

Are Randomized Double-blind Experiments Everything?

Hampel, Frank

Swiss Federal Institute of Technology (ETH), Dept. of Mathematics

CH-8092 Zurich, Switzerland

E-mail: hampel@stat.math.ethz.ch

1. Introduction and overview

In medical statistics, randomized double-blind experiments are one of the most important tools. In order to test a new drug or treatment, it is applied to a randomly selected subgroup of patients, while the others obtain a comparison treatment or no treatment at all. Single blinding means that the patient does not know what he or she gets, and double blinding means that also the doctor who judges the effect does not know what has been applied. Obviously, this is to avoid any subjective biases, and it is necessary if the results are not clearcut. And the randomizing ensures that “in the long run” any significant differences are due to the treatment effects. Because of these nice properties, some medical statisticians strongly demand that all evidence should only come from such experiments, they consider it a panacea and seem to tend to ignore all other information.

However, apart from this limited view, there are also some problems with this approach. One is obviously that the law of large numbers may not yet be applicable. To avoid this, bigger and bigger experiments are proposed, and/or many smaller experiments are combined in a meta-analysis. One danger is that the quality and homogeneity is much harder to ensure in big experiments, even more so in (tendentiously heterogeneous) parts of a meta-analysis. This may concern such at first trivial-looking aspects as definitions(!) (one of the hardest problems in statistics, as is well-known, e.g., from sample surveys). A small, well-done survey is usually better than a big one, even if one tries hard with the latter. It is also known (also empirically) that a small, carefully taken nonrandom sample may be more informative than a random one of the same size, though one cannot combine the former one to form larger samples because of the bias which remains when the random variability eventually (very slowly) goes down. (There were serious discussions among statisticians about random and systematic designs; and one general advice given is to stratify what can be reasonably measured and randomize over the rest.) I get the impression that in big meta-analyses the individual trials are studied only superficially or not at all (e.g., concerning their quality), though I may be wrong. But one big problem is the commercial one: big trials can only be made with a lot of money (and hence usually big monetary interests) behind them, while small firms or research groups normally just cannot afford them.

In this connection it has to be asked whether big trials (mega-trials) and meta-analyses are at least sufficient to guarantee reliable results. Unfortunately, the answer is no, as a big, commercially important meta-analysis recently showed. Amazingly, it was noted only much too late (why?) that in the “official” meta-analysis, most trials were done by the firm producing the drug under investigation which had of course various means of producing the desired results, such as selecting only the favorable trials and keeping the others secret. And in another recent case, the scientific rules were broken to such an extent that it had to be followed by an apology; but the propagandist damage had been done already.

After noting that randomizing by itself is unfortunately in practice not sufficient for reliable results, it may be asked whether it is necessary (cf. also above). The answer varies. After R.A. Fisher introduced randomizing as a very useful integrated tool to agricultural and similar trials, it may have been J. Neyman who dogmatically insisted on it very broadly. As a student, I was a witness when Neyman tried to introduce a statistical consulting service at UC Berkeley. There came a biologist with

a study about mice, with a lot of what seemed to me most interesting statistical problems. Neyman asked only one question: “Did you randomize?” When the answer was “No,” this was the end of the consulting session. I must stress that there are cases where randomizing is absolutely necessary; for example, I did so in the international hail suppression experiment “Grossversuch IV” (cf. B. Federer et al. 1986). And I think it is fair to say that Neyman, with his attitude of a pure mathematician, made statistics respectable for pure mathematicians. On the other hand, Cuthbert Daniel, one of the best applied statisticians who was also guest professor in Berkeley and who worked mainly with technical data, found randomization generally not necessary and usually even impossible for his kind of data. And John W. Tukey, another most famous applied statistician (with a strong background in mathematics, by the way), was so appalled by the Berkeley type mathematical statistics of that time that he coined the new term “data analysis” for the applied statistics he was doing.

