Acceptance address van professor Hamaker bij het in ontvangst nemen van de Shewhart Medal 1979 op 22 mei 1980 in Atlanta, U.S.A.

De American Society for Quality Control kende de Shewhart Medal 1979 toe aan prof.dr. H.C. Hamaker. De motivering van dit besluit luidde als volgt:

'The Shewhart Medal established by the American Society for Quality Control for Outstanding Leadership in the Field of Modern Quality Control awarded to Hugh C. Hamaker for the year 1979.
As a pioneer and original thinker in Quality Control and for his concept of 'indifference quality' in sampling plans, his introduction of statistically designed experiments in European industry, his continuing work in quality standardization, his influential teaching and his distinguished publications.'

De redactie feliciteert hem van harte met dit teken van erkenning. De medaille werd aan hem uitgereikt op de 34th Annual Technical Conference in Atlanta op 22 mei 1980. Bij die gelegenheid hield professor Hamaker de volgende toespraak.

Seeing myself as a Shewhart Medalist?

Since I received the happy message that I was to be the Shewhart Medalist for 1979 I have been wondering how I could justify myself as such. I still have my doubts, but I will do my best.

Between 1925 and 1930, while I was studying experimental physics, I started regularly reading the leading journal 'Biometrika'; what prompted me to do this I cannot remember. In the Netherlands a thesis has to be supplemented by a number of propositions showing that the prospective doctor has some interest and ideas of his own on subjects outside that of his main field of research. My first proposition reads as follows: 'It is desirable that in text-books and in the teaching of probability theory attention is paid to the concepts and methods of mathematical statistics.' That was in 1934, the year in which I joined the Philips Research Laboratories in Eindhoven as a research physicist. I was therefore early in recognizing the importance of statistical theories.

Up to and during World War II I was engaged in various research projects in physics and in colloid chemistry, all the time, however, keeping my interest in statistics alive; so that, when the war was over, I became a member of a small committee that had to supervise the introduction of QC and Sampling Inspection in the Philips factories in the Netherlands. The stimulation of the use of these new techniques on the factory floor was the task of the Department of Technical Efficiency and Organisation, while I was able to assist by solving or studying the more theoretical problems.

In the British Philips factories a sampling inspection procedure had been adopted in which a lot was sentenced on the basis of the highest number of defectives observed in one out of five independent samples drawn from it. It was in order to prove the inefficiency of this method that I first used the combination of the indifference quality and the relative slope as the two parameters for specifying an OC curve instead of using two points arbitrarily adopted [1]. Since the citation on the Shewhart Medal mentions only the 'indifference quality' I wish to emphasize here that the concept of the 'relative slope' is equally important and perhaps the less obvious of the two.

In various sections of the Philips industry research is carried out in direct relation to production problems. The next step, after QC, was to demonstrate the usefulness of a statistical analysis of the extensive data collected in the past by this kind of research. By applying analysis of variance I could establish or confirm conclusions which without the use of statistics had taken years of experience. In another case a discussion with an electrical engineer, G.J. Levenbach, of a life testing problem made him realise that several times in his career he had come up against problems where a statistical analysis had been needed, had he but known about it. He then switched to statistics, and when he emigrated to the U.S. was engaged by the Bell Telephone Laboratories as a statistician.

More important perhaps were the cases where I could demonstrate that the information required was not available in the data obtained, because the experiments had not adequately been planned. In this way, by acting as a consultant to the research groups in the factories, and by lecturing about the results achieved, I stimulated the application of designed experiments in the Philips industries. My position was a particularly fortunate one, because these industries produce a very wide variety of products including apart from lamps and electronics, also glass and pharmaceuticals; which led automatically to an equally wide variety of useful examples.

My main source of statistical inspiration must have been the books by Fisher [2], Cochran and Cox [3], and by O.L. Davies and others [4], plus articles in the statistical journals. But if my work in this field has been particularly fruitful, this is, I believe, chiefly due to the mentality in which I tackled the practical problems.

Early in his famous book W.A. Shewhart [5], p. 5 cites 'Nature' of January 1926: 'A large amount of work has been done in developing statistical methods on the scientific side, and it is natural for anyone interested in science to hope that all this work may be utilized in industry and commerce.'

Then a few pages further on (p. 18) he adds: 'The available statistical machinery referred to by 'Nature' is, as we might expect, not an end in itself, but merely a means to an end. In other words, the fact that the criterion which we happen to use has a fine ancestry of highbrow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works.'

