PREDICTION OF CANCER AND CORONARY HEART DISEASE AS A FUNCTION OF METHOD OF QUESTIONNAIRE ADMINISTRATION [1]

R. GROSSARTH-MATICEK

Institute for Preventive Medicine
European Centre for Peace and Development
University for Peace, established by the United Nations

H. J. EYSENCK AND P. BARRETT

Institute of Psychiatry
University of London

Summary.—We have tested the hypothesis in this study that different methods of administering a questionnaire produce differential approximations to truthful admission of undesirable personality traits and behaviours. Four different methods of administration produced different levels of *trust* and *understanding*, using the current prediction among healthy subjects of death by cancer or coronary heart disease 13 years later as the criterion. There were significant differences in the accuracy of the predictions, depending crucially on the method of administration of the questionnaires. Best predictions were achieved for subjects when both trust and understanding had been increased by interviewers' suitable participation; worst results were achieved for subjects when no special effort was made to increase either. Intermediate results were found for procedures which increased *either* trust *or* understanding. It is argued that the success or failure of studies investigating the influence of personality and stress on diseases like cancer and coronary heart disease may depend crucially on the adopted method of interrogation.

There is a large literature on the relation between personality and stress on the one hand and cancer and coronary heart disease on the other (Eysenck, 1991). Le Shan (1959) has traced the belief in such a connection back to the ancient Greeks and has shown that it was accepted as axiomatic by European physicians until the beginning of this century when the scientific advances made possible by Pasteur and others led to its abandonment. It is only comparatively recently that concepts like "Type A" and its link with coronary heart disease have been taken seriously (Booth-Kewley & Friedman, 1987). Le Shan (1959, 1961; Le Shan & Reznikoff, 1960) and Grossarth-Maticek (1979, 1986, 1989) have performed the same service for cancer (Temoshok & Dreher, 1992). Le Shan (1977) claimed to have found a "pattern of loss of hope" in between 70 and 80% of his cancer patients and in only 10% or so of the control group, and Grossarth-Maticek has claimed a somewhat less extreme success rate in his extensive prospective studies (Eysenck, 1991).

Reviewers of the literature (e.g., Fox, 1978, 1981a, 1981b, 1983) are often skeptical of such claims and quote other studies, sometimes showing no differences between groups with cancer and coronary heart disease and various types of controls. Such skepticism may be mistaken for a variety of reasons. Unsuccessful attempts to test personality-related theories about the origin and progression of cancer or coronary heart disease may have used questionnaires or other instruments irrelevant to the issue; this is not an un-usual happening. As one of us has pointed out (Eysenck, 1990a), one and the same study (Schmale & Iker, 1971) may provide both positive and nega-tive evidence with respect to an hypothetical link between cancer and per-sonality. Strong positive evidence was found for a procedure wherein inter-viewers probed for evidence of "feelings of hopelessness," which was the hypothetical variable posited by Schmale and Iker to be related to cancer fol-lowing Le Shan's theory (1959). Negative evidence was found when the MMPI or a projective technique was used. But these latter techniques are quite irrelevant to the theory and would not be expected to produce evidence in favour of a theory they were not designed to test. The majority of nega-tive findings seem to be based on such doubtful and possibly irrelevant in-struments. When Grossarth-Maticek's (1986, 1989) claims were indepen-dently tested using his own scales, largely positive results have been reported (e.g., Quander-Blaznik, 1991; Schmitz, 1992; Ranchor, Sanderman, & Bou-ma, 1992; Amelang & Schmidt-Rathjens, 1992; Sandín, Chorot, Navas, & Santed, 1992).

However, this may not be the whole story. The method of administra-tion of the personality test may also be of crucial importance. Thus, pre-diction of coronary heart disease on the basis of *interviews* has been consis-tently more accurate than predictions based on *questionnaires* (Price, 1982) although the reasons for this are not too clear. It may be that, as often sug-gested, the tone of voice used in the interview may be more informative than what words are actually being spoken or it may be that the act of inter-viewing which entails some form of social interaction may encourage the respondent to give more truthful or revealing answers (Eysenck, 1990b). The failure then of many researchers to find strong evidence of personality-dis-ease interaction may reflect their failure to acquire the trust of the proband or their failure to clarify questions that arise (Eysenck, 1990b). Note that in the Schmale and Iker study (1971) the interview was successful in predicting coronary heart disease, while the MMPI administration gave no positive evi-dence.

It is well known that even simple questions may not be answered truth-fully when there is no attempt to establish an atmosphere of trust. As Lee (1988) has shown in an analysis of over 100 studies, a simple question like "Do you smoke cigarettes?" may be answered wrongly in a sufficient num-

ber of cases to invalidate the claims and relevance of passive smoking to lung cancer. It seems possible that the much more intimate revelations called for in the case of disease-related personal material (feelings of hopelessness, of anger, of irritability, of failure) will be more likely to be made under certain conditions than under other less trust-establishing and information-providing conditions. This is the issue investigated in the present study.

