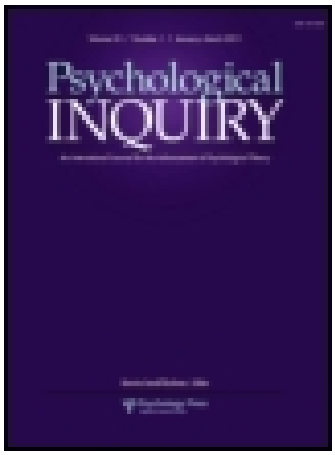

Psychological Inquiry: An International Journal for the Advancement of Psychological Theory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hpli20

Reply to van der Ploeg, Vetter, and Kleijn

Hans J. Eysenck

Published online: 19 Nov 2009.

To cite this article: Hans J. Eysenck (1993) Reply to van der Ploeg, Vetter, and Kleijn, Psychological Inquiry: An International Journal for the Advancement of Psychological Theory, 4:1, 70-73, DOI: 10.1207/s15327965pli0401_16

To link to this article: http://dx.doi.org/10.1207/s15327965pli0401_16

Psychological Inquiry
1993, Vol. 4, No. 1, 70–73

AUTHOR'S RESPONSE

Reply to van der Ploeg, Vetter, and Kleijn

Hans J. Eysenck
Institute of Psychiatry
University of London

As on previous occasions, I have asked Professor Grossarth-Maticek to provide an answer, to which I will then add some comments. Before providing his answer, Grossarth-Maticek makes an important general statement that serves to set the whole debate into its proper background:

> In our prospective studies from 1972 to 1978, we investigated more than 20,000 people, aided by more than 100 paid scientific assistants who carried out the interviews. Our general principle was that, in the organization of data collection, documentation, computerization, statistical treatments, and final diagrammatic or tabular presentation, sources of error not exceeding 6% could be tolerated. For this reason, controls were introduced in the collection of data, documentation, and analysis. In the Heidelberg study, on reexamination we discovered error sources between 0.5% and 6.0%. As an example, in a reassessment of the sex of the subjects in the first follow-up study in 1982, it was found that, for 23 subjects in the stressed group (1.4%), the wrong sex had been given originally.
>
> In conversation with the highly respected epidemiologist A. Lilienfeld, I mentioned that in our studies we could not guarantee accuracy overall of better than 95%. He laughed and said that in such large-scale epidemiological studies an error rate of 5% was considered not only acceptable, but good; errors of that magnitude were unavoidable when so many co-workers were involved whom one could not control individually at all times.
>
> One general principle underlies large-scale population studies: Errors are unavoidable when large samples are under investigation. Modern epidemiologists start with the belief that studies with only 5% errors are recommendable. The second principle, however, makes it clear that such errors must be random; they must on no account be in line with an improvement in the overall correlation between theory and data. It is to safeguard against directional (as opposed to random) errors that, whenever errors were found on reexamination of the data, new statistical investigations were initiated in which probands in whose documentation errors were found were eliminated. We have regularly found that eliminating the error did not affect the final result—or affected it only to a completely negligible extent.

Critics like van der Ploeg and Vetter search systematically for minimal sources of error in order to publicize the results, regardless of the fact that these errors do not in fact improve the final results and hence are not systematic. Publication of nonsystematic errors has little point, as they fail to substantiate the only possibly important factor of data manipulation, which would be associated with systematic error. Van der Ploeg seems unaware of the many possibilities of random error inherent in large-scale epidemiological studies and commits the error of treating such studies as if they were exactly comparable to controlled psychological experiments involving very small numbers of subjects. Hence, he interprets unavoidable error in documentation as intentional data manipulation without being in the position of proving that these errors have actually influenced or improved the final result. At one international conference, he reported that, for 3 subjects of 2,661, age had not been correctly reported, and he accused me of data manipulation as a consequence! In actual fact, age had been correctly stated, even for these 3 subjects. Van der Ploeg failed to answer an epidemiologist who inquired whether he would seriously maintain that the whole outcome of the study would be affected because three age statements had been in error.

Using different statisticians, we have made a systematic search for errors; these searches resulted in only 10 successful discoveries. Vetter, who analyzed the whole set of data from all our studies, found errors of any importance in only 1% of his critical studies, and these have of course been corrected. This suggests that the quality of the data is excellent and that the proportion of errors is no higher than is usual in modern epidemiological studies. One might even claim that my critics have done me an unwitting service. They agree, as Amelang (1991) wrote in his commentary on Eysenck's (1991b) target article, that one would have to search for many small sources of error in hope that these would have a synergistic effect—obviously despairing of finding large sources of error. Critics must certainly be grieved by the fact that, over the years, it has been possible for me to answer all their questions.

