STATEMENT OF DR. JEROME CORNFIELD, CHAIRMAN, DEPARTMENT OF STATISTICS, THE GEORGE WASHINGTON UNIVERSITY, WASHINGTON, D.C.

Dr. CORNFIELD. I appreciate the opportunity to appear before this committee, and like Dr. Prout, will go directly to my statement.

In 1970, the UGDP reported its finding "that the combination of diet and tolbutamide therapy is no more effective than diet alone in prolonging life. Moreover, the findings suggest that tolbutamide and diet may be less effective than diet alone or than diet and insulin, at least in so far as cardiovascular mortality is concerned." With no evidence of efficacy and a definite possibility of toxicity, the investigators concluded that the safety of the patients receiving tolbutamide therapy required its discontinuance, and that the factual basis for this decision needed to be communicated to the biomedical community. This prudent decision and moderately worded conclusion was received by some critics with a hostility that had no discernible scientific basis. In particular, the statistical aspects of the study were attacked as unsound, often by medical critics who had shown no previous interest in principles of sound statistical design and inference in the conduct and interpretation of biomedical studies.

Mr. GORDON. Do you have any examples of that type of criticism?

Dr. CORNFIELD. Well, yes, I could name several. There is a letter that has been submitted to this committee, I believe by Dr. Moss,[1] which criticizes the statistical design and interpretation of the UGDP and quotes some data of his own which in terms of precision and sophistication of design are much less acceptable. For example in the UGDP, there were five treatment groups to which patients were randomly assigned, one of which forms a placebo or control group. But Dr. Moss had no concurrent control group and estimated his control mortality from life insurance statistics. It is curious that somebody that is not embarrassed by having designed such a study would nevertheless find statistical aspects of the UGDP weak.

In the following year I was asked by the editors of the "Journal of the American Medical Association" to analyze these and other criticisms.

The criticisms took several forms. It was charged that the patients assigned to the different therapies were not comparable, and in particular, that those assigned to tolbutamide therapy had the greatest cardiovascular risk and those assigned to diet alone the least, and that this, rather than tolbutamide itself, explained the excess mortality of these patients. It was also charged that the excess mortality with tolbutamide was confined to a small number of clinics and was not typical of what might be found in all clinics in the study or in general medical practice. There was a cluster of charges relating to poor conduct of the study, failure to follow the protocol and finally, charges published in the medical tabloid press of deliberate attempts to mislead by concealment of data. These charges were considered in full detail in an article published in the September 20, 1971, issue of the "Journal of the American Medical Association" and found to be uniformly without foundation. Because it is not possible to enter here into the

[1] See page 11134.

detail considered there, I ask permission to include that article in the record.

[Testimony resumes at page 10792. The information referred to follows:]

The University Group Diabetes Program

A Further Statistical Analysis of the Mortality Findings

Jerome Cornfield

The basic conclusion of the University Group Diabetes Program on tolbutamide is, "the findings of this study indicate that the combination of diet and tolbutamide therapy is no more effective than diet alone in prolonging life. Moreover, the findings suggest that tolbutamide and diet may be less effective than diet alone or than diet and insulin at least in so far as cardiovascular mortality is concerned." With no evidence of efficacy and a definite possibility of toxicity the investigators concluded that the safety of the patients still receiving tolbutamide therapy required its discontinuance and that the factual basis for this decision needed to be communicated to the biomedical scientific community. This prudent decision and moderately worded conclusion has been received by some critics with a hostility which has no discernible scientific basis. The following analysis is largely confined to Dr. Schor's analysis (see page 1671).

In general, independent repetition of a study is the most constructive way of analyzing it. If the evidence against tolbutamide is as weak as some critics appear to believe, an independent repetition of the UGDP would not only be called for, but could be ethically justified. If, on the other hand, no important sources of error or uncertainty in findings can be pinpointed, then a repetition might be considered unethical, and, if requirements for informed consent in this country are taken into account, not even possible. The subsequent analysis is undertaken to illuminate those alternatives and not to defend the UGDP. Its concentration on the strength of the evidence against tolbutamide should of course not be permitted to obscure the more general UGDP finding that lowering of blood glucose level did not appreciably lower the eight-year mortality from cardiovascular disease as compared with patients on diet alone.

Randomization

The results of a clinical trial can be interpreted only if the patients assigned to the different therapies are comparable in all relevant respects. Many of the criticisms of the UGDP appear to stem from failure to realize the role of randomization in achieving such comparability. Those who distrust randomization would prespecify "all" relevant variables and assure comparability by matching with respect to them. But those variables regarded as relevant are not always as critical as some expert opinion held them to be, and if other important but unknown variables are not specified, failure to achieve comparability with respect to them can wreck a study. The major function of randomization, either with or without prior matching or stratification on known relevant variables, is to achieve approximate comparability with respect to all variables, whether known or not. Thus, if 40% of the 409 patients in the UGDP who were assigned to either placebo or tolbutamide were cigarette smokers, then there is approximately only one chance in 50,000 that the actual percents after random assignment would be as different as 30% cigarette smokers receiving placebo and 50% receiving tolbutamide, even though one did not know the smoking histories of the patients assigned. Statements such as Dr. Feinstein's[1] that "randomization cannot prevent major inequalities that may occur prognostically in the 'luck of the draw'" either depend on special definitions of "major inequalities" or reflect an inadequate appreciation of the functions and power of randomization.

In practice, there are two ways in which randomization can fail. First, the randomization scheme may have "broken down," ie, have been deliberately violated in the hope of assigning some patients to a favored therapy. Second, even in the absence of deliberate violations, the "luck of the draw" may have resulted in extreme baseline inequalities. If deliberate violation occurred, study results are best quietly buried, but if baseline inequalities arise from bad luck, poststratification with respect to known variables and statistical analysis can often achieve what randomization failed to.

From the Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh.

Although the new data presented may be considered official UGDP data, the interpretations are my own and do not necessarily represent the point of view of the investigators or their Executive Committee.

Reprint requests to 9650 Rockville Pike, Bethesda, Md 20014 (Dr. Cornfield).

Table 1.—Distribution of Participants by Number of Baseline Risk Factors by Treatment Group*

No. of Baseline Risk Factors	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
0	28	25	22	19
1	60	50	62	76
2	59	58	60	57
3	26	34	34	30
4	10	17	8	4
5	2	4	8	4
6	0	1	1	1
Total	185	189	195	187
Mean No. of risk factors	1.65	1.92	1.88	1.72
Standard error	0.083	0.092	0.089	0.079

*The eight baseline risk factors considered are age ≥ 55 years, hypertension, history of digitalis use, history of angina pectoris, significant ECG abnormality, cholesterol level ≥ 300 mg/100 ml, relative body weight ≥ 1.25, and arterial calcification.

It is difficult to see how deliberate violation could have occurred in the UGDP, since each patient, identified by name and study number, was assigned by the Coordinating Center and not the treating clinic. Subsequent monitoring by the Coordinating Center would have detected any failure to comply with this assignment. But if such deliberate violation did occur, large baseline inequalities should reveal it.

Extent of Baseline Inequalities

In this section I shall examine the baseline inequalities among the treatment groups to see how bad the luck of the draw really was. The original UGDP report did consider individually 14 baseline characteristics, not one of which showed a difference significant at the $P=0.05$ level when all four treatment groups were simultaneously considered and only one of which showed such a difference when the tolbutamide-placebo comparison was considered alone (Table 7, p 800). Furthermore, to check on the possibility that these individually nonsignificant differences had all cumulated in the same direction, the report gave the percentage of patients with one or more of the following risk factors: hypertension, history of digitalis use, history of angina pectoris, significant electrocardiographic abnormality, and serum cholesterol level exceeding 300 mg/ml. These percents were virtually identical, ie, 47.3% for the placebo group and 47.9% for the tolbutamide group. Small excesses in the percentage of patients in the tolbutamide group for each of the last four risk factors were almost exactly balanced by a larger, but still statistically insignificant ($P=0.16$), deficiency in the percentage with hypertension.

It has been suggested that this comparison should have included three other risk factors: arterial calcification, age ≥ 55 years, and relative body weight ≥ 1.25. The distribution by number of possible risk factors (these three plus the previous five) is shown in Table 1 for the 756 participants for whom values for all eight baseline risk factors were available. Of the patients assigned to placebo, 157 out of 185, or 84.9%, had one or more risk factors as compared to 164 out of 189, or 86.8%, receiving tolbutamide. The average number of risk factors present among those assigned to placebo was 1.65 as compared with 1.92 among those receiving tolbutamide, an excess of about one-fourth a risk factor. All in all, the luck of the draw does not seem to have been too bad.

