

6/22/82

WHEN DID THE RUSSIAN HEALTH CRISIS START?  
MICROCOMPUTER STRATEGIES FOR AN ANSWER

Irwin D.J. Bross, Ph.D.  
Director of Biostatistics  
Roswell Park Memorial Institute

Abstract

In an ongoing investigation of the current health crisis in the Soviet Union, a key question arises: When did the death rates start going up? This is a typical practical solution having no ready-made computerized procedure. The advent of the "microprocessor revolution" provides a new strategic approach to an answer.

Key Words

U.S.S.R. mortality rates, microcomputers in statistics, estimation of the starting point of a change

Running Title

Russian Health Crisis

Reprint order form should be addressed to:

Dr. Irwin D.J. Bross  
Director of Biostatistics  
Roswell Park Memorial Institute  
666 Elm Street  
Buffalo, New York 14263

## WHEN DID THE RUSSIAN HEALTH CRISIS START?

### MICROCOMPUTER STRATEGIES FOR AN ANSWER

#### A SCIENTIFIC QUESTION

During the past two years the media has carried occasional news stories on the health crisis in the Soviet Union that is now occurring. Most of the accounts have emphasized the turnaround in infant mortality in the U.S.S.R. Whereas most of the advanced technological societies have infant mortality rates that are gradually decreasing (which was also the Russian pattern in the 1960's), there was a rapid increase in infant mortality in the Soviet Union in the 1970's. While the causes of the Russian health crisis are still being debated, it is important to try to pinpoint these causes as quickly as possible. Something has gone wrong in the U.S.S.R. If the causes can be identified, other advanced technological societies may be able to avoid the fate of the Soviet Union.

While the striking effects for infant mortality have justifiably received most of the attention, there is also indication of an upturn in mortality rates for adults. If, as seems likely, the causes for the increase are technogenic ("techno-" = technology, "genic" = caused by) and related to genetic degradation in the Soviet population, then the health effects would be expected to occur in both infants and adults. In trying to pinpoint the causes, one of the first questions in an epidemiological inquiry would be: When did the Russian health crisis start? When did the death rates start to go up?

Because of the declining infant mortality rates in the 1960's (as compared to stable adult mortality rates), it is somewhat easier to answer this question with the adult data. Furthermore, it is desirable here to have answers that are statistically objective rather than answers subjectively obtained by casual inspection of the data tables. Various hypotheses about the starting time might be formally tested. Since data is available on age-specific mortality rates in the adults, the starting date can be estimated for the different age groups. An important hypothesis to test is that the starting point is similar for all age groups. Some possible causes or etiologies can be confirmed or disconfirmed by the pattern of the starting points.

Strategically speaking, the situation that the analyst encounters with this problem is a common one in many practical applications. In theory, the problem is not particularly difficult. In practice, because of the nature and limitations of the actual data and the special needs of the scientific context, a ready-made solution is not at hand. The working statistician has long had several different strategic options in looking for a practical answer in such a situation. In recent years, a new option has emerged with the much-publicized "microprocessor revolution".

Since many statisticians have worked with the larger computers (often called "mainframes"), they have probably been less impressed than most others with the tactical or technical capabilities of the new microcomputers. Indeed, the micros have some serious inadequacies relative to mainframes. Because of the focus on tactical aspects, the

strategic potential of the "microprocessor revolution" in statistics has sometimes been overlooked. The strategic aspects are the focus here.

To see how microcomputers offer a new strategic option to working statisticians in situations where there is no ready-made solution to an immediate problem, consider briefly two of the older options. One traditional option is to go to the statistical literature. This option is favored by academics but is not as attractive to a working statistician under time pressure. In this example, there would be many articles in the literature on the general theme of finding the starting point of a situational change. However, to get, read, and try to adapt the published results to this data would be a time-consuming and not-necessarily-productive approach. Another older option (available to those who work with mainframes and with software packages such as SPSS, Minitab, or others) would be to try to put together an estimator for the starting point by combining the pieces of an existing software package. As with other options, the feasibility of this approach depends on the experience and skills of the analyst. Thus, while some may be adept at exploiting internal capabilities of a given software package, analysts who are not fully familiar with the architecture of the programs would not find this an easy option.

The option that has emerged with the "microprocessor revolution" is to find a de novo solution to the problem that is tailored to the needs of a particular situation. One strategic potential of the microcomputers is their "friendly" nature. They may have helpful, automated programs designed to guide a particular user through computer operations

(generally programs on floppy disks). This reduces the level of programming skills needed to do a job.