The most serious problem for any such study is, of course, that of obtaining a valid criterion. In the present study we have used as our criterion the accuracy of the prediction made of death and cause of death for the probands following any of four very different administrations of the personality questionnaire used to make these predictions. This would appear to guarantee an objective criterion which can be related to the conditions of manipulation of administration.

METHOD

Questionnaire

The measuring instrument used in this study has been discussed and given in full in a previous publication (Grossarth-Maticek & Eysenck, 1990; also, cf. Eysenck, 1991). It consists of 71 questions in the original German form, one of which was dropped from the English translation to equate number of questions for the six types measured. These six types are Type 1 = cancer-prone, Type 2 = coronary-heart-disease-prone, Type 3 = hysterical personality, Type 4 = healthy autonomous personality, Type 5 = rational/antiemotional, and Type 6 = psychopathic. Our main interest here is the demonstration (in agreement with several large-scale prospective studies using a four-type scale but omitting Types 5 and 6) that Type 1 predicts cancer, Type 2 coronary heart disease, and Type 4 the absence of both. We are less concerned with the other types which are largely irrelevant to cancer and coronary heart disease, except perhaps Type 5 which is also related to disease. It was suggested that the general formula, i.e., Types $(1 + 2 + 5) - (3 + 4 + 6)$, would best predict all-round mortality (Eysenck, 1991).

Procedure and Sample

The studies reported by us have been prospective, i.e., healthy probands have been tested and followed for periods varying from 10 to over 20 years; questionnaire scores were then correlated with mortality and cause of death as shown on death certificates. The method of administration was always by a trained interviewer, who spent one hour explaining the purpose of the questionnaire administration, assured the subject of confidentiality, answered questions both general and specific, and tried to win the trust and confidence of the subject by showing interest and concern, being friendly and courteous, giving information, and being responsive to questions. Independent replications have usually simply administered the questionnaire without

such participation by interviewers to groups of well and ill subjects in an at-
tempt to relate types to cause of death or diagnosed illness (e.g., Schmitz,
1992; Ranchor, *et al.*, 1992; Amelang & Schmidt-Rathjens, 1992; Sandín, *et
al.*, 1992). In spite of these differences, replications have been supportive of
our main conclusions. Note, for example, Schmitz (1992) reported in a study
of 100 married women with an average age of about 30 years that, of 6 can-
cer patients who could be assigned to a type, all were of Type 1, while of 6
patients with cardiac infarction 4 were of Type 2, and 1 of Type 5.

The sample of the present experiment consisted of a fairly random
sample of 1958 males whose ages were between 55 and 57 years, all living in
the small German university town of Heidelberg. Selection was based on
random choice from population records, with age and sex qualifications add-
ed. Individuals were invited to take part in a scientific study of psychosocial
factors in health and illness. Of those approached, 237 declined the invita-
tion, leaving a sample of 1721. The study began in 1975, and the follow-up
was carried out 13 years later in 1988. All participants were asked to com-
plete the Short Interpersonal Reaction Inventory (Grossarth-Maticek & Ey-
senck, 1990), but this administration was carried out differentially using ran-
domized assignment to four groups.

Group A.—In this group (*n* = 338) the questions were read aloud by the
interviewer, and explanations were given after each question as to the precise
meaning of the question following any queries by the subject. This group is
designated the *explanation* group, the hypothesis being that such an explana-
tion would help many (particularly the less well educated) to understand the
meaning of the questions better.

Group B.—In this group (*n* = 348) we tried to manipulate the variable of
trust. Participants were invited to talk with the interviewer for 45 to 60 min-
utes, discussing in the first part positive and negative events of their lives
and their typical reactions to these situations. Following this part of the in-
terview, participants were asked in the second part if they trusted the
purpose of the questionnaire administration or if they still had some ques-
tions to ask. The interviewer did his best to answer such questions as were
raised, and only began administration of the questionnaire when the subject
stated that the interviewer as well as the purpose and also design of the
questionnaire administration were trusted.

Group C.—For Group C (*n* = 348), the *explanatory* method for Group A
was combined with the *trust-evoking* method for Group B so that, following
the discussion devoted to gaining the trust of the subject, the interviewer
would continue with the explanation of all the items in the questionnaire.
This group then enjoyed both the explanatory and trust-producing para-
digms.

Group D.—This group (*n* = 687) constituted the control group, receiving

neither the explanatory nor the trust-producing interaction with the interviewer. Instead, they were given the questionnaire and asked to fill it in without prior discussion or explanation of the meaning of the questions.