It is argued by Schor that comparisons based on the entire group of patients are irrelevant and should be confined to those seen for four or more years. But Table A-1 on p 817 of the Report shows that more than 800 of the 823 patients were observed for at least four years. The effect of removing the handful of remaining patients from the comparison can therefore be nothing but trivial. For this reason the basis for Dr. Schor's statement that the baseline differences for those seen four or more years are "unbelievably large" is hard to understand. He refers to data taken from preliminary, unpublished documents. I am familiar with only one unpublished document in which such data were given, the progress report to the National Institutes of Health dated July 30, 1966, at which time cholesterol level measurements were available on only 27 patients receiving placebo and 22 patients receiving tolbutamide who had been observed four full years. Although irrelevant to the interpretation of final results, they show that six of the 27 patients receiving placebo and seven of the 22 patients given tolbutamide had serum cholesterol levels of 300 mg/100 ml or over, which scarcely corresponds to the alleged difference of almost threefold. Thus, the basis for Schor's comments on early differences remains unknown. Table 2 is a recompilation of the data in Table 6 of the Report (p 799), confined to the subgroup of patients with a date of entry prior to Oct 7, 1965, who were thus available for at least four years of follow-up at the date of the Report. It is the only set of data relevant to Dr. Schor's statement and does not support it in any way.

He also states that "all of the excess deaths in the tolbutamide group occur in only three clinics." I shall consider this point later, but immediately relevant is his further statement that "it would appear to any reasonable statistician that for some reason or other the randomization procedure broke down in these three clinics over some period of time but possibly not over the whole study." This point too can be factually investigated. Table 3 gives the equivalent of Table 6 of the report (p 799) for the three clinics with the largest

Table 2.—Percent of Patients With Date of Entry Prior to Oct 7, 1965 With Selected Baseline Characteristics*

	Placebo	Tolbutamide	Insulin Standard	Insulin Variable	All
Demographic Characteristics					
Age ≥ 55 yr	42.3 (201)	48.2 (197)	46.2 (208)	46.2 (199)	45.8 (805)
Male	30.8 (201)	31.5 (197)	26.9 (208)	22.1 (199)	27.8 (805)
Nonwhite	49.8 (201)	47.2 (197)	51.0 (208)	41.2 (199)	47.3 (805)
Baseline Cardiovascular Risk Factors					
Hypertension	37.1 (197)	29.5 (193)	31.1 (206)	28.7 (195)	31.6 (791)
History of digitalis use	4.5 (198)	7.8 (192)	5.8 (206)	5.1 (195)	5.8 (791)
History of angina pectoris	4.5 (198)	7.2 (195)	7.7 (207)	3.6 (195)	5.8 (795)
Significant ECG abnormality†	3.1 (196)	4.1 (195)	5.3 (208)	4.1 (196)	4.2 (795)
Cholesterol level ≥ 300 mg/100 ml	8.8 (194)	15.5 (193)	16.6 (205)	13.8 (196)	13.7 (788)
One or more cardiovascular risk factors listed above	47.0 (183)	47.8 (186)	50.5 (200)	42.3 (189)	47.0 (758)
Other Selected Baseline Characteristics					
Fasting blood glucose level ≥ 110 mg/100 ml	68.8 (199)	71.6 (197)	64.3 (207)	68.7 (198)	67.9 (801)
Relative body weight ≥ 1.25	53.2 (201)	58.9 (197)	57.7 (208)	54.3 (199)	56.0 (805)
Visual acuity (either eye ≤ 20/200)	4.3 (186)	5.3 (187)	6.2 (195)	5.9 (186)	5.4 (754)
Serum creatinine level ≥ 1.5 mg/100 ml	2.6 (190)	2.6 (193)	2.0 (204)	2.1 (195)	2.3 (782)
Arterial calcification‡	14.1 (199)	19.3 (192)	17.4 (201)	16.3 (190)	16.8 (782)

*Denominators given in parentheses.
†Major or minor Q-waves (codes 1-1-1 through 1-2-7), S-T depression (code 4-1), T-wave inversion (code 5-1), complete heart block (code 6-1), left bundle branch block (code 7-1), or ventricular tachycardia (code 8-2).
‡Evidence of calcification noted by two independent readings of soft tissue x-ray films of right lower limb.

Table 3.—Percent of Patients in Boston, Minneapolis, and Williamson Clinics With Selected Baseline Characteristics*

	Placebo	Tolbutamide	Insulin Standard	Insulin Variable	All
Demographic Characteristics					
Age ≥ 55 yr	46.3 (60)	55.1 (63)	49.2 (63)	50.8 (63)	53.4 (249)
Male	45.7 (60)	36.5 (63)	25.4 (63)	23.8 (63)	32.9 (249)
Nonwhite	8.3 (60)	6.3 (63)	12.7 (63)	4.8 (63)	8.0 (249)
Baseline Cardiovascular Risk Factors					
Hypertension	32.8 (58)	29.0 (62)	32.3 (62)	33.9 (59)	32.0 (241)
History of digitalis use	3.3 (60)	7.9 (63)	6.5 (62)	4.9 (61)	5.7 (246)
History of angina pectoris	8.3 (60)	9.5 (63)	11.1 (63)	1.6 (61)	7.7 (247)
Significant ECG abnormality†	1.7 (60)	6.3 (63)	6.3 (63)	6.5 (62)	5.2 (248)
Cholesterol level ≥ 300 mg/100 ml	13.6 (59)	9.7 (62)	14.5 (62)	16.1 (62)	13.5 (245)
One or more cardiovascular risk factors listed above	47.4 (57)	42.6 (61)	55.0 (60)	49.1 (57)	48.5 (235)
Other Selected Baseline Characteristics					
Fasting blood glucose level ≥ 110 mg/100 ml	54.2 (59)	73.0 (63)	71.4 (63)	73.0 (63)	68.1 (248)
Relative body weight ≥ 1.25	43.3 (60)	50.8 (63)	57.1 (63)	44.4 (63)	49.0 (249)
Visual acuity (either eye ≤ 20/200)	10.2 (59)	8.1 (62)	6.5 (62)	1.6 (61)	6.6 (244)
Serum creatinine level ≥ 1.5 mg/100 ml	1.7 (60)	0.0 (63)	0.0 (63)	0.0 (62)	0.4 (248)
Arterial calcification‡	20.3 (59)	24.2 (62)	16.9 (59)	17.5 (57)	19.8 (237)

*Denominators given in parentheses.
†Major or minor Q-waves (codes 1-1-1 through 1-2-7), S-T depression (code 4-1), T-wave inversion (code 5-1), complete heart block (code 6-1), left bundle branch block (code 7-1), or ventricular tachycardia (code 8-2).
‡Evidence of arterial calcification noted in both of two independent readings of the same set of soft-tissue x-ray films of the right lower limb.

reported tolbutamide-placebo excess in mortality from *all causes* (Boston, Minneapolis, and Williamson clinics) and Table 4 gives the corresponding P values. The numbers involved are smaller than for all 12 clinics and the variation by treatment group consequently larger. There is nevertheless nothing in these tables that would support the hypothesis that randomization broke down in these clinics or to implicate excess baseline inequalities in these clinics as an explanation of their larger reported excess mortality. If anything the proportion of patients with cardiovascular risk factors tends to be low in the tolbutamide group—the case of sex and of cholesterol level ≥ 300 mg/100 ml being perhaps the most striking.

Although the point has not been raised, it is also pertinent to inquir about the distribution of baseline ris factors in the four clinics accountin for the bulk of tolbutamide-placel excess in mortality from *cardiovascular disease* (Birmingham, Boston, Cincinnati, and Minneapolis Table 5 gives the baseline distribution for these clinics and Table 6 th corresponding P values. It will be observed that the patients given pl

Table 4.— *P* Values Based on Chi-Square* Tests of a Distribution of Baseline Characteristics in Patients in Boston, Minneapolis, and Williamson Clinics

	Tolbutamide-Placebo	Insulin Standard-Placebo	Insulin Variable-Placebo	All Four Treatments
Demographic Characteristics				
Age	0.66	0.92	0.79	0.20
Sex	0.25	0.01	0.008	0.02
Race	0.67	0.43	0.42	0.39
Cardiovascular Risk Factors				
Hypertension	0.66	0.95	0.90	0.95
History of digitalis use	0.27	0.43	0.66	0.72
History of angina pectoris	0.82	0.60	0.09	0.21
Significant ECG abnormality†	0.19	0.19	0.18	0.56
Cholesterol level ≥ 300 mg/100 ml	0.50	0.88	0.69	0.75
One or more cardiovascular risk factors listed above	0.60	0.41	0.85	0.89
Other Baseline Characteristics				
Fasting blood glucose level ≥ 110 mg/100 ml	0.03	0.05	0.03	0.07
Relative body weight ≥ 1.25	0.41	0.13	0.90	0.38
Visual acuity (either eye ≤ 20/200)	0.69	0.46	0.05	0.27
Serum creatinine level ≥ 1.5 mg/100 ml	0.30	0.30	0.31	0.37
Arterial calcification‡	0.61	0.64	0.70	0.74

*Uncorrected for continuity. When continuity corrections are used, *P* values are increased.
†Major or minor Q-waves (codes 1-1-1 through 1-2-7), S-T depression (code 4-1), T-wave inversion (code 5-1), complete heart block (code 6-1), left bundle branch block (code 7-1), or ventricular tachycardia (code 8-2).
‡Evidence of arterial calcification noted in both of two independent readings of the same set of soft-tissue x-ray films in the right lower limb.