Another strategic potential of the micros is their ability to do many statistical jobs by "brute force"--by programs that may be inelegant or inefficient but nevertheless will deliver the results (perhaps in minutes instead of seconds). A specific example of this capability is "brute force minimization" (which will be illustrated here).

The strategic import of this minimization capability lies in this fact: Most effective biostatistical techniques are mathematically derived as extremal solutions (e.g., maximum likelihood, least squares, minimum Chi-square, etc.). The extremal solutions for many practical problems (such as estimation of the starting points) are mathematically complicated. They are not something that can always be ground out quickly and surely. By using brute force computing power now readily available in the microcomputers, de novo extremal solutions to practical problems become computationally feasible alternatives to the traditional mathematical derivations. Admittedly, the mainframes also have this potential. However, it has not been very much exploited, perhaps because most computer centers do not offer the "friendly" features of the micros.

#### A MICROCOMPUTER ANSWER

Most of our knowledge about the Soviet health crisis comes from an extraordinary publication of the U.S. Bureau of Census on "Rising Infant Mortality in the U.S.S.R. in the 1970's" issued in September 1980

(Davis and Feshbach, 1980). The authors have done an outstanding job of reconstructing the vital statistics of the U.S.S.R. from various official sources. For technical reasons, they have used two-year moving averages of yearly rates which express the number of deaths per 1000 in a given age group. Table 1 of this paper is extracted from Table 1 of the original paper titled "Reported and Estimated Age-Specific Death Rates in the U.S.S.R.: 1958 to 1976".

As can be seen from Table 1, there are often peculiarities of actual data that make it difficult to use the general extremal solutions in the statistical literatures. For example, the two authors note that an inexplicable jump in the death rates of persons 70 years of age or older occurred between 1968/69 and 1969/70. One way to handle this is to eliminate this age category. However, other difficulties, such as the gap in the rates for 1961/62, are less tractable. To illustrate how such missing data can be handled with microcomputer strategies, the missing year (and 1960/61) have been retained in Table 1.

Conceptually, the estimation of the point at which the rise in Soviet mortality rates started is simple enough. In principle, it is only necessary to consider each possible year as potentially the "break point" or the start of the upturn. A straight line could be fitted to the data prior to this point and a different line to the data after it. The two lines are constrained by the requirement that they intersect at or near the break point. Deviations from the fitted lines can then be calculated. A plausible extremal condition would be the usual least squares criterion, minimizing the sum of squares of the deviations.

Here, the brute force minimization is relatively easy since it merely requires calculating the sum of squares of deviations for each potential break point and seeing which year (or, strictly speaking, two-year average) has the smallest numerical value. This kind of job is relatively easy to program on a microcomputer (even for an inexperienced programmer like the author). However, with only the computing resources available earlier in this century, this would be a tedious job. Thus, to answer the questions about the starting points at different ages, it is necessary to repeat the estimation procedure a number of times. If the starting points are similar, it would then be possible to get a single estimate for all ages.

Hence the option provided by the "microprocessor revolution" is not one that is "new" in the sense of being previously impossible. Rather it is one that is "new" because the conceptual solution now becomes relatively easy to implement in practice. The approach could have been tried 30 years ago--but relatively few analysts would have wanted to do the arithmetic. Now it is a job which many working statisticians with micros could do in a few hours or days.

The tactical or technical implementation of the brute force strategy can be done in various ways, but with the availability of microprocessor technologies the simplest and most direct methods are likely to be as good (or nearly as good) as theoretically optimum methods. Moreover, they are more easily communicated to the public if the issue is one which is likely to get into the public domain (as the example here almost certainly would). Here the easiest way to fit two straight

lines that intersect near the break point is this: Pass the first line through the two data points, the initial death rate in the time sequence and the death rate at the potential break point. Similarly, pass the second line through the value at the break point and at the final death rate in the time sequence.

Since the fit is "perfect" at the three points, three degrees of freedom in the sum of squares of deviations are lost. However, the number of degrees of freedom (i.e., Number of deviations - 3) is the same for all break points and this permits fair comparisons in the minimization process. Note that the potential break point is always excluded from the sum of squares. This simplification makes it relatively easy to combine sums of squares over age groups (or in other ways) for estimation and hypothesis testing.

The simplified procedure is not optimal under the usual normal distribution assumptions with equal variance (but it is optimal under other assumptions). The loss in efficiency is a price that many analysts may be willing to pay for ease of programming and simplicity of explanation. For most data sets, the efficiency loss is not likely to be serious.