I was quite surprised when, at a workshop on the foundations of statistics, two statisticians from quite different schools (Bayesian, and nonadditive probabilities) independently claimed that in the future those responsible for the use of the Neyman-Pearson statistics in clinical trials may have to pay huge compensations for the avoidable losses caused by the use of this (superficial and often wasteful) type of statistics (for example, by introducing some medicine far too late).

While these are practical considerations, one can also bring philosophical arguments about the “existence” of randomness (which in practice is generally only pseudo-randomness) and about the dangers of superficially imposing pure mathematics onto the real world (“Every word a lie”), cf. Hampel (2002).

There is a lot of informal practical experience which often turns out to be surprisingly accurate (though sometimes the interpretation may be difficult). In particular, there is often a valuable body of medical experience which should not be ignored, even if it cannot be explained. (For example, my then doctor once pushed me to take an MRI test; I still remember the embarrassed face of the specialist when he came with the result: it was an unexpected triviality which fitted much better to my condition, and I am still grateful to the specialist that he admitted they had already many such cases, although these were “scientifically impossible.”)

It must not be forgotten that many of the most important discoveries in medicine were done by single cases, or by the creative analysis of a single haphazard data set, such as the discovery of penicillin, or of the connection between rubella and pregnancy.

Personally, I find unintentional (!!) single pieces of information often most valuable, especially if they have the “ring of truth”, even though they have to be checked back, of course. But they may lead to new theories and sometimes even a new world view (cf. Hampel 2009).

On the other hand, some scientists cling so strongly to their old “background beliefs” that they will do anything to defend them, even if it were not necessary. An amusing example is a slowly descending chain of lights many people saw after a big firework, in a hazy night, about 10 km away. There was a big uproar because it could not be explained, and the organizer of the firework was attacked because he admitted that he simply didn’t know. Finally, the “explanation” was found (and believed!): a chain of ordinary burning candles under balloons from a birthday party (10 km away!). So there was no danger anymore that anybody might believe in UFOs with the little green men. In reality, it was the headlights of the cars coming down from a mountain lookout.

A more serious question occurs when a phenomenon cannot be explained by current scientific theories, as with acupuncture and homeopathy. I shall discuss acupuncture in the later parts of this paper, both describing my own experiences and discussing in detail the criticisms by Singh & Ernst (2009). (I selected this book and concentrated on it because it was written for a broad audience and in part by one of the leading critics of alternative medicine.)

As to homeopathy, I don’t have major personal experiences, and I agree that one has to be very skeptical as it seems to contradict present scientific experience; but one should be openminded in

both directions. After all, some well-established “facts” derived from genetics were recently suddenly false when epigenetics was discovered. Much or most of homeopathic work is said by homeopathic specialists to be of very low quality and also shows unclear or no effects; but when in a meta-analysis the group of studies of the highest scientific quality showed a clear effect, this was without any further scrutinizing dismissed as “humbug” (Goldacre 2010, p. 84f). May be it is; but this is the kind of prejudice that makes candles out of car headlights. It fits to the famous lines of the German poet Christian Morgenstern: “Weil, so schliesst er messerscharf, nicht sein *kann*, was nicht sein *darf*” (from *The Impossible Fact*; in the English translation by Max Knight (Morgenstern-Knight (1963) 1966, p. 35): “for, he reasons pointedly, that which *must* not, *can* not be”).

When just somebody gets well after taking homeopathic pills, this does not mean anything. Even if there are many such cases, it would need a careful statistical analysis. But if, as is claimed, quite a few (though by no means all) chronic cases given up by standard medicine as incurable (“austherapiert”) get well again, this should give rise to serious thoughts. For example, what is the rate of “spontaneous healings” without this or any other alternative treatment in such cases?

Fortunately for the patients, more and more hospitals allow and support the cooperation between standard and alternative doctors, usually against strong resistance in the beginning, but with convincing success. Both groups of methods have their limitations (and some people are so disappointed that they switch to the other extreme), but they often complement each other and in this way are able to do the best for the patients.