Table 5.—Percent of Patients in Birmingham, Boston, Cincinnati, and Minneapolis Clinics With Selected Baseline Characteristics*

	Placebo	Tolbutamide	Insulin Standard	Insulin Variable	All
Demographic Characteristics					
Age ≥ 55 yr	50.7 (73)	55.4 (74)	48.1 (77)	43.1 (72)	49.3 (296)
Male	41.1 (73)	28.4 (74)	20.8 (77)	25.0 (72)	28.7 (296)
Nonwhite	38.4 (73)	33.8 (74)	48.1 (77)	31.9 (72)	38.2 (296)
Baseline Cardiovascular Risk Factors					
Hypertension	31.4 (70)	32.4 (71)	30.3 (76)	33.8 (68)	31.9 (285)
History of digitalis use	9.7 (72)	12.5 (72)	6.6 (76)	5.6 (71)	8.6 (291)
History of angina pectoris	9.6 (73)	9.6 (73)	7.9 (76)	1.4 (70)	7.2 (292)
Significant ECG abnormality†	4.2 (72)	4.1 (73)	5.5 (77)	8.5 (71)	5.8 (293)
Cholesterol level ≥ 300 mg/100 ml	7.0 (71)	20.5 (73)	21.3 (75)	21.1 (71)	17.6 (290)
One or more cardiovascular risk factors listed above	45.5 (66)	55.1 (69)	54.2 (72)	52.2 (67)	51.8 (274)
Other Selected Baseline Characteristics					
Fasting blood glucose level ≥ 110 mg/100 ml	66.2 (71)	82.4 (74)	67.1 (76)	69.0 (71)	71.2 (292)
Relative body weight ≥ 1.25	47.9 (73)	59.5 (74)	62.3 (77)	55.6 (72)	56.4 (296)
Visual acuity (either eye ≤ 20/200)	6.1 (66)	5.6 (71)	7.0 (71)	6.2 (65)	6.2 (273)
Serum creatinine level ≥ 1.5 mg/100 ml	1.6 (63)	0.0 (70)	1.4 (73)	0.0 (69)	0.7 (275)
Arterial calcification‡	19.2 (73)	26.0 (73)	21.6 (74)	20.9 (67)	22.0 (287)

*Denominators given in parentheses.
†Major or minor Q-waves (codes 1-1-1 through 1-2-7), S-T depression (code 4-1), T-wave inversion (code 5-1), complete heart block (code 6-1), left bundle branch block (code 7-1), or ventricular tachycardia (code 8-2).
‡Evidence of arterial calcification noted in both of two independent readings of the same set of soft-tissue x-ray films of the right lower limb.

cebo have an apparently significant excess of males ($P=0.01$) and a nearly significant deficiency of hypercholesterolemic patients ($P=0.06$) and that none of the other differences among the four treatment groups is significant. The sum of the 14 chi-squares with 3 degrees of freedom is 46.9, only slightly above the value of 42 expected on the hypothesis of no baseline differences among the treatment groups in these four clinics. This value of chi-square corresponds to $P > 0.25$ and also does not support the hypothesis of a breakdown in randomization even in the subgroup of clinics accounting for the majority of cardiovascular deaths.

Effect of Random Baseline Inequalities

To what extent did the random, nonsignificant baseline inequalities contribute to the significant excess

Table 6.— *P* Values Based on Chi-Square Tests of a Distribution of Baseline Characteristics in Patients in Birmingham, Boston, Cincinnati, and Minneapolis Clinics

	Tolbutamide-Placebo	Insulin Standard-Placebo	Insulin Variable-Placebo	All Four Treatments
Demographic Characteristics				
Age	0.57	0.75	0.36	0.51
Sex	0.11	0.007	0.04	0.04
Race	0.56	0.23	0.42	0.17
Cardiovascular Risk Factors				
Hypertension	0.90	0.88	0.76	0.97
History of digitalis use	0.60	0.48	0.36	0.44
History of angina pectoris	1.00	0.71	0.03	0.19
Significant ECG abnormality†	0.99	0.53	0.29	0.63
One or more cardiovascular risk factors listed above	0.02	0.01	0.02	0.06
	0.26	0.31	0.43	0.68
Other Baseline Characteristics				
Fasting blood glucose level ≥ 110 mg/100 ml	0.02	0.91	0.72	0.10
Relative body weight ≥ 1.25	0.16	0.08	0.36	0.32
Visual acuity (either eye ≤ 20/200)	0.92	0.82	0.98	0.99
Serum creatinine level ≥ 1.5 mg/100 ml	0.29	0.92	0.29	0.56
Arterial calcification‡	0.32	0.71	0.80	0.78

*Uncorrected for continuity. When continuity corrections are used, *P* values are increased.
†Major or minor Q-waves (codes 1-1-1 through 1-2-2), S-T depression (code 4-1), T-wave inversion (code 5-1), complete heart block (code 6-1), left bundle branch block (code 7-1), or ventricular tachycardia (code 8-2).
‡Evidence of arterial calcification noted in both of two independent readings of the same set of soft-tissue x-ray films of the right lower limb.

cardiovascular mortality among the patients given tolbutamide? Since there is no uniquely best way of answering this question, I shall first consider two ways favored by Dr. Schor. One involves comparison of cardiovascular mortality in the different treatment groups only for those with no cardiovascular risk factors present at baseline. Table 8 of the Report (pp 802 to 803) shows that for this subgroup 2.0% of the group given placebo but 9.0% of the group given tolbutamide died of cardiovascular causes. Dr. Schor writes, "This method I find difficult to argue with if done properly. However, not all of the important risk factors were taken into account. Neither obesity nor arterial calcification was classified as a cardiovascular risk factor by the investigators." The data have been recompiled, redefining as risk factors the five original ones used in Table 8 of the Report and the two additional ones suggested. Of those with none of these seven risk factors, the percent dying of cardiovascular causes becomes 2.6 for placebo, 10.0 for tolbutamide, 0.0 for insulin standard, and 0.0 for insulin variable. Although the numbers on which these percents are based are small (39, 40, 36, and 34 patients, respectively), and no one would draw sweeping conclusions from them, the fact remains that the excess cardiovascular mortality among patients assigned to tolbutamide persists. Although the importance of arterial calcification as a risk factor, and its somewhat higher prevalence in the tolbutamide group, have been pointed out, the fact remains that the tolbutamide-placebo excess in cardiovascular mortality occurred both for those with and without arterial calcification (Table 8, p 802). The excess cannot therefore be explained by baseline inequalities in this risk factor. (The percent cardiovascular mortality in Table 8 for those with arterial calcification is in error, as noted by Dr. Feinstein. Corrections for this and other minor inaccuracies appeared in the April issue of *Diabetes*.).

A second method suggested is the use of cardiac score, since "the difference in the number of people having cardiac scores greater than 0 in the two groups could alone have caused the entire difference in mortality." The percent mortality from cardiovascular causes among those with a cardiac score of zero at baseline, about 80% of the total patient population, are 1.8 for placebo, 8.3 for tolbutamide, 4.1 for insulin standard, and 3.1 for insulin variable. Neither of the two suggested methods of correction therefore lends any support to the idea that baseline inequalities explain the excess cardiovascular mortality with tolbutamide.

An alternative analysis is to tabulate the cardiovascular mortality in the different treatment groups by the number of risk factors present, as defined for Table 1. This is shown in Table 7, with the bottom line giving mortality standardized by the direct method, using as a standard population the distribution of all 756 patients by number of risk factors present. Again the excess cardiovascular mortality with tolbutamide persists. The entire effect of adjusting on this basis for the random baseline differences with regard to tolbutamide shown in Table 1 is seen to be, by comparison of the last two lines of the present table, a reduction in the tolbutamide-placebo excess mortality from 7.3 to 5.8 percentage points, so that the excess is still more than 100%.

Table 7.—Percent Mortality From Cardiovascular Causes by Number of Baseline Risk Factors and Treatment Groups*

No. of Baseline Risk Factors	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
0	3.6	8.0	0.0	0.0
1	0.0	6.0	0.0	1.3
2	1.7	8.6	5.0	5.3
3	19.2	17.6	14.7	10.0
4	20.0	35.3	12.5	25.0
5	0.0	25.0	37.5	75.0
6	...	0.0	100.0	100.0
All	4.9	12.2	6.7	6.4
Standardized	5.1	10.9	5.9	7.2

*Denominators are given in Table 1.