The results of the brute force minimizations are shown in Table 2. The table includes two age groups whose inclusion might be questioned. The oldest age group has the unexplained artifact noted by the compilers of the table (Davis and Feshbach, 1980), and the youngest age group does not show much change over time.

One of the strategic advantages of microprocessor operations is that it is often little extra trouble to calculate unnecessary subgroups or to do jobs in two or more different ways. Hence, the need to make "iffy" analytic decisions before the analysis is reduced. Sometimes difficult decisions before the fact are easy; after the fact (for instance when rival statistical analyses give similar results). Here, for instance, even with the artifact in the oldest age group, the break point coincides with that of the other age groups. Also the exclusion of the youngest group can be justified by the analytic results.

The column labeled MSS in Table 2 shows the sum of squares of deviations for the line passing through the first and last mortality rates in the time series. If there is a break point, this parameter should substantially reduce the sum of squares (as can be seen by the minimum values in Table 2). The reduction divided by the residual mean square (F in Table 2) can be tested approximately with the usual F table.

The F tests are only approximate because of a technical difficulty that arises from the use of moving average (i.e., data points are not independent). In borderline cases the results might be in doubt. Here the effects are so pronounced that there can be little question that there are break points in the Soviet mortality rates.

#### A FEW COMMENTS ON STRATEGY

This note was a byproduct of an ongoing investigation into the causes of the health crisis that is now going on in the Soviet Union.

The existence of basic data in English and in familiar Bureau of Census formats for mortality rates in the U.S.S.R. was called to my attention by Gary Groesh during a hearing by the Nuclear Regulatory Commission (Docket No. 50-382) on proposed nuclear power plants on the lower Mississippi. My testimony dealt with genetic damage from low-level ionizing radiation, a focus of our research over more than a decade. It discussed in particular the synergistic effects reported in our three most recent papers (Bross and Natarajan, 1980; Bross and Driscoll, 1981; Bross, 1982). As a working hypothesis, I suggested that these synergistic effects may be one of the major causes of the health crisis in the Soviet Union.

The siting policy of the U.S.S.R. technocrats has been to concentrate most of the heavy industry (including chemical and nuclear plants) and population along the extended Russian river systems. This policy tends to be forced by the fact that the long river systems are the primary Soviet water resources. The end result is to produce a mix of chemical and radiological contaminants with a potential for synergistic effects.

Davis and Feshbach (1980) had originally identified pollution as a probable major cause of the Soviet health effects. This suggested a reanalysis of their mortality data to see if it would confirm or disconfirm the more specific working hypothesis that synergistic effects from low-level radiation might be causing the problems in Russia. Until this hypothesis has been checked out, it does not seem prudent for NRC

to permit the Soviet siting policy to be used on long U.S. river systems such as the Mississippi.

While this study situation, like any other, is unique, some of its strategic aspects are commonly encountered. Getting answers to questions about the starting point is only the first step in a continuing investigation. Thus it may be possible to develop more conclusive tests of the working hypothesis by correlating the geographic and temporal information on Soviet nuclear facilities with geographical data (by provinces) reported by Davis and Feshbach. The 1968/69 starting date for the upturn is consistent with what is generally known about Soviet nuclear development and the build-up of radiological pollution over time. The similarity of the starting dates in the age groups also fits in with the working hypothesis. Hence, some useful preliminary information has come out of this analysis. As in many analytic situations, an answer to a preliminary question (about starting points) is needed for analytic decisions in the subsequent analysis.

When there is need for prompt answers, the analyst cannot afford to get "hung up" on preliminary questions, however interesting, and must get on with the main job. The microcomputers here and elsewhere have opened a strategic option for getting "over the hump" by brute force minimization or other direct routes when ready-made or more elegant techniques are not readily available.

There is another, broader but more subtle, strategic advantage of the "microprocessor revolution" in statistical practice that the example also illustrates. The strategic concept here is related to

"getting out of the rut" or "freedom" or "teaching an old dog new tricks". We are often unaware of how often the "principles" on which we habitually act are rationalizations of lines of action which were forced by past circumstances but where options now exist. One might wonder whether Newton would have developed his theory of fluxions if he had a micro-computer at his disposal. Calculus, which we admire as an intellectual achievement, might have been only an asymptotic form of finite differences.

Similarly the emphasis on "bestness" or "uniqueness" of statistical methods is more a reflection of the preoccupation of mathematicians with necessary and sufficient conditions in formal proofs than criteria for what is desirable in practice. Indeed, "bestness" requires rigid assumptions which are never demonstrable in the real world. In practice, a "robust" technique is usually preferable to the "best" method. In my experience, simple and direct analyses are generally more "robust" and, as illustrated here, the "microprocessor revolution" has made it feasible to approach statistical problems more directly (and, therefore, "robustly").