2. Are meridians a fiction?

When I obtained the Corgi edition of Singh & Ernst (2009), the first page I opened (after the Index) was p. 107, and the second pages were p. 57f. The reasons were very simple: I was curious about what the authors have to say about the alternative method I have a little personal experience with, namely acupressure; and I was generally curious about how Oetzi the Iceman (whose well-preserved 5000 years old corpse had recently been found on a glacier in the Alps) made it into this book.

I was very surprised. On top of p. 107 (Conclusions, point 1) it is claimed that “there is no evidence at all to demonstrate the existence of Ch’i or meridians.” As this is a basic assumption or belief from which the authors derive many conclusions in large parts of the book, and as it contradicts my own personal experiences, I think it is worthwhile to describe my main experience in as much detail as possible, so everybody can form his or her own opinion about the matter.

Moreover, I was even more surprised when (p. 57f) the authors seemed to accept the claim that 5000 years ago, about 3000 years before the first Chinese book about acupuncture (the Huangdi Neijing) and in a different part of the world, there apparently existed not only detailed knowledge about acupuncture points, but also on which treatments to use for which conditions. As F. Bahr et al. published in *Lancet* in 1999, the body of Oetzi showed many tattoo markings which were precisely the points to be used in acupuncture for his two main ailments (which were found by different investigations). I think it is highly unlikely (though not impossible) that such a sophisticated cultural transfer took place over time and space, between the Alps and China; I find it more plausible that both the methods of treatment developed independently, based on the same empirical observations about the live human body. But then it is practically impossible that the same methods were developed if Ch’i, meridians and acupuncture points were a mere fiction, as the authors claim elsewhere.

Now let me describe the background of my experience, which may be considered relevant for its interpretation. I had learnt some shiatsu and acupressure; I knew the (alleged) course of the meridians and some important points; but I was rather doubtful about some more theoretical claims such that there should be a complete circuit of Ch’i in the 12 ordinary meridians, the more so as the Chinese for some time knew only 11 meridians (Kubiena et al. 1991, p. 32). I was also wondering about the

meridian (“Triple Warmer”) to which belonged no organ, and I could not believe that a smooth flow should make sharp right angles. I had not learnt the alleged sequence of the 12 meridians by heart, but I could reconstruct it mentally in a slow and cumbersome way if needed.

One day, I went for treatment to a Chinese doctor who had also obtained the “regular” Swiss medical doctor FMH. Since I had told him I knew a few acupuncture points, he told me to press at home a particular point (N7 in German, Kidney 7, Fuliu) “as long as possible.” So at home at night, I made myself comfortable in front of the TV and started pressing this point (close to the ankle) with my thumb.

For a long time, nothing happened, but suddenly I noticed a tingling feeling under my thumb which slowly, millimeter by millimeter, crawled upward in my leg, creating a thin tingling line which more and more clearly marked the course of the kidney meridian. The line went up the belly and the breast, and then it stopped at the end of the meridian, as half expected. I felt the whole kidney meridian starting from N7 and thought: “This was it.” But after quite a while, I suddenly felt a tingling in a different part of the breast; it also started growing, mm by mm, to run along the inner arm and to extend all the way to the middle finger. I was totally surprised. The new tingling line (while I still felt the old one) followed the Pericardium. I remembered that I had heard something about a sequence of meridians, and wondered whether the Pericardium could be the next one after the kidney meridian in this sequence; since I did not want to interrupt my pressing of N7 in any way, I tried to reconstruct the circuit mentally instead of getting up and looking it up; and in fact, the Pericardium was the next meridian.

Now I started to get the pattern. After another long pause after the end of the Pericardium had been reached, I felt the tingling also in the ring finger and its extension along the Triple Warmer all the way to the face, and so on. By now, I decided to sit or rather press it through, although it took many hours, until I felt all 12 meridians simultaneously in my body and was back at the beginning of the kidney meridian. Especially amusing was the feeling in my leg, namely like criss-crossing spaghetti. All sensations followed exactly the meridians as described in the books, and even the sharp right angles did occur. Moreover, I now had experienced the complete circuit of the 12 meridians which I had doubted so much.