Table 8.—Patients at Baseline by Treatment Group and Probability* of Cardiovascular Death, P

Probability of Cardiovascular Death	Treatments				
	Placebo	Tolbutamide	Insulin Standard	Insulin Variable	All
< 0.0065	49	36	40	39	164
0.0065-0.0140	37	38	46	44	165
0.0141-0.0295	44	35	40	46	165
0.0297-0.0665	34	54	39	38	165
≥ 0.0673	41	41	45	37	164
Total	205	204	210	204	823

*Probability for each patient computed from equation 3 of Report (p 822) with $x_1 = x_2 = x_3 = 0$.

We have finally the method of correction described in pages 823 to 824 of the Report—the use of a multiple logistic function. This is a method that has generated considerable interest in recent years in cardiovascular epidemiology.[1] The method of cumulation used in this analysis assigns a weight to each possible risk factor and lets the cumulated measure of risk depend upon the product of the weight and the intensity of the risk factor, summed over all risk factors. Weights are derived from the data themselves in such a way as to give the best possible prediction of mortality in the patients studied. This method may be preferable to that summarized in Table 7 of this communication, since it does not treat all risk factors as of equal importance. A cumulative score can be calculated for each patient. It gives the probability of a cardiovascular death, P, for a patient receiving placebo with that particular patient's combination of risk factors. A frequency distribution of the values of P has been calculated for all 823 patients, and the quintiles of the distribution determined. Table 8 shows for each treatment group the number of cases in each quintile at baseline. The distribution of patients among the quintiles is much the same among the treatment groups, thus again failing to uncover any indication of a breakdown in the randomization or bad luck in the draw.

Table 9 shows for each treatment group the observed and expected number of cardiovascular deaths by quintile, the expected number being the sum of the calculated risks, P, for all subjects in the quintile. The increase shown in observed cardiovascular mortality from lowest to highest quintile is very marked for all treatment groups, thus indicating that at least for these 823 patients the score provides a sensitive index of cardiovascular risk. It will be noted that the observed excess mortality with tolbutamide persists and that the theoretical expectations given by the function agree reasonably well with the actual deaths, thus indicating that in the aggregate the adjustment is about right. It is not clear how Dr. Schor's speculation concerning over and under adjustment for individual variables can be checked or what the relevance of this might be for the evaluation of the study.

The major point of Table 9 is that the expected number of deaths among tolbutamide patients is 10.7, indicating that their baseline characteristics could in fact account for excess over placebo of 0.7 of a cardiovascular death, as compared with an actual excess of 16. No one would claim that this is the uniquely best way of cumulating cardiovascular risk factors; it does seem to be the most useful method that has emerged so far after 15 years of cardiovascular epidemiology.

The coefficients for diastolic blood pressure and relative body weight in the regression function are negative, contrary to expectation. The sign for systolic blood pressure is positive, however, and the combined effect of the two, reflecting the overall effect of blood pressure, is dominated by the systolic. The negative sign for diastolic may reflect the difficulty of disentangling by purely statistical methods the effects of such highly correlated variables. The negative sign for weight is perhaps a result of the difficulty of detecting the very small effect of weight on risk. It is known in Framingham, for example,[2] that weight is less important than cholesterol level, blood pressure, amount smoked, and ECG status in men, and has not been demonstrated to have any effect on risk in women, who, after all, constitute almost three fourths of the UGDP study patients. Statisticians and epidemiologists with whom I have discussed Dr. Schor's statement about the effect of the unknown 18th variable know of

41-746 O - 75 - 3

Table 9.—Observed and Expected* Cardiovascular Deaths, by Treatment Group and Probability† of Cardiovascular Death, *P*

Probability of Cardiovascular Death	Treatments									
	Placebo		Tolbutamide		Insulin Standard		Insulin Variable		All	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
≤ 0.0065	0	0.2	0	0.1	0	0.2	0	0.2	0	0.6
0.0065-0.0140	1	0.4	1	0.4	0	0.5	0	0.5	2	1.7
0.0141-0.0295	0	0.9	4	0.7	1	0.9	1	1.0	6	3.6
0.0297-0.0665	2	1.6	5	2.4	3	1.7	1	1.6	15	7.3
≥ 0.0673	7	6.9	12	7.1	9	7.3	10	7.9	38	29.3
Total	10	10.0	26	10.7	13	10.6	12	11.2	61	42.5

*Obtained by summing *P*, the probability of a cardiovascular death, for all patients in each interval by treatment group.
†Probability for each patient computed from equation 3 of Report (p 822) with $x_1 = x_2 = x_3 = 0$.

no basis in theory or empirical evidence for his claim.

The published data as summarized in Table 10 appear to show that the excess cardiovascular mortality for patients given tolbutamide is largest among white women, a point raised in discussions of the UGDP results at meetings of the Society for Epidemiological Research and of the Epidemiology Council of the American Heart Association. How can one be sure that the white women were balanced with respect to baseline risk factors? Application of the same risk function to each of the 16 cells of Table 10 yields the expected number of deaths shown in Table 11. The results show an excess of observed over expected number of deaths for white women in the group given tolbutamide (and for each of the other three race-sex groups), indicating that again baseline inequalities cannot explain the observed excess.

None of the previous adjustments takes account of risk factors that were not measured, such as smoking history. But, one can calculate the consequences of assuming that the one in 50,000 chance mentioned previously in this communication did occur and that the prevalence of cigarette smokers among tolbutamide patients exceeded that among placebo patients by 20 percentage points. Since cigarette smokers have an excess mortality from cardiovascular disease of about 80%, even such an improbable difference could account for an excess cardiovascular

Table 10.—Percent Mortality From Cardiovascular Causes by Treatment, Race, and Sex*

	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
White men	10.6 (47)	17.5 (40)	15.1 (33)	6.5 (31)
Nonwhite men	12.5 (16)	17.4 (23)	0.0 (24)	0.0 (15)
White women	1.8 (56)	16.2 (68)	4.3 (70)	7.8 (90)
Nonwhite women	2.3 (86)	5.5 (73)	6.0 (83)	4.4 (68)
Total	4.9 (205)	12.7 (204)	6.2 (210)	5.9 (204)

*The number in parenthesis is the total number of patients on which the percent is based.

mortality of only 16%, or 1.6 additional cardiovascular deaths in the group given tolbutamide.

In summary, one must say that there is no reason to believe that the randomization broke down, and no evidence that the random baseline differences that did occur contributed in any important way to the adverse effects with tolbutamide. If one questions that UGDP on this matter of baseline differences, one must question the entire concept of the randomized therapeutic trial.

Special Clinic Effects

The second major line of statistical criticism is that the excess mortality was in fact confined to a small number of clinics, and because of this could not be generalized to all clinics in the study or to what might be expected in general medical practice. This criticism appears to give too little weight, if not to overlook entirely, the small number of patients enrolled in any one treatment group in any one clinic, ie, 22 or less. This corresponds to an average of two or fewer cardiovascular deaths per treatment group per clinic. Because of these small numbers, considerable variation in the apparent treatment effect from clinic to clinic is inevitable. It is precisely because of the small number of patients available for treatment in any one clinic that a collaborative multiclinic trial is a necessity and why the cumulation of results over many clinics is required to obtain interpretable results.

To what extent is the observed variation in treatment effect from clinic to clinic actually explained by small numbers? The investigators address themselves to this question on pages 825 to 826 of the Report and conclude on the basis of a standard statistical analysis for this type of question "that the observed distribution of drug-placebo differences in mortality among clinics for a given drug-placebo comparison was not at variance with the hypothesis that the effect of that drug on mortality was the same in all twelve clinics," ie, that small numbers alone could explain the entire observed variation.

A similar attention to the problem posed by small numbers would allay

Table 11.—Observed and Expected* Cardiovascular Deaths by Treatment, Race, and Sex

	Placebo		Tolbutamide		Insulin Standard		Insulin Variable	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
White men	5	4.0	7	3.0	5	3.2	2	2.5
Nonwhite men	2	0.9	4	1.7	0	1.0	6	1.0
White women	1	2.6	11	4.0	3	4.3	7	5.4
Nonwhite women	2	2.5	4	1.9	5	2.1	3	2.3
Total	10	10.0	26	10.7	13	10.6	17	11.2

*Obtained by summing P_i, probability of a cardiovascular death, for all patients in a race-sex-treatment group. P_i computed from equation 3 of report (p 822) with $x_1 = x_2 = x_3 = 0$, $x_4 = 1$ for men and 2 for women and $x_5 = 1$ for whites and 2 for nonwhites.