Our hypothesis was that the correlation between disease and personality would emerge most clearly for Group C and least clearly for Group D, with estimates for the other two groups being intermediate. No prediction was made with respect to possible differences between Groups A and B.

For the analysis we used proportions as this was the most direct way of expressing the differences between groups and types. The test made for the difference between proportions is that given by Armitage and Berry (1991), based upon the approximation of the binomial distribution by the normal distribution, yielding a standardised normal deviate as the test value. Given a 2 × 2 chi-squared table, analysis yields a chi squared which is the square of the equivalent standardised normal deviate; the two statistical tests are actually mathematically equivalent in this special case. Thus, in the light of this fact, it is immaterial whether 2 × 2 chi squared or a normal approximation to the binomial test is used.

Occasionally, Yates' correction was applied for reasons which are based upon the derivation of the test as outlined above. Armitage and Berry do not provide a correction for unpaired, proportionate test data. Since the test made is, however, mathematically equivalent to a 2 × 2 chi squared and since it is conventional to correct the chi squared if any observed frequency is below 5, we decided, in tables where any observed frequency was less than 5, to use chi squared with Yates' corrected calculations as a conservative estimate of the significance of the differences.

RESULTS

Criterion for Prediction

We must next discuss the nature of the criterion used to carry out this comparison. From our theory and previous studies, we predicted a number of relationships; if these were not found to hold over-all, *i.e.*, for the four groups combined, the whole experiment would, of course, founder since the basic premise was the *differential* support of the groups' scores for the predictions. It was, of course, possible that some but not all predictions would be borne out. Below is given a list of the specific predictions used, together with a statement of the *over-all* validity of the prediction for this sample. (Tests for all six predictions were carried out by chi squared.)

Prediction 1.—Type 1 is associated with cancer significantly better than all other types. Of Type 1 scorers, 35 out of 326 died of cancer which is 10.74%; of all other types, 45 out of 1395 died of cancer which is 3.23%. The difference between the two proportions is 7.51%, which is significant at $p < .00001$ one-tail, with 95% confidence limits between 4.02% and 11.00%. This difference is significant and in line with prediction.

Prediction 2.—Type 2 is associated with coronary heart disease significantly better than all other types. Of Type 2, 39 out of 304 died of coronary heart disease which is 12.83%; of all other types, 46 out of 1417 died of coronary heart disease which is 3.25%. The difference between the two proportions is 9 58% ($p < .00001$ one-tail test with 95% confidence limits between 5.71% and 13.45%). The difference is significant then and in line with prediction.

Prediction 3.—Type 1 scores are more closely associated with cancer than are Type 2 scores. Of 326 Type 1 subjects, 35 died of cancer which is 10.74%. Of 304 Type 2 subjects, 25 died of cancer which is 7.57%. The difference between the two proportions is 3.17% which is not significant ($p = .09$) on a one-tail test with 95% confidence limits between −1.32% and 7.77%. Although the difference is not significant, it is in the predicted direction.

Prediction 4.—Type 2 scores are more closely associated with coronary heart disease than Type 1 scores. Of 304 subjects classified as Type 2, 39 died of coronary problems which is 12.83%. Of 326 subjects classed as Type 1, 24 died of coronary heart disease which is 7.36% The difference between the two proportions is 5.47% which is significant at $p = .00112$, one-tail. The 95% confidence limits lie between 0.76% and 10.18%. Note the difference is statistically significant.

Prediction 5.—Types 3, 4, and 6 scorers are less likely to suffer cancer or coronary heart disease than are Types 1, 2, and 5. Out of 911 subjects of Types 1 + 2 + 5, 142 have died of cancer or coronary heart disease (15.59%), while of 810 subjects of Types 3 + 4 + 6, only 23 have died with such a diagnosis (2.84%). The difference between the two proportions is 12.75%, significant at $p < .00001$ on a one-tail test. The 95% confidence limits lie between 10.13% and 15.37%.

Prediction 6.—Types 1, 2, and 5 scorers are more likely than Type 3, 4, and 6 scorers to have died of diseases other than cancer or coronary heart disease, as compared with the surviving control group. Out of 911 subjects scoring as Types 1 + 2 + 5, 161 died of other causes (17.67%). Out of 810 subjects classified as Types 3 + 4 + 6, 106 died of other causes (13.09%). The difference between the two proportions is 4.59% and significant ($p = .0044$, one-tail test) with 95% confidence limits between 1.19% and 7.98%.

Comparisons Between Administrations

These results, all significant (with one exception) and in line with theory and previous work suggest that we can use them to formulate our criteria against which to test the hypothesis that these trends would be best established in Group C, worst in Group D, and show intermediate success in Groups A and B. Table 1 gives the numbers and percentages for the four groups; further analyses, like the preceding ones, are based on these tables.

