not U.S. Population tables? One could argue that the SOLVD subjects are, if anything, a "sick" subset of the general US population, and thus ought to be compared to US population life table rates, as expected mortality. But Singer observed that the number and type of exclusions to which study participants were subjected amounted to fairly rigorous selection. Since "SOLVD eligibles plus selection" constitute the study subjects, with their observed mortality rates, "a general population plus selection" would be an appropriate comparison group for expected mortality. Such a select U.S. population is what is found in the Basic Select tables. Since the study period was 1986 to 1991, the 1975-80 Select tables are the most nearly contemporaneous tables.

Singer used his presumed age-distribution to derive initial mean q''s. He entered the basic table opposite the various ages, for male and female separately, and combined these fractional contributions to get an overall mean q' for the first year. He advanced mean q' for each duration by reentering the table under the next policy year and recalculating the age-and-sex-weighted par-

tial contributions. The q' values tabulated in Singer's Tables 370M1-2 and M1-3 are the final product of this exacting approach.

With mean q' now available for each duration, and for both Treatment and Prevention groups, Singer could at last derive MR's and EDR's. By means of a life-table reconstruction, he thus developed exposure data that was otherwise unobtainable, and could base his MR's and EDR's on aggregate rather than geometric rates.

A Closer Look at the Stool Guaiac Study

The source-article on Colorectal Cancer and Stool Guaiac Screening also furnished a mixture of summary and interval data. Exposures were reported in summary fashion (all durations combined) for the three sub-groups followed, so aggregate rates could be calculated for overall mortality experience. Could aggregate rates for *interval* experience be reconstructed, much like Singer did? While the 11.84 year average duration of FU in this 13 year study suggests that those "withdrawn alive" affected exposure only in the late durations of the study (w approximately zero through the first 11 years of FU), one would need all-cause mortality by duration in order to attempt reconstruction of the complete double-decrement table. The reason for this is that when cause-specific mortality (eg. colorectal cancer mortality) is being studied, other causes of death are equivalent to censoring (withdrawal from study). In this respect, a cause-specific mortality study is something like a morbidity study, with a category of deaths being treated as withdrawals.

Lacking sufficient data for interval exposure derivation, one cannot calculate aggregate interval mortality rates. But geometric rates are still possible, and in Table 991M1-4 they are shown (q) for three 4-year intervals. An overall aggregate rate (q) for all durations combined is also shown, since overall exposure data were reported. Table 991M1-2 gives geometric 5 and 13 year annualized mortality rates, and Table 991M1-3 uses cumulative exposure data to report some pertinent aggregate annual rates.

The Stool Guaiac Study also examined the predictive value of fecal occult blood testing. They used as their gold-standard for true positives any positives that resulted in discovery of colorectal cancer within the next year. The cancer need not have been diagnosed at the time of stool positivity, but at any time within the next year. This relatively wide time-window probably enhanced the positive predictive value (PPV), but the authors did not comment on how much of an effect this had. They did comment on the decrease in PPV that was

associated with the rehydration of stool cards before testing. Rehydration enhanced sensitivity (reduced false-negative results) at the expense of lowering specificity (increasing false-positive results). This trade-off is inevitable when a change in a decision point (positive/not positive) repartitions the underlying populations being tested (those having the condition in question — colorectal cancer — and those not having the condition in question — negative for colorectal cancer, although possibly having other conditions associated with occult bleeding). Some results can be "false-positive" *for colorectal cancer* screening and yet still be meaningfully "positive" for polyps, ulcers, diverticular disease, and other conditions of clinical significance. In other words, while the occult bleeding is not necessarily false, the conditions causing it may be false-positive when the definition of true-positive is narrowed to colorectal cancer. How sensitivity, specificity, and predictive value inter-relate is discussed in more detail in a prior *Journal* article, "Drawing Conclusions from Test Results."[8]

Loose Ends

In the last issue, a study of "Abdominal Aortic Aneurysm and Risk of Rupture" was presented as an abstract. The same week that abstract was published in this *Journal*, the *New England Journal of Medicine* published Ernst's review, "Current Concepts: Abdominal Aortic Aneurysm."[9] In addition to current information on demographics and health care costs of this condition, Ernst summarizes the natural history data available from a number of large longitudinal studies. The risk of rupture is only briefly commented on, but 80% of aneurysms are said to increase progressively in diameter, and 20% at a rate of more than 0.5cm per year. Five-year rupture rates for aneurysms of 4-5 cm initial size are given as 3-12%. Screening is considered cost effective at higher ages (where the prevalence is great), especially in those with vascular risk factors (hypertension, femoral or popliteal aneurysm, or family history of AAA). Mortality data for aneurysm repair, with rupture and without, are summarized in author's tables 1 and 2. Late survival after repair is given in Table 3 for four of the largest studies to date.