Table 12.—Patients With High Adherence* Dead, by Cause of Death

	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
No. at risk of death	143	151	121	92
Cardiovascular causes				
1. Myocardial infarction	0	10	3	0
2. Sudden death	2	2	4	2
3. Other heart disease	1	5	1	0
4. Extracardiac vascular disease	2	5	3	2
Total	5	22	11	4
Noncardiovascular causes				
5. Cancer	4	3	3	0
6. Cause other than 1-7	2	2	1	1
7. Unknown cause	0	0	1	0
Total	6	5	5	1
All causes	11	27	16	5
Percent dead				
Cardiovascular causes	3.5	14.6	9.1	4.3
All causes	7.7	17.9	13.2	5.4

*Patient received ALL of prescribed study medication ≥ 75% of all follow-up periods. A patient was regarded as receiving ALL of his assigned study medication in a given follow-up period if he received medication in the dosage specified for at least 75 days of that three-month period, and if he was not receiving any hypoglycemic agent other than his assigned study medication during that period.

some of Dr. Schor's other concerns. The variation from treatment group to treatment group in percentage of deaths from all causes from cardiovascular diseases in patients who were hospitalized or autopsied can be calculated from the data in Tables B-1 through B-4 (pp 827-830) to be no greater than would be expected with numbers this small ($P > 0.10$ for all four comparisons) and cannot be considered as evidence that the definition of a cardiovascular death varied by treatment group. (It is hard to see how it could, since the causes of death were coded by a committee that was blind with respect to individual treatment assignment.) Similarly his question, "Is it not just as logical for the investigators to have claimed that tolbutamide lowers the risk of cancer death," appears to overlook the statistical significance of the cardiovascular excess and the nonsignificance of the cancer deficiency ($P > 0.15$). (Feinstein also discusses this point. When the analysis is restricted only to patients who died, seven out of 21 deaths in the group given placebo and two out of 30 deaths in the group given tolbutamide were of cancer, leading to an exact P value in Fisher's test of 0.023. Because the denominator is the total number of deaths and not the number of patients at risk, this significant deficiency among patients given tolbutamide can be significant either of a reduced mortality from cancer or an increased mortality from other causes. For this reason such a comparison has long been recognized as logically incapable of yielding interpretable information about differences between populations in specific causes of death. Feinstein also analyzes on the basis of all patients at risk, which does yield interpretable results, obtaining a chi-square, uncorrected for continuity of 2.81, and a P value between 0.05 and 0.10, but does not give the exact P value yielded by Fisher's test. The use of uncorrected chi-squares has been challenged.[7] When corrected chi-squares are used, the P value is, as stated above, > 0.15, as is the exact P value given by Fisher's test.)

Dropouts and Nonadherence

For this complex problem, the UGDP has followed the generally accepted practice of comparing the mortality experience of the originally randomized groups, and of not eliminating dropouts or nonadherers from the analysis. This practice is conservative in that it dilutes whatever treatment effects, beneficial or adverse, are present. The comparison based on patients with good adherence only is generally considered unsafe because the original comparability provided by randomization may be impaired. But since the point has

been raised, attention is called to Tables 12 and 13 which show the separate mortality experience of those with high adherence. It will be noted that the excess mortality of the patients given tolbutamide not only persists, it is intensified. The 2½-fold elevation in cardiovascular mortality shown in Table 1 of the original Report (p 790) is now fourfold (Table 12) and the nearly threefold elevation after eight years is almost sixfold (Table 13). The analysis suggested by Dr. Schor, difficult to interpret though it is, does nothing to weaken the UGDP finding with respect to tolbutamide and in fact tends to strengthen it.

Table 13.—Cumulative Mortality per 100 Population at Risk, and Standard Errors by Year of Follow-Up

Year of Follow-Up	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
All causes				
		Cumulative Mortality		
1	0.0	0.0	1.6	2.0
2	2.1	3.3	2.5	4.1
3	4.9	6.6	4.1	4.1
4	6.3	9.9	7.4	4.1
5	7.0	13.0	9.2	4.1
6	8.0	14.8	10.2	4.1
7	8.0	20.3	15.2	4.1
8	8.0	22.9	15.2	9.8
		Standard Errors		
1	0.7	0.7	0.8	0.9
2	1.4	1.4	1.5	1.7
3	1.8	1.8	2.0	2.2
4	2.1	2.1	2.3	2.6
5	2.3	2.3	2.6	2.8
6	2.5	2.4	2.7	3.1
7	2.9	3.0	3.2	3.6
8	3.4	3.4	3.7	4.3
Cardiovascular causes				
		Cumulative Mortality		
1	0.0	0.0	0.8	2.0
2	0.7	2.0	1.7	4.1
3	3.6	4.7	2.5	4.1
4	3.5	7.4	5.0	4.1
5	3.5	11.2	6.8	4.1
6	3.5	13.1	6.8	4.1
7	3.5	17.3	10.3	4.1
8	3.5	20.0	10.3	4.1
		Standard Errors		
1	0.6	0.6	0.7	0.8
2	1.2	1.1	1.2	1.4
3	1.6	1.5	1.7	1.9
4	1.8	1.8	2.0	2.2
5	2.1	2.0	2.3	2.5
6	2.2	2.1	2.4	2.7
7	2.6	2.6	2.8	3.1
8	2.9	2.9	3.2	3.6

*Patients with high adherence. See Table 12 for definition.

Characteristics of Study Population

Using rule of thumb ratios, Dr. Schor concludes "the group given placebo is remarkably healthy from a cardiovascular viewpoint and all comparisons of cardiovascular deaths in other treatment groups with this group given placebo would be biased." It is true that the cardiovascular mortality of the group given placebo is below that expected for a normal population of the same age-sex-race composition, 6.6% for the placebo group after eight years of follow-up (Table A-2, p 819), compared with an expectation of 10.4% after eight years of follow up for the US life table population (Table A-3, p 819). This is also true for both groups given insulin, however. The comparable percents, 7.7 for the insulin standard and 8.1 for the insulin variable, are both below that for the US life table population. That individuals willing to participate in an intervention study have a lower than expected mortality is not unique to the UGDP. The National Diet-Heart Feasibility Study, for example, reported an annual incidence of new coronary heart disease of 0.5% compared with an expectation for a comparable general population of 1.0%.[1] Dr. Schor's earlier remark that the comparability of the treatment groups is much more important than their representativeness seems relevant here. While fully agreeing that the study population is not representative of the general adult-onset diabetic population, it does not follow that comparisons with the groups given placebo are therefore biased. Indeed the earlier analysis of the extent and effect of baseline inequalities argues the contrary.

Clinical Questions

Several clinical questions, on which my comments will be brief, have also been raised. The UGDP finding of an adverse mortality experience with a fixed dose of tolbutamide can be contrasted to the mortality experience of patients treated with a fixed dose of insulin. Despite the virtually identical baseline characteristics and blood glucose level control, they differed widely in mortality experience. Whether the finding for tolbutamide can be extended to other populations and dosages is in the present state of knowledge a matter of individual judgment. It is not wholly irrelevant to note, however, that the variable dosage of insulin, while much more effective than the fixed dosage in lowering blood glucose level, was no more effective in lowering mortality.

Dr. Schor also disagrees with the judgment of the clinic physicians in screening patients for life-endangering conditions so as to obtain patients with a minimum life expectancy of five years. He points to