TABLE 1

Type	Deaths							Surviving	
	Cancer		Coronary		Other		Total	n	%
	n	%	n	%	n	%	n		
Group A (explanation only)									
1	6	40.00	4	25.00	11	19.64	21	47	18.73
2	4	26.67	7	43.75	5	8.93	16	42	16.73
3	2	13.33	1	6.25	8	14.29	11	31	12.35
4	1	6.67	1	6.25	12	21.43	14	82	32.67
5	2	13.33	3	18.75	14	25.00	19	38	15.14
6	0	0.00	0	0.00	6	10.71	6	11	4.38
Totals	15		16		56		87	251	
Group B (trust only)									
1	7	46.67	3	17.65	13	19.40	23	43	17.27
2	4	26.67	9	52.94	14	33.40	27	38	15.26
3	2	13.33	1	5.88	7	10.45	10	32	12.85
4	1	6.67	2	11.76	10	14.93	13	90	36.14
5	1	6.67	1	5.88	16	23.88	18	38	15.26
6	0	0.00	1	5.88	7	10.45	8	8	3.21
Totals	15		17		67		99	249	
Group C (explanation and trust)									
1	12	70.59	5	29.41	17	36.96	34	51	19.03
2	3	17.65	11	64.71	15	32.61	29	57	21.27
3	0	0.00	0	0.00	1	2.17	1	50	18.66
4	0	0.00	0	0.00	3	6.52	3	69	25.75
5	2	11.76	1	5.88	7	15.22	10	33	12.31
6	0	0.00	0	0.00	3	6.52	3	8	2.99
Totals	17		17		46		80	268	

(continued on next page)

Note.—Figures in percent columns are the row numbers expressed as a percentage of the column total.

TABLE 1 (Cont'd)

PERSONALITY TYPE AND MORTALITY: FOUR GROUPS GIVEN DIFFERENT ADMINISTRATIONS OF THE QUESTIONNAIRE

Type	Deaths							Surviving	
	Cancer		Coronary		Other		Total	n	%
	n	%	n	%	n	%	n		
Group D (neither trust nor explanation)									
1	10	30.30	12	34.29	18	18.37	40	67	12.86
2	12	36.36	12	34.29	19	19.39	43	52	9.98
3	1	3.03	1	2.86	4	4.08	6	65	12.48
4	4	12.12	5	14.29	40	40.82	49	220	42.23
5	6	18.18	5	14.29	12	12.24	23	102	19.58
6	0	0.00	0	0.00	5	5.10	5	15	2.88
Total	33		35		98		166	521	

Note.—Figures in percent columns are the row numbers expressed as a percentage of the column total.

The method of analysis may be illustrated as follows. The illustration is based on the argument that, if dissimulation is an important factor in interrogation whether oral or by questionnaire, then Types 1 (cancer group) and 2 (coronary heart disease group) should be most affected, Type 4 (healthiest) least, because items of Scale 4 have much higher social desirability than items of Scales 1 and 2, which invite admissions of failure in interpersonal relations and inability to deal with stress. The ratio of numbers of Type 4/ Types 1 + 2 individuals should therefore be highest in Group C, lowest in Group D, and intermediate in Groups A and B. It may be noted that Schmitz (1992) found a significant *positive* correlation with neuroticism for Types 1 and 2 and a significant *negative* one for Type 4. The evidence suggests that neuroticism and social desirability are negatively correlated (Eysenck & Eysenck, 1985).

We used the methods suggested by Armitage (1955) and Armitage and Berry (1991) to test first by chi squared that there is no significant difference between Groups A and B and then the proposition that there is a difference such that Group D > Group AB > Group C, there being a linear trend within this dataset.

The ratio of Type 4 individuals to Types 1 + 2, within each group, A to D, is as follows; for Group A the ratio is 96/126 or 0.76, for Group B the ratio is 103/131 or 0.79, for Group C the ratio is 72/171 or 0.42, and for Group D the ratio is 269/202 or 1.33. Initially, we tested for the equivalence of intermediate Groups A and B: difference in proportions was − 0.0244 which is not significant ($p = 0.64$, two-tail). For Groups A and B the differences are statistically equivalent, so we combined these groups for comparative purposes: the Group AB ratio was 199/257 or 0.77.

Now, the hypothesis to be tested is that Group D > Group AB > Group C. Using a test for ordered proportions, defined by Armitage (1955) and based upon partitioning a chi-squared variable into an over-all chi squared due to a linear trend to test the hypothesis above (D > AB > C), we have an over-all chi squared of 50.62 ($p < .000001$, $df = 2$). The χ_1^2 due to linear trend was 50.61 ($p < .000001$); χ_1^2 testing for departure from linear trend was 50.62 − 50.61 = 0.008 ($p = .929$, which is a nonsignificant value). There is thus a definite linear trend within this dataset.