In another recent issue of the *New England Journal*, two articles discuss the survival, with and without surgical intervention, of an 87-year-old woman with aortic stenosis and concurrent coronary artery disease. The first article, titled "Too Old for What?,"[10] is a clinical problem-solving piece with an accompanying discussion of the decision analysis issues pertaining to surgical intervention. The discussant focuses on the quality of life issues, believing the effects on survival to be modest,

given the patient's advanced age. The second article, titled "You're Never Too Old,"[11] is a more formal deci-sion-analysis of the therapeutic options. Using a Markov model, the authors find improved survival (longevity) to actually be the principal benefit. Their conclusion contains the following acknowledgment: "We are indebted to Richard B. Singer, M.D. for his contributions to our understanding of the mortality experience used to estimate the survival of healthy elderly populations."

The authors — Wong, Salem, and Pauker — are mem-bers of the Division of Clinical Decision Making (Deci-sion Analysis Lab) at Tufts-New England Medical Center in Boston. The Committee on Morbidity and Mortality (CMM) of the Academy has enjoyed a fruitful collaboration with Pauker's group over the last few years. Members of the Academy have attended Deci-sion-Analysis Workshops sponsored by the Tufts Lab, and been introduced to Markov modeling, decision trees, and the DEALE (Declining Exponential Approxi-mation to Life Expectancy). Pauker and some of his staff attended last year's Advanced Mortality Methodology Workshop and made significant contributions to dis-cussions of survival methodologies and life expectancy estimations. A particularly memorable exchange oc-curred on the subject of when rates are probabilities, and when they are not. When Pauker's Lab needed data on survival among healthy elderly lives, the CMM responded with a variety of life insurance, annuity study, and social security experience, and Singer con-tributed his personal expertise to the selection of appro-priate tables. Future collaborations are likely, and Wong has promised an article on hazard functions and sur-vival distributions in a forthcoming issue of this *Journal*.

The next Advanced Mortality Methodology Workshop will be held in Toronto October 6-8 at the conclusion of the joint AAIM/CLIMOA meeting. It will be given at the Toronto Marriott-Eaton Centre, the same hotel where the Academy meeting is being held, and not at a different hotel as an earlier notice incorrectly stated. Otherwise, the notice and application form on page 315 of the Winter 1992 issue of the *Journal* accurately set forth the requirements and expectations of the course. It will be limited to 8 students, and several spaces are still open at this time. Attendance from qualified actu-aries and underwriters, as well as medical directors, is invited, and at this writing, four medical directors and one actuary have registered. If you are interested and believe that you qualify, a registration form should be completed and submitted now. Some pre-work needs to be completed by each student before the course be-gins. Participants are assured an interesting educational experience, and by the conclusion of the course will have completed much of the work necessary to produce their own publishable abstract

References

1. Kita MW. Durations, truncations, and Kaplan-Meier. *J Insur Med* 1993;25(1):15-18.

2. Singer RB, Kita MW. Guidelines for the evaluation of follow-up articles and the preparation of mortality abstracts. *J Insur Med* 1991;23(1):21-29.

3. Kita MW. Morbidity/Mortality abstraction — Finding suitable articles. *J Insur Med* 1990;22(4):287-8.

4. Kita MW. Mortality among workers exposed to ethylene oxide: 16-Year average follow-up of over 18,000 workers. *J Insur Med* 1991;23(3):176-79.

5. Kita MW. Mortality in patients with asymptomatic carotid bruit. *J Insur Med* 1991;23(4):270-71.

6. Singer RB. Life table reconstruction from annual data for survival rate, P, and entrants, l. *J Insur Med* 1991;23(4):274-6.

7. Lew EA, Gajewski J, eds. *Medical Risks: Trends in Mortality by Age and Time Elapsed*. New York, Praeger, 1990.

8. Kita MW. Drawing conclusions from test results. *J Insur Med* 1990;22(4):270-78.

9. Ernst CB. Current concepts: Abdominal aortic aneurysm. *New Engl J Med* 1993;328(16):1167-72.

10. Thibault GE. Too old for what? *New England J Med* 1993;328(13):946-50.

11. Wong JB, Salem DN, Pauker SG. You're never too old. *New Engl J Med* 1993;328(13):971-4.