In fact, van der Ploeg sent his article (van der Ploeg & Vetter, this issue) to *Psychological Inquiry* but not to me, which constitutes a definite breach of the agreement—signed by van der Ploeg—that he would show his analyses to me first in order to give me the opportunity to reply. Vetter too sent his article (this issue) to

Psychological Inquiry without giving me the opportunity to reply. This is not the behavior of scientifically objective critics.

Such lack of objectivity is equally obvious in the Vetter article (this issue), which claims to discover lack of dose–response relation between type and lung cancer. In his analysis, Vetter used the sum of Types 1, 2, and 5, yet only Type 1 shows a specific relation with cancer. The simultaneous presence of Type 2 and Type 5 would serve only to confound such an effect. Indeed, the behavior of Type 2 has been shown to be very different from—and in some ways antagonistic to—the behavior of Type 1. The obvious method of analysis would have been to use Type 1 scores possibly with Type 4 scores (negative) in order to test the hypothesis of a linear dose–response relation. The generalized sum of Types 1, 2, and 5 is relevant to overall mortality but certainly not to lung cancer specifically. When asked to calculate the dose–response relations for Types 1 and 4, Vetter refused, perhaps because he did not want to correct his hasty dissemination of a profoundly faulty analysis. His critical paper should have included the relevant data for Type 1 and Type 4; to submit it without these data is not responsible scientific criticism.

I now return to van der Ploeg's article (this issue), which at first appeared to have Vetter as co-author, then had Vetter's name removed, and finally has it restored. Vetter (not van der Ploeg) had noticed in going over the data that, for some subjects, the pattern of answers to the questions at the beginning of the questionnaire was identical. Van der Ploeg used this discovery to write his article with Vetter as co-author. Vetter carried out extensive statistical calculations but failed to discover any improvement in the predicted personality versus disease correlations when the offending subjects were eliminated, disproving the possibility of data manipulation, which would have led to an improvement in the results. Having discovered this negative result, Vetter pulled out of the co-authorship but apparently agreed to be put back again, but without insisting on a mention of the only really important finding—namely, that the faulty data made no difference to the result. [In a May 11, 1992, letter to me, Vetter confirmed that "the effect ... on the overall prediction of mortality is negligible"!—Eysenck.]

The cause or meaning of the partial identity of answers is not clear. Only some of the psychosocial questions were answered in an identical fashion; the remainder and all answers to questions concerning the physical variables differ to a considerable extent. The identical sets of answers show neither a cancer-prone nor a coronary-heart-disease–prone pattern, but rather a Type 4 pattern. How this arose, whether it is a statistical accident in a small number of a very large sample, or whether a lazy interviewer fabricated data will probably never be known. The main point is that the faulty data did not improve the outcome and hence are clearly evidence of lack of data manipulation.

Turning next to the van der Ploeg and Kleijn article (this issue), it may be seen that the errors they noted amount to 4.8% with respect to cholesterin or 5.6% with respect to lymphocytes. The errors arose from the complexity of the study. The intention was to measure cholesterin up to seven times in 2,561 people, at intervals of several months, as well as percent of leukocytes. Sixty-three scientific co-workers controlled groups of 40 people at any given time, asking them at regular intervals to give blood. Some refused, others allowed this to be done on one, two, three, ... up to seven occasions. Every subject had a special number, and, after measurement was completed, the assistants handed over the data to five documentation workers who transferred them into a central data source.

When the study was completed in 1977, having been begun in 1972, probands were assigned new numbers. In 1987 an assistant was asked to select from the highly stressed groups probands for whom all seven measures had been completed—in order to answer the question of what proportion of subjects showed a variation between the lowest and highest value that exceeded 80mg percent. Our hypothesis was that such people would be more susceptible to cancer. Due to the change in numbers, errors were made in 1.6% of the cases when cholesterin and lymphocyte values were attributed to the wrong people. These errors occurred only in the highly stressed group because the normal sample was examined by assistants who knew about the renumbering, whereas the stressed sample was examined by assistants who seemed ignorant of the change.