Table 14.—Percent of Patients With Specified ECG Findings at Baseline

ECG Abnormality*	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
Q/QS patterns				
Major Q/QS (code 1-1-X)	0.0	2.0	0.0	1.5
Moderate Q/QS (code 1-2-X)	1.0	0.0	2.0	1.0
Minor Q/QS (code 1-3-X)	3.0	3.0	1.0	1.5
Any of above	4.0	5.0	3.8	4.0
T-wave abnormalities				
Major T (code 5-1)	1.0	1.0	0.5	1.0
Moderate T (code 5-2)	6.0	9.0	8.1	9.4
Minor T (code 5-3)	3.5	8.0	4.8	8.4
Borderline T (code 5-4)	4.0	2.0	2.9	5.0
Any of above	14.6	19.9	15.5	23.8
S-T Depressions				
Major S-T (code 4-1)	1.5	2.5	1.4	1.0
Moderate S-T (code 4-2)	1.0	1.0	1.4	2.5
Minor S-T (code 4-3)	2.5	6.0	4.3	7.4
Junction S-T (code 4-4)	0.0	0.5	0.0	0.5
Any of above	5.0	10.0	7.2	11.4
S-T elevation				
S-T elevation (code 9-2)	0.5	1.5	1.9	2.0
A-V conduction defects				
Complete A-V block (code 6-1)	0.0	0.0	0.0	0.0
Partial A-V block (code 6-2)	0.0	0.0	0.0	0.0
Prolonged P-R (code 6-3)	2.0	2.5	4.3	3.0
WPW syndromes (code 6-4)	0.0	0.0	0.0	0.0
Short P-R intervals (code 6-5)	2.5	1.0	0.5	3.0
Any of above	4.5	3.5	4.8	5.0
Ventricular conduction defects				
Complete LBBB (code 7-1)	0.0	0.0	1.0	0.5
Complete RBBB (code 7-2)	1.5	0.5	1.4	1.5
Incomplete RBBB (code 7-3)	1.5	1.0	1.0	0.5
Intraventricular conduction defects (code 7-4)	1.0	0.0	1.4	0.0
R-R patterns (code 7-5)	1.5	0.5	0.1	0.5
Incomplete LBBB (code 7-6)	0.0	0.0	0.0	0.0
All ventricular conduction defects	5.5	2.0	5.7	3.0
Arrhythmias				
PBS ≥ 1 to beats (code 8-1)	3.5	3.0	2.4	2.0
Ventricular tachycardia (code 8-2)	0.0	0.0	0.0	0.0
Atrial fibrillation (code 8-3)	1.5	0.5	0.5	0.0
Supraventricular tachycardia (code 8-4)	0.0	0.0	0.0	0.0
Ventricular rhythm (code 8-5)	0.0	0.0	0.0	0.0
Nodal rhythm (code 8-6)	0.0	0.0	0.0	0.0
Sinus tachycardia (code 8-7)	0.5	1.5	2.9	3.0
Sinus bradycardia (code 8-8)	0.5	1.5	1.0	1.0
Any of above	6.0	6.5	6.7	6.0
No. of patients*	199	201	206	202

*WPW signifies Wolff-Parkinson-White; LBBB, left bundle branch block; RBBB, right bundle branch block; PBS, premature beats.

two elderly patients who died during the course of the study as examples of patients who, in his judgment, should not have been admitted. Whatever the merits of this position, these patients did conform to protocol, requirements for eligibility, and both did, in fact, survive more than five years.

Changing Definitions

Dr. Schor points to the fact that definitions of baseline characteristics varied throughout the study and comments "one can only conjecture as to why this was done." One must first emphasize that none of these definitions had any effect on findings.

The final judgment concerning the principal cause of death for each deceased patient was made by a special review team without knowledge of the treatment group to which the patient had been assigned. This team consisted of the chairman of the UGDP Mortality Committee and a consultant pathologist. Their decision regarding principal cause of death was based on information in the detailed death report prepared at the study clinic (Report, p 790).

Definitions of baseline risk factors, including ECG abnormality, were used by the Coordinating Center solely to classify study patients in the hope of further elucidating the cause for the elevated eight-year cardiovascular mortality for those receiving tolbutamide. But this finding depended only on the brute facts of the UGDP experience and was beyond the ability of anyone to influence by manipulating definitions of baseline characteristics.

In a study as complex as the UGDP

the analysis elucidating the major findings must be to some extent experimental and it is hardly to be expected that the first analysis used would invariably be the most perspicuous. In an initial attempt to combine risk factors, the investigators used a cardiac score, which was subsequently criticized for not including hypertension, and was therefore not used in the final report. As shown above, its use does not alter study findings in any way. Similarly, the ECG classification was being used while Dr. Henry Blackburn, one of the study's advisors on ECGs, was revising the Minnesota code and relating the various abnormalities to prognosis in Framingham. The abnormalities believed to be most closely related to prognosis were the ones used in the final *Diabetes* report. That no "damaging" baseline inequalities were concealed is indicated by Table 14, which displays the classification of patients by complete ECG findings at baseline. Using this table one is free to define abnormality as one wishes.

The point at which one dichotomizes a continuous distribution is necessarily arbitrary. The dichotomizations given in the Mortality paper appear, if anything, to have been selected as to emphasize, and not to minimize, the apparent imbalances among treatment groups. For example, tables in Appendix I of the Baseline paper, which give complete distributions, show much less apparent imbalance in age (p 777) or serum cholesterol (p 782) than does Table 6 of the Mortality paper (p 799).

There have been charges published in the medical tabloid press of deliberate attempts by the investigators to mislead by concealment of data, as for example year-by-year cumulative mortality. These very data are given in parts A and B of Fig 2 of the Report (pp 794-795). The data on dropouts and nonadherence which might be considered to strengthen the evidence against tolbutamide, were available when the Mortality Report was being prepared, but were not included because their interpretation is not unequivocal. Similarly, Feinstein remarks that the "UGDP statisticians were commendably fair in their decision to keep" 68 patients who did not fulfill criteria for admission in the analysis since "the subsequent comparison of the death rates for the smaller denominators in the PLBO and TOLB groups would have magnified the existing differences in these rates." One can only conclude from all this that if the investigators did wish to conceal "damaging" data they are hopelessly inept at the task.

Table 15.—Cumulative Cardiovascular Mortality per 100 Population at Risk and Standard Errors by Year of Follow-Up

Year of Follow-Up	Placebo	Tolbutamide	Insulin Standard	Insulin Variable
Patients in Clinics Not Administering Phenformin				
		Cumulative Mortality		
1	3.6	9.0	3.0	2.2
2	5.7	8.2	6.0	4.8
3	5.6	4.6	0.7	5.0
4	4.4	7.5	2.9	5.0
5	4.4	11.5	5.7	5.0
6	6.1	14.1	4.5	5.9
7	9.3	18.6	5.9	7.0
8	7.3	20.6	6.9	9.0
		Standard Errors		
1	0.6	0.6	0.6	0.6
2	1.1	1.1	1.1	1.1
3	1.5	1.5	1.5	1.6
4	1.6	1.9	1.6	1.8
5	2.0	2.1	2.0	2.0
6	2.3	2.5	2.2	2.3
7	2.6	2.7	2.6	2.6
8	2.9	3.0	2.9	2.9
Patients in Clinics Administering Phenformin				
		Cumulative Mortality		
1	3.0	0.0	1.4	1.5
2	3.0	3.0	2.9	1.5
3	3.0	1.5	2.9	3.1
4	3.0	3.0	2.5	3.1
5	1.9	3.0	4.6	3.1
6	3.4	5.7	4.6	3.1
7	3.4	5.9	13.4	3.1
8	...	5.9	13.4	...
		Standard Errors		
1	1.1	1.0	1.0	1.1
2	1.3	1.3	1.3	1.2
3	1.7	1.6	1.6	1.7
4	1.9	1.8	1.8	1.9
5	2.2	2.0	2.3	2.2
6	2.5	2.4	2.4	2.5
7	4.4	4.1	4.1	4.3
8	...	4.8	8.8	...

*Includes first 32 patients in Boston.

The Decision to Discontinue

Dr. Schor has deplored the failure to continue treatment until the excess mortality with tolbutamide had been demonstrated over a longer period of time in the last five clinics included in the study. Had this been done, Dr. Schor writes, "very few people would argue with the conclusions as currently stated." Before the decision to discontinue was made the investigators did in fact compare, on a life-table basis, the mortality in those clinics in which phenformin was used (last five plus the last 54 patients in Boston) with those in which it was not. The results, summarized in Table 15, show that mortality of patients given tolbutamide exceeded that of patients given placebo from the third year on in both groups of clinics. The number of patients who had been ob-

served for eight full years was considerably smaller in the clinics in which phenformin was used and the standard errors are larger. The demonstration would no doubt have been strengthened if the investigators had waited longer. But in investigations involving human subjects, one is not obliged to continue treatment until a conclusive demonstration of the mortal effect of a supposedly therapeutic agent has been achieved, particularly when there is no possibility of demonstrating a positive effect on mortality by continuing.

Comment

An accepted basis for treating adult-onset diabetes is that lowering blood glucose to normal levels will reduce the incidence of cardiovascular complications. It is only natural that any study which casts doubt on this widely held postulate, and furthermore suggests that one of the hypoglycemic agents may be harmful, should be subject to intense scrutiny for possible sources of error. The preceding sections, particularly the new results presented there, indicate that none of the possible errors suggested so far do in fact account for the UGDP findings. Although further investigation, particularly if undertaken in a nonadversary framework, may still be useful, it seems likely that a point of diminishing returns may not be far off, and that continued analysis of the UGDP, in the hope of finding errors which alter the conclusions, will become increasingly unrewarding.

It does not follow, of course, that the UGDP results, particularly the tolbutamide effect, must be accepted as conclusively established. No single clinical trial no matter how well-designed and executed can do that, particularly if a drug is discontinued, as it must be, as soon as (if not before) adverse findings became significant. Continuation of UGDP patients on the two insulin treatments and on a regimen of phenformin will provide further evidence on the benefits to be expected, if any, from the prophylactic use of hypoglycemic agents in asymptomatic patients. Further evidence on the possible adverse effects of tolbutamide will be harder to obtain, although epidemiological study of experience with it is still a possibility. Clinical trials abroad, now being considered, may also contribute additional knowledge. At this point the UGDP findings cannot easily be dismissed, and if they are eventually rejected, it will only be because a large body of scientifically defensible evidence against them, not now available, has been accumulated, and not because of continued exegesis of current results. Until, and unless, that point is reached, the findings must stand, in my opinion, as the best available on the effects of hypoglycemic agents on cardiovascular complications.