Consciously or unconsciously it is in this spirit I have always acted. All too often theoretical statisticians see the development of new models and new theorems as an end in itself; in applications statistical techniques must be seen only as a tool, and the tool must be adjusted to the practical purpose which it has to serve.

I have, for example, always been critical of tests of hypotheses. Because a small and unimportant technical effect can turn out to be statistically significant by a sufficiently large number of data; while an important effect can be found insignificant when the data are too few.

Likewise I have refused to apply statistical ideas when they were not needed. In one case my advice was requested for designing an experiment with 5 factors, while my clients could not supply any information at all concerning the effects of these factors when varied singly. The basis for a multiple factor design is then missing, and I told my clients to carry out some one-factor experiments first. They never came back because they had soon found out what was the source of their troubles.

Some basic understanding of the subject and purpose of an experiment is also of fundamental importance. The chemical department of the University of Technology in Eindhoven asked me to design an experiment for the combustion of a sludge suspended in water by a continuous process for which an experimental set-up had been constructed. The chief factors involved were 1) the pressure, 2) the temperature, 3) the concentration and 4) the rate of flow of the suspension, 5) the rate of supply of oxygen, and 6) the rate of stirring. From chemical principles it was evident that for some of these factors the yield curves would have horizontal asymptotes, so that a simple linear response function would not do. The fact that I knew enough about chemistry to ask for a reaction equation which could serve as a model was a great help in this case. A first trial by a 3 x 3 experiment revealed that one of the parameters could not be estimated at all. But my assistant at that time, H.N. Linssen, is now in constant demand by the chemical department to assist in designing and analysing experiments for the non-linear situations chemical reaction equations lead to.

Similarly I have found that in a multifactor situation a series of 2-factor or 3-factor experiments based on a judicious choice of the factors successively introduced may be simpler and more effective than a full scale experiment involving a larger number of factors [6]. But here again it is the judicious choice that matters and that requires a good understanding of the technical problem to be solved.

And when an analysis of variance shows a significant interaction, the next problem is to explain this interaction in technical terms, and if you succeed in this respect it may be best to keep the analysis of variance to yourself.

Teaching by lecturing in the university I have always found a difficult matter, and I doubt whether I have been very successful in that respect. But I am a great believer in training people on the job. Many of my assistants in the Philips Research Laboratory have, after working three or four years under my supervision, been taken over by one of the factories where they now hold independent positions as applied statisticians. And the students in the university who wanted to major in applied statistics have always been posted somewhere in industry for their final research project. One case was concerned with a printing-office producing the envelopes for cigarette packages. Over a long period covering many thousands of items inspected the producer had found 4.5% defectives, the consumer 36.2%. There were 24 possible defects but only for two of these the terms used were the same; the terminology of the producer related to the production stage where the defects occurred, that of the cigarette manufacturer to the impression the defects would make on the ultimate consumer, the cigarette smoker. To sort out such a situation with the aim of proposing an acceptable sampling inspection procedure is an excellent exercise for training an applied statistician.

I reached retiring age from my industrial position already 13 years ago and such more recent practical problems as Reliability, Product Liability, Multivariate Analysis, Forecasting and Prediction, or the Use of Computer Packages never came my way. With regard to these subjects I am an old man living in the past.

As to Europe as a whole the situation is much the same as it is in the U.S. We have a European Society for Quality Control (EOQC) organizing Annual Conferences and we have an annual Meeting of European Statisticians organized by the Bernoulli Society in cooperation with the national societies in different countries. There are chairs in statistics in many European universities, either in the mathematical or in some other department. In the Netherlands for instance, we have a professor of statistics in three of our medical faculties. But my own part in these organisational developments has been negligible; I am first and foremost a scientist, not a manager.

My feeling with respect to present day literature has recently been aptly worded by Sir Claus Moser in his acceptance speech as President of the Royal Statistical Society in England [7]. I quote: 'As statisticians we quickly learn of the crosses we have to bear. We know how people move away from us at parties when they learn of our profession, we know that look of incredulity, amusement and resigned boredom. ..... So a general sense of inadequacy, at least in the eyes of others, is part of our life. But for statisticians like me there is an additional problem. When at the party others move away so that one is left talking to the only other statistician present, the chances are that if he is a mathematical statistician, even he will move away. The gap between theoretical and applied statisticians is wider than ever. Most applied statisticians now will understand only a small part of what appears in the journals or at conferences. A decade or so ago, we would have been able to understand all of the Journal of the Royal Statistical Society A, some articles in the Society's other journals, and a fair amount of the Review of the International Statistical Institute or the Journal of the American Statistical Association. This is no longer so. Considerable mathematical ability is now needed to understand more than a small proportion of any of these journals. Moreover much of what is written is irrelevant to, and oblivious of, applications in business, government or other important fields. More often papers deal with theories looking for data rather than with real problems needing theoretical treatment.'

