

Theory and clinical use of probabilities in Germany after Gavarret. Part 2: assessments of the state-of-the-evaluative art

Ulrich Tröhler

Institute of Social and Preventive Medicine, University of Bern, CH-3012 Bern, Switzerland

Corresponding author: Ulrich Tröhler. Email: ulrich.troehler@ispm.unibe.ch

This article continues our series on probabilistic thinking and the evaluation of therapies, 1700–1900

The series will continue to appear as separate articles in forthcoming issues of JRSM

Two comments from outside and inside Germany

The contributions made by German writers discussed in the previous article in this series had not passed unnoticed. A conscientious young medical historian doctor sensed that issues of evaluation and probability were in the air: **Julius Petersen** (b.1840) of Copenhagen gave a substantiated description of the situation in his *Hauptmomente in der geschichtlichen Entwicklung der medicinischen Therapie* (Key moments in the historical development of medical therapy, 1877). In 29 pages, he dwelled on Poisson, Gavarret, Louis, Wunderlich and others, and on the numerical method. He quoted Gavarret, rightly (and at length), saying that there was much loose verbiage about probability, whereas only the calculus of probabilities could really help to estimate the worth of mean values (averages). Albeit still far from being perfect, this method was important for future developments (p. 179). Petersen clarified the confusion between Benecke and Vierordt: the former explained the effect of a cure, the latter thought that it was being demonstrated statistically. Again, this was the old debate between rationalism and empiricism. I think we can trust him when he said in the mid-1870s that in France, England and Germany polypharmacy and the *post hoc ergo propter hoc* fallacy prevailed, Louis's principles were lost sight of, but some British followers of Bacon were still eclectics and indulged in common sense (Petersen, 1877).

A young German insider engaged in an overview of the methods available in clinical research, particularly in therapeutics. **Friedrich Martius** (b.1850), while a military doctor and later an assistant physician at the Berlin Charité University Clinics, published two lengthy articles on the subject: *Die*

Principien der wissenschaftlichen Forschung in der Therapie (The principles of scientific research in therapy, 21 pages, Martius, 1878) and the even more erudite *Die numerische Methode (Statistik und Wahrscheinlichkeitsrechnung) mit besonderer Berücksichtigung ihrer Anwendung auf die Medicin* (The numerical method [statistics and calculus of probabilities] with special reference to its application in medicine, 41 pages, Martius, 1881). Later he became professor of internal medicine at Rostock and, typically, did not publish any longer on the subject.

As Oesterlen and Schweig had done some 30 years previously, Martius begun by clarifying the confused terminology. He analysed the French and German works (Radicke was not mentioned!) by putting them in the wider historical context of the theories of cognition: the term 'induction', he wrote, was often used without distinguishing whether 'logical', 'numerical' or 'experimental' induction was meant. Numerical induction he understood as being based on statistics and probability calculus. For, although they had been developed separately, statistics and the calculus of probabilities could be summarised under the common term 'numerical method'. Indeed, they complemented each other

necessarily and happily [...]. The calculus of probabilities needs for its application materials collected according to the strict rules of statistics and the latter, without the calculus, would not [...] always find its critical utilization and the elaboration of which it is capable (Transl. from Martius, 1881, p. 349).

Consequently, aware of the Paris deliberations of 1835, 'which have since acquired fame', Martius regretted that Gavarret, in his enthusiasm, had

disparaged statistics in favour of the calculus of probabilities (Martius, 1878, p. 1185; Martius, 1881, p. 243). Gavarret was following Laplace, who had declared that all knowledge was based upon probability (Martius, 1881, pp. 347–348). This was obviously not true. One had just to think of anatomy. Of course, one had to be familiar with basic mathematical principles to be able to discuss the appropriateness of conclusions arrived at by the ‘numerical method’, for it was often used simply to prove what one wanted to prove (Martius, 1881, pp. 338–339). But doctors’ continuing aversion to the mathematical approach stemmed from their ‘*mathematical incapability*’ (Transl. from Martius, 1881, p. 346).

As to the fundamentals on which the true ‘numerical method’ rested, Martius identified some open questions. He wrote that Gavarret’s famous probability ratio of 212/213 (99.5%) had been chosen arbitrarily on the basis of Poisson’s formulas, as had Hirschberg’s simplification by fixing a ratio of 9/10. This showed the arbitrariness and unreliability of such ratios and of the whole process: which of these haphazardly proposed probabilities excluded the hazard? Here was Martius’s answer:

To remedy this undeniable drawback, Liebermeister now intends - by dropping Poisson’s formulas completely and, departing from other preconditions - to develop new formulas, that can serve to calculate, with certitude and precision, the degree of probability with which the hazard is excluded. And this for any [...] observational material, be it ever so small [...] provided] the comparability of the cases, this eternal crux of all statistical data collections, can be demonstrated (Transl. from Martius, 1881, pp. 375–376).

If Liebermeister’s formulas were more easily applicable, they were also ‘more unscientific’ than Poisson’s, for

...they completely neglect the law of large numbers, and they offer nothing but the reflection, expressed in numbers of probability, that when one ignores the nature of the process in course the best thing to do is, *faute de mieux*, to stick to true, existing successes (Trans. from Martius, 1881, p. 376).