This investigation was supported by Public Health Service research grant GM-15004.

Genell Knatterud, PhD, Coordinating Center, UGDP, undertook the additional statistical analyses presented here. Jacob E. Bearman, PhD; Byron W. Brown, Jr., PhD; Curtis L. Meinert, PhD; Thaddeus E. Prout, MD; and Theodore B. Schwartz, MD, contributed by reviewing the manuscript.

Nonproprietary and Trade Names of Drug

Phenformin hydrochloride—*DBI, DBI-TF*

References

1. The University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. *Diabetes* 19(suppl 2):747-830, 1970.
2. Feinstein AR: Clinical biostatistics. VIII. An analytic appraisal of the University Group Diabetes Program (UGDP) study. *Clin Pharmacol* 12:167-191, 1971.
3. Hill AB: *Principles of Medical Statistics*, ed 7. London, Oxford University Press, 1961.
4. Epstein F: Multiple risk factors and the prediction of coronary heart disease. *Bull NY Acad Med* 44:1014-1035, 1968.
5. Truett J, Cornfield J, Kannal W: A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chron Dis* 20:511-524, 1967.
6. *Smoking and Health*, Report of the Advisory Committee to the Surgeon General of the Public Health Service, Public Health Service publication 1103, US Dept of Health, Education, and Welfare, 1964.
7. Dorn HF: Some applications of biometry in the collection and evaluation of medical data. *J Chron Dis* 1:658-664, 1955.
8. Mantel N, Greenhouse SW: What is the continuity correction. *Amer Stat* 22:27-29, 1968.
9. National diet-heart study. *Circulation* 37(suppl I):1-113, 1968.

Dr. CORNFIELD. The original charges were characterized by some inaccuracies, some grossly inflated language and a strong tendency to score debater's points, rather than to examine the full record. The rejoinders to my JAMA article have been in the same vein. A number of years ago the journal "Cancer" published an article proposing ground rules for statistical criticism in connection with the question of smoking and lung cancer. It remarked that "much that nowadays passes as statistical criticism is superficial and sophomoric in character and serves to obscure scientific discussion rather than to clarify it."

The statistical critics of the UGDP have not always acted in a way that suggests appreciation of the rules proposed in that article.

One of the initial charges against the UGDP had been that it changed its original electrocardiographic criteria in order to conceal damaging baseline inequalities. The JAMA article provided a full and detailed tabulation of the new criteria, which have gained worldwide acceptance and explained that they had been developed by Dr. Henry Blackburn, of the University of Minnesota, an internationally recognized authority in the field, and consultant to the World Health Organization, on the basis of information not available at the initiation of the UGDP. But in an article devoted to rebutting the JAMA article the old charge is simply repeated as follows, "Why the later definition, if valid, was not used originally, or why the revised definition is medically any better than the original one for patients with diabetes is not explained." Similarly, this rebuttal points to what it terms an excess of 62 risk factors among the tolbutamide patients and wonders whether this excess might not readily account for the excess of 16 cardiovascular deaths, in the tolbutamide group. But this very question was addressed in table 7 of the JAMA article in which it was shown that even allowing for this apparent excess of risk factors, the mortality with tolbutamide was more than twice that of the placebo patients.

Several methods of appraising the effects of the alleged baseline inequalities were considered in the JAMA article, and the rebuttal pounces on the fact that they do not lead to identical results. Two different procedures employed are said to lead to a "striking" and later a "major inconsistency." On examination it appears that one definition of low risk groups leads to one death in the low-risk placebo group and two in the low-risk tolbutamide group, and the other definition of low risk to one death in the placebo group and one in the tolbutamide group. To call this shift of one patient a "major inconsistency" particularly when both definitions show that the excess cardiovascular mortality with tolbutamide persists under either definition of risk seems to support the characterization of much statistical criticism in the Cancer article quoted above.

One is inevitably reminded of Max Planck's remark, "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it." Planck was talking about theoretical physics, a field uncomplicated by the existence of special economic interests, and the difficulty of informing a patient that his past treatment was ineffective and probably harmful. The UGDP, like any other scientific enterprise is not without weaknesses, but no one has yet suggested how any of them contributed to

its findings of an excess cardiovascular mortality with the oral hypoglycemic agents. Unless a large body of scientifically defensible evidence that contradicts them is eventually accumulated, they must stand as the best available on the subject and a landmark in the design and execution of experiments for the scientific evaluation of medical therapy.

Mr. GORDON. Thank you very much, Dr. Cornfield.

I have a couple of questions.

Do you know of any study approaching the thoroughness and sophistication of the UGDP study that supports the use of these oral antihypoglycemic agents?

Dr. CORNFIELD. I know of no studies of these agents which approaches the UGDP in thoroughness and in care and the length of time covered.

Mr. GORDON. Then, why weren't these other studies attacked?

Dr. CORNFIELD. Well, as a statistician, I always feel more comfortable when I stick to the facts, and I have to conjecture a little bit to answer your question, but I would conjecture that in part the other studies were not attacked, not because of their high scientific quality, but because their results were acceptable.

I would also note that several of the critics of UGDP have been paid consultants to the manufacturers, and as far as I know, there have been no consultants paid to undertake a thorough analysis of any of the other studies.

Mr. GORDON. The Committee on the Care of the Diabetic proposed that in the labeling of these drugs the following paragraph should be included and I quote: "Other long-term prospective studies in which patients were randomly assigned oral hypoglycemic treatment, tolbutamide or a placebo, showed no difference in the incidence of cardiovascular mortality or complications." And it listed Keen, et al., Feldman, et al., Garcia et al., Hooper and others, Moss and so forth.

I wonder from your statistical and critical review of the UGDP data, how would you compare those results with the other studies mentioned in the quoted statement?

Dr. CORNFIELD. Well, I have already commented on Dr. Moss' data which as far as I know has not been published in any scientific journal. Any journal that used any system of critical refereeing would be unlikely to publish such a poorly designed study.

Now, I am not familiar with all the studies mentioned, but I do know something about some of them. The study by Paasikivi had 52 diabetic patients in it.

Mr. GORDON. How many did the UGDP study have?

Dr. CORNFIELD. Somewhat over a thousand.

I do not have the number of patients included in Dr. Keen's study, but I think it was well below the numbers involved in the UGDP.

If one examines the question of baseline inequalities which were alleged to exist in the UGDP, similar baseline inequalities of perhaps even greater magnitude can be found in Dr. Keen's study. His study did not go on for the 10 years the UGDP went on, and my recollection is that his definition of cardiovascular events is a good deal fuzzier and less clearcut than the UGDP definition.

I think your earlier question as to why these studies have not come in for some of the criticism is a very good one.

Mr. GORDON. Dr. Chalmers, would you like to add to his statement concerning these other studies? I think you mentioned it in your prepared statement.

Dr. CHALMERS. Yes, there are two other studies which have been mentioned, and I am looking them up now because I cannot pronounce the names. Well, Feldman I can pronounce. His study is in much milder diabetics, and there have been only two or three deaths so far, so it is impossible to detect whether there would have been a difference. The other one, the one that I cannot pronounce, had to do with phenformin, and that was a study in patients who they hoped would demonstrate a protective effect with regard to arteriosclerosis, and the patients did not have diabetes.

Dr. CORNFIELD. Mr. Gordon, Dr. Prout has shown me the relevant table from Dr. Keen's article, and his study includes 248 patients.

Mr. GORDON. How about the studies which are consistent with the results of the UGDP study?

Dr. CORNFIELD. Well, there are several such, and again, I am not an expert in this field, but there is a case-control study by Dr. Hadden and his colleagues that was published in 1972, I believe, which showed an elevated risk of myocardial infarction.

Mr. GORDON. That is similar to the UGDP study, is it not?

Dr. CORNFIELD. Yes. The results are almost the same as the UGDP results, showed a 2½ fold effect in cardiovascular mortality.

Mr. GORDON. The Hadden study if I recall correctly had over 600 people in it, isn't that right?

Dr. CORNFIELD. Yes, it did. But it was not a landmark study and if one is critical of some of the studies that disagree with the UGDP, I think one should also note that in terms of design, although the results are consistent with the UGDP, its designs were not nearly as precise or as careful as the UGDP.

Mr. GORDON. Neither were the others.