What can one say in evaluating the additional criticisms here published, and Grossarth-Maticek's reply? In the first place, he is right in saying that, in large-scale epidemiological inquiries, it may be taken for granted that a certain proportion of random errors is unavoidable; several well-known practitioners have told me so, although they might be less willing to say so in print. Perhaps this is the reason why such a large proportion of psychologists are unwilling to allow anyone access to their poor data (Craig & Reese, 1973; Wolins, 1962). Note in contradistinction that Grossarth-Maticek has, since the beginning of his work, allowed any interested scientist to inspect and use his data. Vetter has played with these data for 18 years, van der Ploeg for 6, and others too have had complete access to this rich treasure trove without let or hindrance. This alone speaks strongly for the genuineness of the data.

In the second place, Grossarth-Maticek argues that random errors are not a great danger to the conclusions of epidemiological research as long as they remain below the 6% level; it is systematic errors that suggest actual manipulation of data. In spite of high motivation, Vetter, van der Ploeg, and all other statisticians, epidemiologists, and psychiatrists who have had access to Grossarth-Maticek's data have not been able to discover data showing systematic bias that would be sufficient to change the conclusion significantly. In one or two cases, reanalysis with methodological improvements actually improved the significance of the

results. This surely must suggest that, although the data are not immaculate, they certainly show no evidence of manipulation. If an outstanding and very critical statistician like Dr. Vetter, after 18 years of continuing analysis of Grossarth-Maticek's data, has been unable to find any evidence of malpractice, but only very occasional evidence of random error unrelated to final outcome, it is difficult to refuse the conclusion that there simply is no such evidence. The same is true of van der Ploeg, who failed equally, after much searching, to discover any critical errors. Who of us, in allowing a hostile critic to reanalyze literally millions of pieces of data, would be able to lay his hand on his heart and say "There will be no error found"?

My third reaction is that, as a hard-boiled experimentalist to whom all errors are anathema, I still feel horrified when even quite small errors are detected in any research with which I am connected. I know that, in research in behavioral genetics, there is a built-in 5% error (zygosity ascertainment), but, although I know cognitively that this will not affect conclusions beyond an ascertainable degree, emotionally I still hark back to the happy old days when I was working with small numbers of twins, and zygosity could be ascertained with 100% accuracy—an ideal obviously unattainable now that we are dealing with 12,000 or even 15,000 pairs of twins. Clearly, at least some of the random errors in Grossarth-Maticek's work could have been avoided by closer supervision and scrutiny, and I have great difficulties in becoming reconciled to the existence of these errors on the grounds that they are random. To fail to make sure that all co-workers are informed about a change in the numbering seems to me inexcusable and showing a deplorable lack of supervision. Nevertheless, I do feel that the immense contributions made by these well-designed prospective studies—unique in their theoretical relevance, their large size, and their well-supervised final documentation of mortality—make any such emotional reactions irrelevant. If the data are genuine, we simply cannot disregard them. Discovery of minor sources of random error, however painful to me personally, cannot change that situation.

Right from the beginning, when I first became involved with the analysis of Grossarth-Maticek's data, the question of their genuineness has been uppermost in my mind, and my main effort has been to devise a foolproof method of ascertaining the truth about this question. I believe that a final verdict is now possible on objective grounds, and I am glad to say that the verdict is positive. It will be remembered that the 10-year follow-up of the Heidelberg group was continued for another 4½ years in order to test the genuineness of the original data and that the test proved positive (Eysenck, 1991). Vetter (1991c) threw some doubt on the adequacy of this proof in terms of his "morbidity hypothesis" but was forced to retract this hypothesis on the basis of his own calculations (Vetter, 1991a, 1991b). In essence, his test involved comparing 1982–1986 follow-up subjects whose illness was discovered and diagnosed before or after the beginning of the follow-up period. Using a small group of subjects, he found convincing evidence for the genuineness of the data but concluded that "it is deemed highly desirable ... to extend the test to all (or most) of the deaths that occurred from 1982 to 1986 by collecting information from doctors and/or relatives when the disease that led to death first became known" (Vetter, 1991b, p. 323).

This has now been done, the information being collected independently by a Heidelberg clinic, and the results are given in Table 1. (Analysis is along the same lines as in Vetter, 1991b.) It will be seen that the results are the same as before and, as they were calculated by the critic (Vetter) himself, may be accepted as authentic. It is difficult to know what more needs to be done to establish the genuineness of the Grossarth-Maticek data. All original test data were in the hands of Vetter and of two independent university departments when the follow-up started, and all determinations of mortality and of time when the disease first became known were carried out independently. I know of no other epidemiological study that has received a fraction of the care and attention to establish its genuineness than that received by the Grossarth-Maticek study.