During later repetitions of the experience, I also found out what happened in between the meridians, as there were always strange gaps between the last point of the previous meridian and the first point of the next one. In between, there was a fine, tender, 3-dimensionally spread-out web in the body, until the Ch’i collected itself again in the concentrated stream of the meridian. I never read or heard anything about these intermezzos.

Pressing other points, I also could feel each one of the 8 “extraordinary” or “wonder” meridians (2 of which are commonly used in acupuncture), but I want to leave it at that here. -

I was now convinced that suitably gifted people all over the world can feel the flow of Ch’i in the meridians, as is demonstrated by the agreement of the descriptions. On the other hand, this gift of perception may be rather rare, as the Chinese for some time knew only 11 meridians.

But someone without this gift and the patience of using it might easily deny the existence of these phenomena if he or she is not willing to trust other persons that do have this gift. The problem here is to find out who is trustworthy; obviously, there are also many false claims in the world. (A nice objective example for the coexistence of right and false claims is an experiment in astronomy (Hunt & Moore 1982, p. 32f) where about 200 persons falsely claimed to be able to see the crescent of the planet Venus close to lower conjunction with the naked eye, while 2 persons actually saw it.) I can only say that I, as a mathematician and scientist, and a rather critical and skeptical, but also openminded person, experienced the whole (outer) meridian system, very much to my surprise, and partly against my own prejudices.

Compare also the story of Jiro Murai (Teeguarden 1978) who in Japan before and around the

middle of the last century (on the way to his death, as he thought) rediscovered the Ch'i and meridians in his body, was completely healed, and developed the healing technique Jin Shin Jitsu.

This book (Teeguarden 1978) also mentions in general form many other cases with this experience.

It is interesting to note that the concept of Ch'i (Ki, Prana, Mana, ...) is known in many different cultures; I once heard that also in Europe, Paracelsus knew it under the name *Munia*. The fact that our most modern technical apparatus cannot measure it (Singh & Ernst, 2009, p. 72) only shows the limitations of these apparatus. Not much more than 100 years ago, many phenomena could not be measured which are now an integrated part of science.

3. Some more remarks on acupuncture

We shall now look at some more arguments against the existence of Ch'i, the meridian system, and the possibility of real (as opposed to placebo) effects in Singh & Ernst (2009).

On page 60, the authors try to discredit the meridian system by claiming that some acupuncturists believe in 14 and others only in 12 meridians. This is a profound misunderstanding. There are 20 meridians (cf., e.g., Kubiena et al. 1991): the 12 "regular" or "ordinary" meridians through which flows the circuit described above, and the 8 "extraordinary" or "wonder" meridians, which are only occasionally used; two (and only two) of these have their own acupuncture points, and if one wants to use them, one has to work with 14 meridians. This may not always be necessary, so that 12 meridians may suffice. But perhaps the sources for the two authors only meant that the regular circuit of Ch'i only flows through the 12 ordinary meridians, which is of course true. (The 12 regular meridians may be likened to rivers, and the 8 other ones to canals between the rivers, which may be used for special effects.) Acupuncture is an art based on a science; different acupuncturists may prefer different suitable points and combinations of points, in the same way in which Western medical doctors may prefer different drugs and treatments.

It is clear that yin and yang (p. 60) always play at least a background role. They may be divided further, e.g., according to the position of the meridian considered in the circuit mentioned above. And the Ch'i can be divided into a number of different forms, depending on origin and purpose. However, I am not so familiar with these forms, so I do not want to comment on them.

The number of ordinary meridians is 12, no matter whether there are or were 12 great rivers in China or not (cf. p. 70, where it is suggested that the number of rivers determined the number of meridians "believed" by the Chinese). And the number of (main) acupuncture points is *about* 361, not necessarily 365, since the definition of some points (e.g. double-points or very weak points) seems somewhat ambiguous. (To equate their number with the number of days of the year seems indeed artificial.)