We next test Prediction 1. Out of 326 Type 1 individuals 35 (10.74%) died of cancer and 45 out of 1395 (3.23%) other types died of cancer:

Type	Group D	Group A	Group B	Group C
1	10	6	7	12
2, 3, 4, 5, 6	23	9	8	5

Testing of the equivalence of Groups A and B, using the same test for the difference between proportions (which yields a χ_1^2 2 × 2 table-equivalent probability since a chi squared with 1 $df = z^2$), we have a difference between proportions for Groups A and B of 6.67%, which is nonsignificant ($p = .71$, two-tail).

Collapsing across Groups A and B

Type	Group D	Group AB	Group C
1	10	6.5	12
2, 3, 4, 5, 6	23	8.5	5

and testing for Group D < Group AB < Group C, we find the over-all χ_2^2 was 7.40 ($p = .025$). χ_1^2 due to linear trend was 7.17 ($p = .007$). The χ_2^2 testing for departure from linear trend was 0.23 ($p = .63$), which is nonsignificant. There is thus a definite linear trend within this dataset.

Testing for Prediction 2, we find that 46 out of 1417 (3.25%) other types died of coronary heart disease and 38 out of 304 (12.83%) Type 2 individuals died of coronary heart disease.

Type	Group D	Group A	Group B	Group C
2	12	7	9	11
1, 3, 4, 5, 6	23	9	8	6

Testing of the equivalence of Groups A and B by using the same test for the difference between proportions (which yields a chi squared 2 × 2 table-equivalent probability since a chi squared with 1 $df = z^2$), we have a difference between proportions for Groups A and B of 9.91%, nonsignificant at $p = .60$, two-tail.

Averaging across Groups A and B

Type	Group D	Group AB	Group C
2	12	8	11
1, 3, 4, 5, 6	23	8.5	6

and testing for Group D < Group AB < Group C, we find that over-all χ_2^2 was 4.37 ($p = .11$). χ_1^2 due to linear trend was 4.36 ($p = .04$). χ_1^2 testing for departure from linear trend was .0050, $p = .94$, which is nonsignificant. There is thus a definite trend within this dataset.

We next test Prediction 3, noting that 35 out of 326 (10.74%) Type 1 individuals died of cancer, and 23 out of 304 (7.57%) Type 2 individuals died of cancer. Testing of equivalence of Groups A and B, using the same

Type	Group D	Group A	Group B	Group C
1	10	6	7	12
2	12	4	4	3

test for the difference between proportions (which yields a χ_1^2 2×2 table-equivalence probability since a χ_1^2 with 1 $df = z^2$), we have a difference between proportions between Groups A and B of 3.6%, which is not significant at $p = .78$, two-tail (using Yates' correction for continuity).

Collapsing across Groups A and B

Type	Group D	Group AB	Group C
1	10	6.5	12
2	12	4	3

and testing for Group D < Group AB < Group C, the over-all χ_2^2 was .46 ($p = .11$). χ_1^2 due to linear trend was 4.45 ($p = .03$). χ_1^2 testing for departure from linear trend was 0.01 ($p = .96$) and nonsignificant. There is thus a definite linear trend within this dataset.

Our next test is of Prediction 4. Out of 304 Type 2 individuals 39 (12.83%) died of coronary heart disease. Also, 24 out of 326 (7.36%) Type 1 scorers died of coronary heart disease.

Type	Group D	Group A	Group B	Group C
2	12	7	9	11
1	12	4	3	5

Testing the equivalence of Groups A and B and using the same test for the difference between proportions (which yields a χ_1^2 2×2 table-equivalent probability since $\chi_1^2 = z^2$), we have a difference between proportions for Groups A and B of 11.36%, which is not significant at $p = .89$, two-tail, using Yates' correction for continuity.

Collapsing across Groups A and B

Type	Group D	Group AB	Group C
2	12	8	11
1	12	3.5	5

and testing for Group D < Group AB < Group C, the over-all χ_2^2 was 1.95 ($p = .38$). χ_1^2 due to linear trend was 1.57 ($p = 0.21$); χ_1^2 testing for departure from linear trend was 0.38 ($p = .54$, not significant). There is thus a definite trend within this dataset.

We now examine Prediction 5. Out of 810 Types 3 + 4 + 6 subjects 23 (2.84%) died of cancer or coronary heart disease, and 142 (15.59%) out of 911 Types 1 + 2 + 5 subjects died of cancer or coronary heart disease.