Why then the continuing battle? Essentially we are dealing with a conflict between two personality types who will never learn to appreciate the virtues of the other. Grossarth-Maticek is the wide-ranging, creative scientist, working on a large scale, impatient of detail,

Table 1. *Predictor Squared Multiple Correlations, 1982 to 1986 Data*

	Set of Predictors			
	Smaller		Larger	
Sample	R^2	p	R^2	p
Normal				
Prior Group[a]	.011	.090	.022	.100
Later Group[b]	.023	.002	.031	.011
Regression Coefficients[c]				
Prior Group[a]	.000	1.000	.001	1.000
Later Group[b]	.007	.019	.007	.022
Stressed				
Prior Group[d]	.017	.060	.033	.035
Later Group[e]	.067	.000	.076	.000
Regression Coefficients[c]				
Prior Group[d]	.015	.003	.023	.001
Later Group[e]	.054	.000	.056	.000

[a] $n = 3$. [b] $n = 6$. [c] Derived from rest of cancer deaths. [d] $n = 10$. [e] $n = 29$.

concerned with the wider issues, the broad strokes, the major breakthrough. Irritated by doubts and criticisms, conscious of the enormous social and scientific importance of his discoveries, convinced (rightly) that his work and theories are streets ahead of what his critics have to offer, he obviously does not suffer fools gladly, and he may hit out at them in a rather exaggerated way. To Vetter and van der Ploeg, on the other hand, I would apply Madame Curie's famous words: "There are sadistic scientists who hasten to hunt down error instead of establishing the truth." Pedantic to the last degree, any error, however slight, random, and unimportant from the point of view of the grand design, is a sin against the Holy Ghost, to be hunted down, exposed, and eradicated. This battle is age-old, and few creative scientists escape it. I have more sympathy with the attitude of Vetter and van der Ploeg than does Grossarth-Maticek, but I do find, after observing this conflict for many years, that the critics are right on a few minor points but disastrously wrong on the major ones. There seems to be no escape from the conclusion that the Grossarth-Maticek data are genuine—even though at times marred by errors that fortunately make no difference to the overall conclusions. Even Vetter, arch-critic and more intimately acquainted with all the Grossarth-Maticek data than anyone, has been forced to this conclusion. What more can one say?

I would like to end with a word of warning. Medicine has often persecuted its greatest innovators and heroes and driven them to death. The story of Semmelweiss, the great precursor of the antiseptic revolution, who reduced pueperal fever deaths in his Vienna hospital from some 20% to complete insignificance, was disbelieved, ridiculed, and finally driven out; thousands of women died because orthodoxy failed to appreciate the rebel and his genius. I would like to think that we will not allow Grossarth-Maticek, whose message holds out hope for thousands that cancer and coronary heart disease are not inevitable, to share his fate. To those like Vetter and van der Ploeg, I quote Dr. Johnson's famous words: "I can give you an argument, I cannot give you an understanding." Perhaps we may conclude with a syllogism that I shall leave readers to finish. If Grossarth-Maticek's data are genuine, he is a genius. His data have been shown to be genuine; ergo ...

Note

Hans J. Eysenck, Institute of Psychiatry, University of London, Denmark Hill, DeCrespigny Park, London, SE5 8AF, England.

References

Amelang, M. (1991). Tales from Crvenka and Heidelberg: What about the empirical basis? *Psychological Inquiry, 2,* 233–236.

Craig, J. R., & Reese, S. C. (1973). Retention of raw data: A problem revisited. *American Psychologist, 28,* 723.

Eysenck, H. J. (1991a). Appendix A: Analysis of mortality data in the 1972 prospective Heidelberg study by Grossarth-Maticek, covering the period 1982 to 1986. *Psychological Inquiry, 2,* 320–321.

Eysenck, H. J. (1991b). Personality, stress, and disease: An interactionist perspective. *Psychological Inquiry, 2,* 221–232.

Vetter, H. (1991a). Addendum: The "mortality hypothesis" retracted. *Psychological Inquiry, 2,* 323.

Vetter, H. (1991b). Appendix B: Some genuine predictions of Grossarth-Maticek's established. *Psychological Inquiry, 2,* 322–323.

Vetter, H. (1991c). Some observations on Grossarth-Maticek's data base. *Psychological Inquiry, 2,* 286–287.

Wolins, L. (1962). Responsibility for raw data. *American Psychologist, 17,* 657–658.