Thus, like Claude Bernard had done, Martius made clear that progress in identifying constant, determined causal relations required induction through laboratory experiments, not the numerical method. This did not mean that he proposed neglecting statistics. On the contrary, through mass observation and reliable assessment of treatment successes, the

probability of obtaining important indications for practical action increased (Martius, 1881). Yet such probabilities were not, as Gavarret had deemed in his first enthusiasm, the ripest fruit of modern thought, or

the highest and most consummate level of all research methods usable in therapeutics. Rather it is and remains a makeshift, albeit a very important one [...], that is undoubtedly worth an even deeper foundation and more extensive application (Transl. from Martius, 1878, p. 1185).

These were clever insights, and such efforts would effectively be made in the 20th century. But before, new difficulties and, consequently, new desiderata were recognised by two practitioners from Breslau (now Wrocław, Poland), **Alfred Ephraim** and **Ottomar Rosenbach**.

When reading these two testimonies one will realise that formal, mathematical probabilistic reasoning had clearly made an impact on its authors. But on the practical side, the consequences were limited, while on the theoretical, new requirements for scientific evaluation were identified – for the new century.

Towards the *fin-de-siècle*

Unsystematically compiled statistics continued to be worked up and interpreted in the manner of shopkeepers, and without additionally calculating probabilities despite what Martius had called for. So, **Alfred Ephraim** (b.1863) felt once more – just as Wunderlich had 50 years earlier – that therapeutics were chaotic. In his *Über die Bedeutung der statistischen Methode für die Medicin* (On the significance of the statistical method for medicine, Ephraim, 1893), he saw the reason for this desolate state in the oblivion of the provisions stipulated time and again since the Paris discussions of decades ago. The methodology of clinical evaluation was eclipsed by new technical methods of examination. Ephraim noted that a recent discussion between two eminent German physicians made clear that the numerical method continued to have both detractors and supporters. He claimed that the *reversal* of previously statistically founded claims did not help to convince the medical world of the value of such work (Ephraim, 1893, pp. 695–696). That is why Ephraim answered the two eternal questions of the statistical endeavour – (i) what was to be counted? and (ii) how many cases should be counted – by recalling the precepts established by Gavarret (p. 712). But while these theoretical difficulties could be dealt with, one should not overlook the practical ones;

and here he enumerated three new criteria for solid comparisons:

- Diagnoses must have been made using the same diagnostic methods, which is particularly difficult when cases are assembled from various sources.
- Adherence to treatments must be strictly observed.
- Trials of treatment should be conducted over sufficient duration.

Moreover, the quest for untreated cases for the equally necessary comparisons was not new. But, in practice, they were difficult to find. If lack of treatment seemed inhuman, it could be justified because most treatments had actually not been demonstrated to be useful. 'Non-adherence [to these precepts] was being seen every day and lead to delusive therapeutic-statistical conclusions' (p. 715), but they were as difficult to fulfil as they were indispensable.

Ephraim concluded that those who deemed these requirements insurmountable must be aware that they are renouncing trustworthy therapeutic knowledge. He noted that 'to substantiate the efficacy of mercury in syphilis, of quinine in malaria... one might perhaps not need statistics' (p. 711). Yet, reliable identification of less dramatic treatment effects could only be assured by the results of statistical research. However, he did not mention the calculation of probabilities as a complementary method of evaluation, thus once more ignoring terminological precision.

Ottomar Rosenbach (b.1851) had worked since 1874 as a hospital physician at Breslau. By 1896 he had resigned his position as chief of the medical department and his associate-professorship and retired to private practice in Berlin, but continued publishing. It is probable that he knew young Ephraim since they had lived in Breslau at the same time. Certainly, he knew the latter's methodological work for he extended it in two publications, a lengthy 10-page one in three parts on *Serumtherapie und Statistik* (Serum therapy and statistics, Rosenbach, 1896) and a shorter paper on *Der Kampf um die Zahl in der medizinischen Wissenschaft* (The fight about numbers in medical science, Rosenbach, 1899). Both were conspicuously published in the *Münchener Medicinische Wochenschrift* (The Munich Medical Weekly).

Like Ephraim, Martius and many others before them, Rosenbach criticised the misuse of therapeutic statistics:

Although everybody now knows that small numbers prove absolutely nothing, although everyone knows [...Poisson's] law of large numbers, yet people

preferentially use small numbers, and even many of those who with aplomb only exploit large numbers are in error about their bearing in that it is not the large numbers as such that matter, but the circumstances [over time] in which they are generated (Transl. from Rosenbach, 1896, p. 913).

He further emphasised the arbitrarily defined, often inadequate duration of trials, the failure to use modern diagnostic criteria (for example, the use of clinical signs *and* bacteriology), and the differences among cases (an issue that had already been proclaimed innumerable times).