Whereas "uniqueness" has highest priority for mathematicians, in practice there is great diversity. Analysts tend to approach data with a "style" that is comfortable for them as individuals. Some analysts are wedded to particular techniques or families of techniques or to particular software packages or computer languages. They may try to force the data to fit their preferences (e.g., by transformations to formats which will run in a canned program). Hence, the new freedom to rethink old problems which is made possible by microcomputers may not be

welcomed by all analysts. However, some analysts may welcome this strategic freedom and this case history may encourage them to explore new possibilities in their own work.

ACKNOWLEDGEMENT

Mrs. Deborah Driscoll has carried out the calculations for Table 2 and has otherwise assisted in this work.

REFERENCES

Bross, I.D.J. (1982). Letter to the Editor. Regarding "Genetic effects of the atomic bombs: a reappraisal", by Schull, et al. Health Physics, in press.

Bross, I.D.J. and Driscoll, D.L. (1981). Direct estimates of low-level radiation risks of lung cancer at two NRC-compliant nuclear installations: why are the new risk estimates 20 to 200 times the old official estimates? The Yale Journal of Biology and Medicine, 54(5):317-328.

Bross, I.D.J. and Natarajan, N. (1980). Cumulative genetic damage in children exposed to preconception and intrauterine radiation. Investigative Radiology, 15(1):52-64.

Davis, C. and Feshbach, M. (1980). Rising infant mortality in the U.S.S.R. in the 1970's. Bureau of the Census, U.S. Department of Commerce, Series P-95, No. 74, pp. 1-6, issued September 1980.

Table 1. AGE-SPECIFIC DEATH RATES IN THE U.S.S.R.: 1960-1976

Age Group	YEAR															
	60/61	61/62	62/63	63/64	64/65	65/66	66/67	67/68	68/69	69/70	70/71	71/72	72/73	73/74	74/75	75/76
25-29	2.1	(NA)	2.0	2.0	2.0	2.0	2.0	2.1	2.2	2.2	2.2	2.1	2.1	2.0	2.1	2.1
30-34	2.7	(NA)	2.6	2.5	2.5	2.6	2.6	2.7	2.8	2.8	2.8	2.8	2.8	2.8	3.0	3.0
35-39	3.0	(NA)	3.1	3.1	3.1	3.2	3.4	3.5	3.5	3.7	3.8	3.7	3.6	3.6	3.7	3.8
40-44	3.7	(NA)	3.8	3.7	3.8	3.9	4.1	4.3	4.6	4.7	4.7	4.8	4.8	4.9	5.2	5.3
45-49	5.4	(NA)	5.3	5.1	5.0	5.1	5.3	5.5	5.6	6.0	6.0	6.1	6.2	6.4	6.7	6.9
50-54	7.5	(NA)	7.7	7.7	7.8	7.9	7.9	8.0	8.1	8.7	8.7	8.8	8.6	8.8	9.0	9.3
55-59	10.9	(NA)	11.2	10.7	10.8	11.1	11.3	11.5	12.1	11.7	11.8	11.9	12.5	12.3	13.0	13.4
60-64	16.7	(NA)	17.5	17.1	17.2	17.2	17.4	17.8	18.2	18.0	17.9	18.1	18.0	18.2	18.3	18.9
65-69	24.6	(NA)	25.6	24.1	24.4	25.5	25.9	26.3	27.5	27.5	26.9	26.8	27.2	27.0	27.4	28.0
70+	63.0	(NA)	67.7	63.6	64.2	65.8	66.1	66.8	67.3	75.7	74.9	74.8	75.5	73.5	73.3	75.0

Table 2. RESULTS OF MINIMIZATION

Age Group	Starting Year	MSS*	RSS**	F
25-29	62/63	.09	.038	16.42
30-34	68/69	2.65	.271	105.34
35-39	67/68	5.93	.422	156.63
40-44	68/69	35.92	1.967	207.14
45-49	69/70	41.96	3.016	154.73
50-54	69/70	36.77	1.872	223.71
55-59	68/69	92.04	5.166	201.80
60-64	68/69	39.09	2.309	191.15
65-69	67/68	114.50	10.684	116.60
70+	68/69	1327.49	133.700	107.15

\*Initial Mean Sum of Squares

\*\*Residual Mean Sum of Squares