To this I am even inclined to add theories for the sake of theories; or theories in search of a university career, as a friend bluntly suggested. This is a very serious matter. Since my retirement I have been doing some work on behalf of the standardisation of statistical techniques for practical purposes; this as a member of the Technical Committee 69 (TC 69) of the International Standardisation Organisation (ISO) in Geneva. In my view this work is seriously encumbered by the persistent differences between the theoreticians and the practicians. The former are inclined to stick to their theoretical principles and are insufficiently aware of the very crude and approximate nature of practical requirements. Some examples may help to illustrate this point.

Duplicate tests do not as a rule yield identical results owing to the unavoidable errors of observation; and the differences are usually larger when the tests are made in different laboratories than when they are made in a single laboratory. This has led to the concepts of a repeatability, r, a reproducibility, R, and a standard (ISO 5725) for the experimental determination of these two quantities. In that standard R is defined as follows: 'The reproducibility, R, is the value below which the absolute difference between two single test results obtained with the same method on identical test material under different conditions (different operators, different laboratories and/or different times) may be expected to lie with a specified probability; in the absense of other indications the probability is 95%.'

Thus $R = 1.960 \sqrt{2}\, \sigma_R$, where σ_R is the standard deviation between single tests performed in different laboratories. But that is theory!

To determine R one has to send carefully prepared test samples to 15 to 20 laboratories in different countries to be tested by some properly standardized test method. One then has to replace σ_R by an estimate s_R with at best about 20 degrees of freedom and, according to normal theory, a coefficient of variation of the order of $V(s_R) = 1 / \sqrt{40} = 16\%$. Hence s_R may vary between 0,68 σ_R to 1,32 σ_R and the probability associated with R may vary from case to case between 91% and 99.5%. And this cannot be corrected by using a t instead of the factor 1,960, because the t-distribution requires the use of a fresh estimate s_R every time, while here one single estimate is applied repeatedly.

In practice this does not matter. In the manual 'Standard Methods for Testing Tar and its Products' [8], in which r and R are consistently applied, it is only stated: 'Single results submitted by each of two laboratories should be considered suspect if they differ by more than R', where R is given as a numerical value or set of values according to the method of test considered. Whether, if a difference larger than R is observed, any further action is required depends entirely on the circumstances envisaged and cannot possibly be standardized. But disputes as to whether a difference is in some sense significant or not must be avoided and for this purpose a numerical value for R is of great practical value. Whether the associated probability should be 90% or 99% or any other value in between it is impossible to make out; the knowledge that it is higher than 90 suffices for practical purposes.

Another example is provided by the Military Standard 105D, which has recently without change been taken over as ISO 2859 by the International Standardisation Organisation and as IEC 410 by the International Electrical Commission. The binomial and the Poisson distributions have been applied to derive the OC curves and to construct double and multiple sampling plans corresponding to the single sampling plans. But all the other essential features of the standard, viz.:

a) a discrete set of preferred AQL's, } both forming a geometric series
b) a discrete set of sample sizes, }
c) 7 Inspection Levels specifying 7 different relations between lot size and sample size,
d) sample sizes independent of the AQL for a given Lot Size-Code Letter combination,
e) Tightened and Reduced Inspection with the Switching Rules,

are not based on advanced statistical theories, but are the result of practical considerations supported by simple statistical logic.

The numerous publications dealing with sampling inspection as an economic problem have never been applied on a scale worth mentioning. Simplicity is the keystone of success and these theories render the techniques too complicated for practical purposes. Moreover, they are almost invariably based on a false assumption; namely, that the production process can be represented by a stable a priori distribution to which a sampling plan has to be adjusted so as to minimize a certain cost or utility function. The reverse is true; by using a sampling standard the sampling procedure that will be applied is agreed between producer and consumer in advance, and it is the production process that has to be adjusted to it so as to avoid rejected lots. Because these are a source of trouble to both parties, and if too many lots are rejected the producer risks to loose his share of the market.