As a consequence, statisticians' concentration on the *Genesungsquotient* (recovery rate) was misleading, since both numerator (number of cures) and denominator (number of diseased) were often based on variable and subjective criteria. In short, these statistics served only to reinforce preconceived opinions and frequently, when there was no comparison group, to fall into the trap of the *post-hoc-ergo-propter hoc* fallacy (pp. 912–913).

As new elements, he drew attention to bias mechanisms, namely:

- The selection of cases by enthusiasts who, with what they refer to cynically as 'scientific thoroughness', eliminate all unsuitable cases so that, under the new method, deaths must, in reality, no longer occur. (That they still happen is, by the way,... always the fault of unhappy external circumstances, never imputable to the procedure.... Or it is the impossibility of using the panacea sufficiently promptly).
- The historical insight that this procedure of unevenly distributing light [on successful cases] and shadow [on failures] has repeated itself in the history of medicine countless times, and it never loses its impression on credulous minds who do not want, or are unable to understand that highly astounding results can be brought about by the simple 'slight-of-hand' (legerdemain) of a new scientific definition (pp. 912–913).
- A historical comparison was only admissible if the forms of an epidemic remained essentially unchanged over the years. In the case of diphtheria, for example, where Behring's serum-therapy had been introduced since 1893, he demanded 'that one should try once again to obtain a large series of observations [of patients treated] without serum-therapy' over many years (Transl. from Rosenbach, 1899, p. 256).

Of course, many people did not understand or like Rosenbach's method-based objections. And maybe

they did not like him: had he not, while still an aspiring *Privatdozent*, written quite aggressively in a book on the *Foundations, Duties and Limits of Therapeutics* (Grundlagen, Aufgaben und Grenzen der Therapie, Rosenbach, 1891):

Statistics – what would they not have sanctioned in the hands of able arrangers. [And later] The history of medicine furnishes enough examples of friends and foes fighting with equal obstinacy and equal certainty for a dogma established on the basis of such contradictory [statistical] results (Transl. from Rosenbach, 1891, pp. 66,183).

Even when in private practice, Rosenbach indefatigably continued responding to his detractors, writing critically in the *Zeitschrift für klinische Medizin* (Journal of Clinical Medicine) on methodological problems, right up to his death in 1907. In his last paper on this issue – *Die Diagnose als ätiologischer Factor* (Diagnosis as aetiological factor, Rosenbach, 1905) – Rosenbach adduced yet another new criterion for a valuable experiment: the *method of alternation*. Returning to the serum-therapy of diphtheria, he repeated the need for the *experimentum crucis* (the *decisive experiment*), namely,

always to treat one case with and the next without the promoted medicine, whether the medicine is tested in all places at the same time, or in different places one after the other. And of course, this holds not only for the treatment of diphtheria (Transl. from Rosenbach, 1905, p. 233).

If one did not want to, or could not perform this process of evaluation, which Rosenbach felt was easy to carry out, one deprived oneself straightaway of the possibility of doing scientific research (Rosenbach, 1905).¹

Germany by 1900

According to Petersen and many of his contemporaries (and later historians), the overall situation in Germany in the 19th century may have been similar in practice to that in France and Britain: if considered at all, the actuarial method of counting and statistical analysis prevailed. In particular, however, Louis's numerical method as applied to the evaluation of

therapies was also constructively criticised right from the beginning, and appropriate efforts to ameliorate it were made. In the end, selection bias, making results positive by changing criteria and quantification of preconceived ideas were decried as misuses (Rosenbach, 1896). The notion of probability was understood by many clinicians, and a few of them actually struggled to apply formal mathematical probability and its consequences throughout the whole second half of the 19th century. The influence of Poisson and particularly of his medical pupil, Gavarret, was pivotal.

Declarations

Competing interests: None declared.

Funding: None declared.

Ethics approval: Not applicable.

Guarantor: UT.

Contributorship: Sole authorship.

Acknowledgements: My heartfelt thanks to: Iain Chalmers, without whose unflinching encouragement, gentle whip, intellectual and unrenounceable practical help over the years, I would neither have begun nor ever terminated this work; Thomas Schlich, who critically and helpfully read all previous versions; Robert Matthews, whose help with mathematical matters was very welcome; Brigitte Wanner and Christian Wyniger of the Institute of Social and Preventive Medicine, Bern, who helped me, together with Patricia Atkinson, Oxford, with ever so many IT technicalities; my wife, Marie Claude, whose patient love is not probably, but absolutely true.

Provenance: Invited contribution from the James Lind Library.^a

Supplementary file: The reference listed below is chosen as essential to the reading of the article. However, the full list of primary and secondary references is available online both on the Journals's website as supplementary material, and with the original publication at <https://www.jameslindlibrary.org/articles/probabilistic-thinking-and-the-evaluation-of-therapies-1700-1900/>. Except when otherwise mentioned, translations into English are author's own.

Reference

1. Chalmers I, Dukan E, Podolsky SH and Smith D. The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *JLL Bulletin: Commentaries on the history of treatment evaluation*. See www.jameslindlibrary.org/articles/the-advent-of-fair-treatment-allocation-schedules-in-clinical-trials-during-the-19th-and-early-20th-centuries/ (last checked 7 October 2020).