Type	Group D	Group A	Group B	Group C
1, 2, 5	57	26	25	34
3, 4, 6	11	5	7	0

Testing of equivalence of Groups A and B, and using the same test for the difference between proportions (which yields a chi squared 2×2 table-equivalent probability since a $\chi_1^2 = z^2$), we have a difference between proportions for Groups A and B of 5.75%. This value is not significant at $p = .56$, two-tail.

Collapsing across Groups A and B

Type	Group D	Group AB	Group C
1, 2, 5	57	25.5	34
3, 4, 6	11	6	0

and testing for Group D < Group AB < Group C, the over-all χ_2^2 was 6.82 ($p = .03$). χ_1^2 for a linear trend was 4.29 ($p = .04$); χ_1^2 testing for departure from linear trend was 2.53 ($p = .11$) which is not significant. There is thus a definite linear trend within this dataset.

Finally, let us consider Prediction 6. Out of 810 Types 3 + 4 + 6 subjects 106 (13.09%) died of other causes and 161 (17.67%) out of 911 Types 1 + 2 + 5 subjects died of other causes.

Type	Group D	Group A	Group B	Group C
1, 2, 5	49	30	43	39
3, 4, 6	49	26	24	7

Testing the equivalence of Groups A and B, and using the same test for the difference between proportions (which yields a chi-squared 2×2-table equivalent probability since a χ_1^2 with 1 $df = z^2$), we have a difference between proportions for Groups A and B of 10.61%, which is not significant ($p = .23$, two-tail).

Then, collapsing across Groups A and B

Type	Group D	Group AB	Group C
1, 2, 5	49	36.5	39
3, 4, 6	49	25	7

and testing for Group D < Group AB < Group C, the over-all χ_2^2 was 15.92 ($p < .01$); χ_1^2 due to linear trend was 14.80 ($p < .01$). The χ_1^2 testing for departure from linear trend was 1.12 ($p = 0.29$) which is not significant. There is thus a definite linear trend within this dataset.

Table 2 shows the results of testing these six predictions. It is clear that five of six predictions are significant but that the differentiation of death from cancer and death from coronary heart disease is less successful than the other predictions. It seems that stress predicts mortality better than *specific* stress predicting *specific* cause of death. In this our results are not unlike those of previous studies (Eysenck, 1991). Over-all, however, it seems that we may use these predictions as a reasonable criterion for comparing the validity of our four methods of data collection.

TABLE 2

PROPORTION OF CASES IN LINE WITH PREDICTION (+) AND CONTRARY TO PREDICTION (-), TOGETHER WITH TEST OF SIGNIFICANCE OF DIFFERENCES

Prediction	Congruence with Prediction		Difference	p
	+	-		
1	10.74	3.23	7.51	< .01
2	12.83	3.25	9.58	< .01
3	10.74	7.57	3.17	.08
4	12.83	7.36	5.47	.01
5	15.59	2.84	12.75	< .01
6	17.67	13.09	4.58	< .01

Summary and Conclusions

The results of this study are fairly clear-cut. We have shown that, in another sample of 1721 healthy probands followed for 13 years, individuals classified as Type 1 died predominantly of cancer, those classified as Type 2 died predominantly of coronary heart disease, and individuals of Types 1 + 2 + 5 were more likely to die of any causes than were individuals of Types 3 + 4 + 6. Results were very similar to those of previous research when data were summed over all four methods of administering the questionnaire.

Using this set of results as our criterion measure, we tested six specific predictions to the effect that for every prediction the control group, receiving no special attention from the interviewer, did worst and the group receiving special attention to establishing trust and receiving explanations did best in prediction against the other two groups who received either the explanation or the trust-establishing procedures that provide an intermediate administration. This sequence of groups appeared significantly on every test (with one exception), and there was a significant linear trend.

These results have an important bearing on the whole question of epidemiological methodology in the psychosocial field. Typically, data are collected by handing out questionnaires, yet this method of data collection is the least

trustworthy and has in the past given the poorest results. The work on the
Type A-Type B concept in relation to coronary heart disease is an obvious ex-
ample; results using the structured interview have usually been superior to
those using the Jenkins Activity Survey (cf. Eysenck, 1990b) in predicting
coronary heart disease. If this be indeed so, then clearly it is quite unjusti-
fied to combine results from the two methods in some form of meta-analysis
(Eysenck, 1984, 1992). Much preferred is Slavin's (1986) "best evidence
synthesis." He suggested that we should consider the best evidence in any
field coming from studies having the highest internal and external validity,
using well-specified, defined, explicit *a priori* inclusion and exclusion crite-
ria, and favouring size-effect data to statistical significance alone when in-
terpreting the literature reviewed. As Newton (1676) pointed out, "it is not
number of experiments, but weight to be regarded; where one will do, what
need of many?"