It is by this principle that sampling inspection works; its foremost purpose is not to detect lots of poor quality, but to stimulate the production of lots of good quality. If this was not so tightened inspection would be meaningless for it is meant to enforce an improvement in the quality of the lots delivered. In my opinion any sampling standard that has been thoughtfully designed will work satisfactorily. The Philips Standard Sampling System [9, 10], which was considerably simpler than Mil. Std. 105D, was found to be quite effective. The only reason for its discontinuation was, that the use of two different standards for the same purpose created confusion and misunderstanding; and as Philips was forced to use Mil. Std. 105D by many outside suppliers and customers it was the Philips System which had to be given up, much to the regret of many of its users.

Shewhart too was well aware of the need of simple approximate techniques in practice. He proposes the 3σ limits in the following terms (p. 276): 'Experience indicates that t = 3 seems to be an acceptable economic value.' The justification could hardly be simpler, less statistical, and more practical. And there is no mention of normality or a probability level.

Looking at these problems from a more general point of view the purpose of standardizing statistical techniques is, in my opinion, to establish rules for taking routine decisions as they daily occur in industry, commerce and other fields. But here we must clearly distinguish between technical decisions and statistical decisions.

To accept or reject a lot is a technical decision that has important economic consequences; to accept or reject some hypothesis concerning the percent defectives in that lot is a statistical decision, but is from a practical point of view an empty phrase.

The assumption of a random sample, of independence, of the normal or some other distribution, and the adoption of a 5% or 1% significance level must be seen as statistical conventions needed for the derivation of statistical decision rules; which, however, cease to be of importance when these statistical decision rules are converted into technical decision rules. The chief requisite of rules for technical routine decisions is, that the decisions they lead to are accepted without protest or dissension by all parties for whom these decisions have some economic consequences. To this end the rules should preferably be simple, easy to apply, and easy to understand; and, provided it is simple too, the underlying statistical model needs not be more than a fairly crude approximation to the situation envisaged; for routine decisions too much detail conflicts with practical applicability.

To conclude: If we had an equivalent of the Shewhart Medal in Europe I am not sure I would be the first to receive it; there may be others more deserving. I cannot speak for other countries, but as for the Netherlands I would like to mention F.G. Willemze and A.H. Schaafsma, two engineers who introducted QC in the Philips factories in the Netherlands and were the first to write a book in Dutch on 'Modern Quality Control' [10]; and J.D.N. de Fremery, who as executive secretary of the EOQC did a good job in setting that Organisation afoot.

Needless to say that I highly appreciate having been chosen as the Shewhart Medalist for 1979. But if I am the second European to be so honored (L.H.C. Tippett, 1961, was the first) this is, I believe, because owing to the very fortunate circumstances under which I have been working, my contributions to QC have been more conspicuous than those of others. And I hope you will permit me to accept this distinction not only as a tribute to myself, but also as a tribute to my fellow Europeans who might be equally deserving.

References

1. H.C. Hamaker, The theory of sampling inspection plans,
Philips' Technical Review 11, (1950), 260-270.

2. R.A. Fisher, The design of experiments,
Oliver and Boyd, Edinburgh, 1935; many times reprinted.

3. W.G. Cochran and M.G. Cox, Experimental designs.
Wiley and Sons, New York 1950.

4. O.L. Davies and other, The design and analysis of industrial experiments,
Oliver and Boyd, Edinburgh, 1954.

5. W.A. Shewhart, The economic control of quality of manufactured product,

6. H.C. Hamaker, Examples of designed experiments,
Industrial Quality Control, XVII, No. 9, (1961), 1-5.

7. Sir Claus Moser, Statistics and public policy,
Journal of the Royal Statistical Society A, 143, part 1, (1980), 1-33.

8. Standard methods for testing tar and its products, 6th Edition (1967),
Edited by P.V. Watkins, Published by the Standardization of Tar Products Committee, Gomersal, Cleckheaton, Yorks, Great Britain.

9. H.C. Hamaker, J.J.M. Taudin Chabot and F.G. Willemze,
The practical application of sampling inspection plans and tables,
Philips' Technical Review 11 (1950), 362-379.

10. A.H. Schaafsma and F.G. Willemze, Modern kwaliteitsbeleid (Modern Quality Control),
Philips' Technical Library, 1954.