I do not know much about the history of medicine (p. 70ff, where the lack of human dissection in China is stressed), though I would assume that the Chinese would know the organs from dissecting animals. And in exaggerated reply to the arguments in the book one could say that the Chinese studied the *live* human body very thoroughly, while Western medicine (at least in its beginnings) was based on dead corpses and hence could not discover Ch'i in action. This is, of course, not true anymore, but somehow our whole modern technology still seems to be more geared to analysing the last refinements of dead tissue, even though it could also look at live tissue. (Cf. also above, the end of the previous section.)

So far, I have only discussed the existence of Ch'i and the meridian system. Another question is the effectiveness of a medical treatment based on it. Here, I am not so convinced about generally strong effects. More precisely, for me one of the most interesting questions is when a treatment works and when not. (This holds also for conventional medicine; I know that there are cases where the

medical doctor is fully convinced that a certain treatment must work, and it does not.) And I agree with the authors that some effects may just be placebo effects. (But we must not forget that even the general recognition of placebo effects in modern medicine is rather recent.)

Although the authors admit (p. 105) that for principal reasons it will never be possible to conduct a perfect double-blind randomized acupuncture trial (because the person who sets the needles always knows whether it is a treatment or a placebo, and hence is not blinded), the attempts by various groups to come closer to such trials are impressive. A general idea is to use sham acupuncture in one form or other for a placebo (p. 90, p. 104).

The “false” needle is either inserted at a false place or at a wrong (low) depth. And several groups, including a group by one of the authors, had the ingenious idea to use a retracting needle, like a theatrical dagger (p. 103f). These “great minds” (p. 104) seem to believe that only needles at the right place and the right depth should have an effect (if at all), according to the Chinese prescriptions. But acupressure shows that any pressure applied at the point also has an effect (though perhaps a milder one), and my experience shows that a slightly misplaced pressure point can “attract” the connection to the meridian up to some extent. I do not know how far the misplaced needles in the various experiments were from the ideal points, and whether the body in these cases was still able to build “bridges.” But in the other cases, including the telescopic needles, there was pressure applied by the sham acupuncture at the correct points, and hence it was not or only marginally possible to distinguish between a treatment effect and a placebo effect.

The observed effects of proper treatment and sham treatment might both be placebo effects, but contrary to the belief of the authors, they might (partly or fully) both also be true effects. On p. 106, the authors cite the first results of the German mega-trials which show little or no difference between real and sham acupuncture, from which they conclude that both measure only a placebo effect. Only on p. 112, the authors mention that both treatments (proper and sham) are considerably and significantly better than no treatment. So, without any much more detailed and less prejudiced arguments (remember, the authors believe in the nonexistence of Ch'i, without being able to prove it), there might be a considerable true treatment effect.

(And arguing along the lines of the authors: as long as there is doubt that all effects are only placebo effects, even those who find a pure placebo treatment unethical (p. 113), may still use acupuncture.)

Let us now turn to the Conclusions (p. 106ff) about acupuncture in the book by Singh & Ernst. We have discussed point 1 (p. 107) already extensively, and also the following points. One could bring basic arguments against the usual Western discussion of acupuncture (overall, it seems the Chinese can learn Western culture and thinking much better than the other way around), but I do not want to go into all the cultural misunderstandings, except to mention briefly that a Western assessment of acupuncture is from a Chinese viewpoint negatively biased, because the concepts are not the same. I am, like the authors, against exaggerated claims (and moneymaking with them), but I believe an often mild and sometimes dramatic treatment effect is possible. As to the costs of acupuncture (p. 108), they could be considerably decreased by using acupressure. And neither has the negative side effects of many painkilling drugs (p. 108).

The results of the 640 patients with chronic back pain (p. 109) may well be interpreted like the German mega-trials (see above); but even if a true effect of the acupuncture could be excluded in this case, it still would act as a “powerful placebo.” I think the points about complexity and basic philosophy are still relevant, because of the deeper question: For what does a clinical trial work? What is the same disease, what is the same treatment? The concepts of disease and treatment are different in China and the West. (It may be interesting to assess Western medicine by Chinese methods, if it has not already been done; probably it would be too costly.)