There are, of course, drawbacks to the extensive use of interviews as
opposed to administration of a simple questionnaire. Interviewers have to be
carefully selected, have to be trained, and must be paid; in addition, inter-
viewers' administration of tests takes far longer than the more usual methods
of data collection by handing out questionnaires. However, if the issues are
important and the social or scientific consequences of great moment, it may
be a false economy to use the less reliable and valid methods. Particularly in
novel fields, where orthodoxy much prefers committing errors of Type 2 to
the remote possibility of committing errors of Type 1, using the less efficient
method plays into the hands of critics only too eager to deny the value of the
new ideas. The Schmale and Iker (1971) study already mentioned illustrates
how inappropriate methods give nonsignificant results; appropriate methods
give significant results (cf. Eysenck, 1990a). Indeed, to carry out mumpsi-
mus studies may be counterproductive, giving the impression that theories
are false when in fact the lack of support rests on faulty methodology.

It has to be admitted that there may be problems other than those al-
ready mentioned in the use of interviewers. Not all are highly motivated by
scientific ideals, and some may try to obtain their pay by spurious means,
e.g., by filling in questionnaires without actually contacting the supposed in-
terviewees. This has sometimes happened in epidemiological, criminological,
and other enquiries, and makes necessary careful supervision and subsequent
statistical control.

In the present project Dr. W-D. Heller, an independent observer from
the Karlsruhe Institute of Statistics, interviewed a random sample of inter-
viewers to see to what extent they had understood and carried out instruc-
tions. In addition, Dr. H. Vetter, who had put all the data on computer,
carried out many routine checks on accuracy and authenticity of the data. In
the course of doing so, he discovered a small section of data from previous

studies showing suspicious identity of responses, presumably because one interviewer filled in the forms in an identical manner. The elimination of these individuals did not alter the significance of the results; indeed, adding random errors to a meaningful set of data would only serve to make them fall short of the statistical criterion for significance (Eysenck, 1993). Of course, such errors should be avoided, but in large-scale studies they are probably unavoidable; they only impugn the results if they are *directional*, i.e., favour the hypothesis being tested.

Our findings should, of course, not be generalized too far. What applies to issues very important to subjects and likely to force them to admit to socially very undesirable attitudes and behaviours may not apply to the issues covered by ordinary personality questionnaires which do not invoke stress and reaction to stress. Also, the Grossarth-Maticek theories are quite complex so the questionnaires designed by him are often complex and difficult to understand, particularly by less well-educated subjects (who probably constituted the majority of respondents for we are not dealing here with the usual sophomores!).

Crucial to this whole problem is the question of *motivation*. Michaelis and Eysenck (1971) have shown that, as the motivation to dissimulate (make oneself appear better) increases, scores on the Neuroticism scale decline, those on the Lie scale increase and so do the (negative) correlations between Neuroticism and Lie scale scores. In interpreting questionnaire results, one must always bear in mind (and preferably attempt to measure) the motivation of one's subjects. Schmitz (1992) obtained better results than Amelang and Schmidt-Rathjens (1992) possibly because Schmitz's subjects had joined his therapeutic class to obtain help; this would constitute powerful motivation to answer questionnaires truthfully even though they were not administered by an interviewer. He also used a simplified form of the Grossarth-Maticek questionnaire, thereby making understanding easier. Such points are important in understanding differences in the results from different studies, which a meta-analysis can only obscure. Differences in age are also important; a student sample might not have a single case of cancer or coronary heart disease or have much experience of adult types of stress.

We may conclude that the study of illness-proneness by questionnaires is a very complex field in which much of the work done has been methodologically defective. Hence the importance of studies giving positive results and also of the need for independent replication. It must also be stressed, of course, that replication indeed means replication; too often the term is used in an imprecise sense as referring to a study vaguely similar in intent to the one to be "replicated" but with sufficient possibly important differences to make different outcomes uninterpretable. We believe that the issue is sufficiently important to make it desirable that independent replications (true replications!) should be undertaken to support or refute our results.

REFERENCES

AMELANG, M., & SCHMIDT-RATHJENS, C. (1992) Psychometric properties of modified Grossarth-Maticek and Eysenck inventories. *Psychological Reports*, 71, 1251-1263.

ARMITAGE, P. (1955) Tests for linear trends in proportions and frequencies. *Biometrics*, 11, 375-386.

ARMITAGE, P., & BERRY, C. (1991) *Statistical methods and medical research*. London: Blackwell Scientific Publ.

BOOTH-KEWLEY, S., & FRIEDMAN, H. S. (1987) Psychological predictors of heart disease: a quantitative review. *Psychological Bulletin*, 19, 343-362.

EYSENCK, H. J. (1984) Meta-analysis: an abuse of research integration. *Journal of Special Education*, 18, 41-59.