Dr. CORNFIELD. Yes. So I would return to my concluding statement, which is that of all the studies available, on purely objective, scientific grounds, the one whose design commands the most confidence in execution is the UGDP, and other studies, some of which are not in agreement and some of which are in agreement—to me, if one wants the best available information on the effect that the oral agents, one has to turn to the UGDP.

Mr. GORDON. Thank you very much, Dr. Cornfield.

Dr. PROUT. Mr. Gordon, I think a point should be made, returning to our 1 percent per year mortality figure, that it becomes quite clear that studies of smaller numbers of people for shorter periods of time cannot show that kind of difference, yet we consider that an appreciable difference, and one that anyone who undertakes to use the oral agents, and any patient who takes the oral agents should be cognizant of.

Dr. CHALMERS. There is also one other kind of data that should be emphasized here, and that is vital statistics that are gathered by various health agencies throughout the world. They have shown an age specific mortality rate for diabetics which, beginning with the discovery of insulin in 1921, has come steadily down and continued to come down in younger diabetics at the age at which people ordinarily take insulin. However, in older diabetics, from 40 to 50 and 50 to 60

and over, the death rates were coming down steadily until the late 1950's and early 1960's when the decline stopped, and now the rate has started to rise again. So there is a distinct upward bump in those curves for people in the age group who would ordinarily be taking the oral hypoglycemic agents. This suggests, although again we would have to have the kind of corroborative evidence we have in the UGDP, that there may in fact be a large increase in premature deaths in the world as a result of the widespread use of these drugs.

Mr. GORDON. How about the Japanese statistics? There seems to be some correlation between the use of the oral antidiabetics with the death rate in Japan.

Can you comment on that?

Dr. MILLER. Yes. The Japanese, in fact were the first group to note this surprising and disturbing change in mortality rate in the older age group, beginning at a time coincident with the introduction of oral agents. In 1971 or 1972 a paper was published in Japan describing not only the increase in older age groups in their country but also similar data from other countries.

Subsequently a paper from Israel confirmed these observations. More recently a paper from Japan did a rather interesting study in which they looked at the causes of death in diabetics dying in a large group of hospitals in Japan, and particularly looked at the percentage of patients dying from cardiovascular causes.

The interesting way in which the study was designed should be commented upon. The ratios of the mortality in diabetics treated with diet alone, with oral agents or with insulin to the mortality of age and sex-matched non-diabetic Japanese were calculated.

Cardiovascular involvement in diabetes is related to the duration of the disease. In those patients who died with diabetes of less than 5 years' duration, there was no increase in the percentage of patients dying of cardiovascular disease in the two groups treated with diet or insulin. But in the tolbutamide group there was an excess of mortality, and in those groups where the diabetes was in existence for more than 5 years, this same relative increase in mortality from cardiovascular disease was noted only in the tolbutamide group.

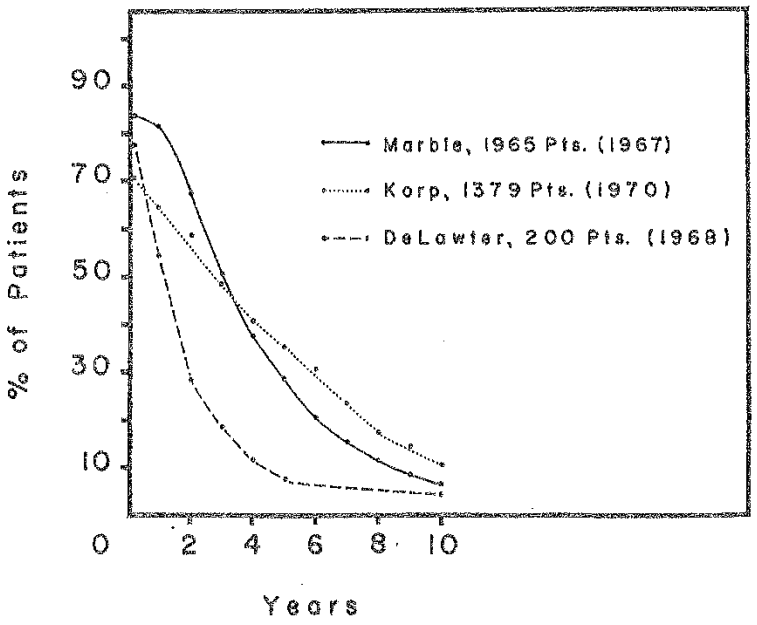
I would like to submit for the record the graphs of the mortality data from Germany which illustrate what Dr. Chalmers has pointed out, that there was an increase in mortality only in the older age groups, coincident with the introduction of the oral agents.

Senator NELSON. It will be received for the record.

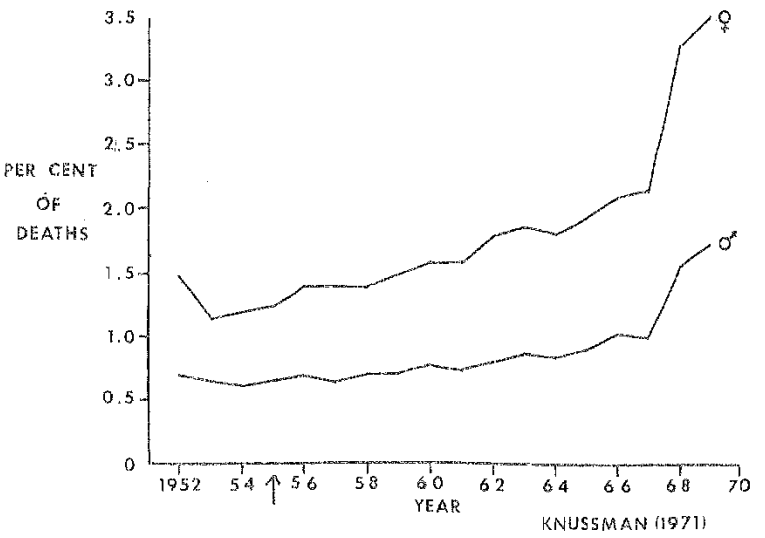
[The information referred to follows:]

Document #24

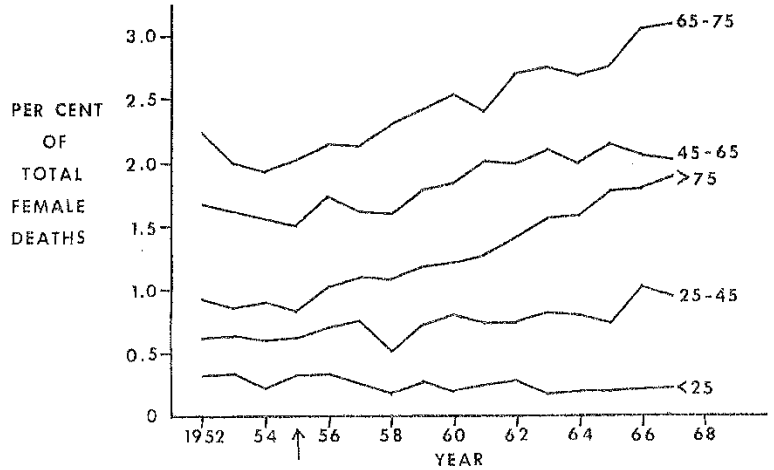
PERCENTAGE OF PATIENTS ADEQUATELY "CONTROLLED" ON S-U OVER A TEN YEAR PERIOD



PER CENT OF DEATHS DUE TO DIABETES IN WEST GERMANY, 1952-1969



PER CENT OF TOTAL FEMALE DEATHS DUE TO DIABETES
ACCORDING TO SPECIFIC AGE GROUPS
IN WEST GERMANY, 1952 - 1967



KNUSSMAN (1971)

Senator NELSON. Our final witness is Dr. Thomas Chalmers, president, the Mount Sinai Medical Center, dean, Mount Sinai School of Medicine of the City University of New York.

Dr. Chalmers, we are very pleased to have you here today. Your statement will be printed in the record, and you may present it however you desire.[1]

STATEMENT OF THOMAS C. CHALMERS, M.D., PRESIDENT, THE MOUNT SINAI MEDICAL CENTER, DEAN, MOUNT SINAI SCHOOL OF MEDICINE OF THE CITY UNIVERSITY OF NEW YORK

Dr. CHALMERS. I shall not read this statement. It is the longest one, and a lot of what is contained in the statement has already been covered by the previous discussion.

Senator NELSON. It will be printed in full in the record. If you want to elaborate on any points that were not covered or those that were, just go ahead.

Dr. CHALMERS. I will summarize the important points and then make some recommendations which I think are important at this time. I am neither a diabetologist nor a statistician as are the four people on my left, and I appear before you as a sometime specialist in clinical trials. I have been interested since 1950 in the conduct and in the design of clinical trials, and the interactions of ethics and clinical care in such trials. I have conducted a number of trials myself, and have recently been serving on a number of policy advisory boards and have

[1] See prepared statement, page 10987.