The last (5th) counterpart discussion (p. 110f) tries to make the arguments of the acupuncturists

look ridiculous, but I have a different opinion. As mentioned, I don't have enough information about the misplaced needles, but I don't think that "a needle anywhere to any depth" (quasi at random) would have a therapeutic benefit, as the authors claim in their counterargument. And there is no (or not only a) philosophy of acupuncture, but basically an empirical science, as rational as Western science. I am really surprised that the authors did not take the (hypothetical, in their eyes) acupressure effects of the telescopic needles more serious. And (depending on the points pressed) "a handshake, a tap on the back or scratching an ear" (p. 111) may well have effects. (The poet Robert Louis Stevenson, a very rational and sober man, described towards the end of "In the South Seas" (Edinburgh 1894-98) a Gilbert Islands medicine man who scratched not his ear, but just the brim of his hat softly with a palm twig, with a tremendous effect, though the method may have been different.) And the effect of pressure on the skin may in fact make us ill (p. 111); I know the case of a person pressing a certain point too much, and there are of course many stories about martial arts.

To sum up the parts on acupuncture: The arguments of the authors are not only strongly biased, but also not convincing, and they contradict at least in part clear empirical evidence. –

While I can share the definition of evidence-based medicine as written on p. 36, I think there can be considerable disagreement on what is considered "current best evidence." And while I share the hope that science and scientific method will eventually converge to the Truth (a term that is irritatingly often used in the book, even in cases where its claims are plainly false), we must not forget that the present state of science is only a transient state in its development, and future generations may laugh about many of our present beliefs.

Acknowledgments. I am grateful to V. Hampel, B. Hulliger, U. Rentsch and W. Stahel for various contributions, though the responsibility for the contents of the paper is entirely mine.

REFERENCES

- Federer, B., A. Waldvogel, W. Schmid, H.H. Schiesser, F. Hampel, M. Schweingruber, W. Stahel, J. Bader, J.-F. Mezeix, N. Doras, G. D'Aubigny, G. DerMegreditchian, and D. Vento (1986). Main results of Grossversuch IV. *J. Climate Appl. Meteorology*, 25: 917-957.
- Goldacre, B. (2010). *Die Wissenschaftslüge. Die pseudo-wissenschaftlichen Versprechungen von Medizin, Homöopathie, Pharma- und Kosmetikindustrie*. Fischer Taschenbuch Verlag, Frankfurt am Main.
- Hampel, F. (2002). Some thoughts about classification. In: K. Jajuga, A. Sokolowski, and H.-H. Bock (eds.): *Classification, Clustering, and Data Analysis. Recent Advances and Applications*. Springer, Berlin, p. 5-26.
- Hampel, F. (2009). How can we get new knowledge? In: T. Augustin, F.P.A. Coolen, S. Moral, and M.C.M. Troffaes (eds.): *Proc. Sixth Internat. Symp. on Imprecise Probability: Theories and Applications*. Durham Univ., Durham, UK.
- Hunt, G.E., and P. Moore (1982). *The Planet Venus*. Faber and Faber, London.
- Kubiena, G., A. Meng, E. Petricek, and U. Petricek (1991). *Handbuch der Akupunktur. Der chinesische und der moderne Weg*. Verlag Orac im Verlag Kremayr & Scheriau, Wien.
- Morgenstern- Knight ((1963) 1966). *Christian Morgenstern's Galgenlieder. A Selection*. Translated and with an Introduction by Max Knight (4th printing). University of California Press, Berkeley and Los Angeles.
- Singh, S., and E. Ernst (2009). *Trick or Treatment? Alternative medicine on trial*. Corgi edition, London.
- Teeguarden, I.M. (1978). *Acupressure Way of Health*. German translation (1989): *Die Kunst der mitfühlenden Berührung*. Droemer (Knaur Nachf.), München.