EYSENCK, H. J. (1990a) The prediction of death from cancer by means of personality/stress questionnaire: too good to be true? *Perceptual and Motor Skills*, 71, 216-218.

EYSENCK, H. J. (1990b) Type of behaviour and coronary heart disease: the third stage. *Journal of Social Behavior and Personality*, 5, 25-44.

EYSENCK, H. J. (1991) *Smoking, personality and stress: psychosocial factors in the prevention of cancer and coronary heart disease*. New York: Springer-Verlag.

EYSENCK, H. J. (1992) Meta-analysis: sense or non-sense. *Pharmaceutical Medicine*, 6, 113-119.

EYSENCK, H. J. (1993) Reply to van der Ploeg, Vetter, and Kleijn. *Psychological Inquiry*, 4, 70-73.

EYSENCK, H. J., & EYSENCK, M. W. (1985) *Personality and individual differences: a natural science approach*. New York: Plenum.

FOX, B. H. (1978) Premorbid psychological factors as related to incidence of cancer. *Journal of Behavioural Medicine*, 1, 45-133.

FOX, B. H. (1981a) Behavioral issues in cancer. In S. M. Weiss, J. A. Herd, & B. H. Fox (Eds.), *Perspectives on behavioral medicine*. London: Academic Press. Pp. 101-134.

FOX, B. H. (1981b) Psychosocial factors and the immune system in human cancer. In R. Ader (Ed.), *Psychoneuroimmunology*. Orlando, FL: Academic Press. Pp. 103-158.

FOX, B. H. (1983) Current theory of psychogenic effects on cancer incidence and prognosis. *Journal of Psychosocial Oncology*, 1, 17-31.

GROSSARTH-MATICEK, R. (1979) *Soziales Verhalten und die Krebserkrankung*. Basel: Beltz.

GROSSARTH-MATICEK, R. (1986) Psychosociale Verhaltenstypen und chronische Erkrankungen. *Der Kassenrat*, 39 (September), 26-35.

GROSSARTH-MATICEK, R. (1989) Disposition, Exposition, Verhaltensmuster, Organverschaedigung und Stimulierung des Zentralen Nervensystems in der Aetiologie des Bronchial-, Magen- und Leber Karzinoms. *Deutsche Zeitschrift fuer Onkologie*, 21, 62-78.

GROSSARTH-MATICEK, R., & EYSENCK, H. J. (1990) Personality, stress and disease: description and validation of a new inventory. *Psychological Reports*, 66, 355-373.

LEE, P. N. (1988) *Misclassification of smoking habits and passive smoking*. New York: Springer-Verlag.

LE SHAN, L. (1959) Psychological states as factors in the development of malignant disease: a critical review. *Journal of the National Cancer Institute*, 22, 1-18.

LE SHAN, L. (1961) A basic psychological orientation apparently associated with neoplastic disease. *Psychiatric Quarterly*, 35, 314-330.

LE SHAN, L. (1977) *You can fight for your life: emotional factors in the causation of cancer*. New York: M. Evans.

LE SHAN, L., & REZNIKOFF, M. (1960) A psychological factor apparently associated with neoplastic disease. *Journal of Abnormal and Social Psychology*, 60, 439-440.

MICHAELIS, H., & EYSENCK, H. J. (1971) The determination of personality inventory factor patterns and intercorrelations by change in real life motivation. *Journal of General Psychology*, 118, 223-234.

NEWTON, I. (1676) Letter to Oldenburg, 18th August.

PRICE, V. A. (1982) *Type A behavior patterns*. New York: Academic Press.

QUANDER-BLAZNICK, J. (1991) Personality as a predictor of lung cancer: a replication. *Personality and Individual Differences*, 12, 125-130.

RANCHOR, A., SANDERMAN, R., & BOUMA, J. (1992) The empirical basis of the personality

types of Grossarth-Maticek and Eysenck. Poster presented at the 6th European Conference of Personality, Groningen, 15-19th June.

SANDÍN, B., CHOROT, P., NAVAS, Z., & SANTED, M. (1992) Estres y enfermedad: inventario de reacciones interpersonales de Grossarth-Maticek y Eysenck. *Revista de Psicología Generale y Applicada*, 45, 391-396.

SCHMALE, A. H., & IKER, H. (1971) Hopelessness as a predictor of cervical cancer. *Social Science and Medicine*, 5, 95-100.

SCHMITZ, P. G. (1992) Personality, stress-reactions and disease. *Personality and Individual Differences*, 13, 683-691.

SLAVIN, R. E. (1986) Best-evidence synthesis: an alternative to meta-analysis and traditional reviews. *Educational Research*, 19, 9-11.

TEMOSHOK, L., & DREHER, H. (1992) *The Type C connection*. New York: Random House.