

The ‘personal equation’ as observer bias, and proposed methods to contain it in Anglo-American medicine

Rory Brinkmann¹ and Scott H Podolsky^{1,2}

¹Harvard Medical School, 25 Shattuck St, Boston, MA 02115, USA

²Center for the History of Medicine, Countway Medical Library, Boston, MA, USA

Corresponding author: Scott H Podolsky. Email: Scott_Podolsky@hms.harvard.edu

Introduction

Arising as a concept for differences in astronomers’ observations in early 19th-century Europe, the ‘personal equation’ is a crucial piece of the pre-history of what would later be technically termed ‘observer bias.’ The term ‘personal equation’ spread into a variety of fields, including medicine, where it was used widely and variously from the late 19th century to the middle of the 20th century.¹ We have elsewhere described the complexities of the use of the term in Anglo-American medicine between the mid-19th and mid-20th centuries, which reflected evolving concerns over the perceived art and science of medicine.²

A principal use of the term ‘personal equation’ reflected concern about observer bias. It thus serves as a useful marker for examining the variety of methods invoked to reduce or remove bias and so promote fair assessments. Medical professionals adopted the ‘personal equation’ term to denote such bias in many types of observations and in many different facets of medicine. These included assessments of symptoms and physical examinations, laboratory data, emerging technologies (such as X-ray), diagnosis and classification of diseases, and estimates of therapeutic effects.

The sources of observer bias associated with the ‘personal equation’ were manifold, as John Shaw Billings suggested in 1886:

Almost all men suppose they think scientifically upon all subjects; but, as a matter of fact, the number of persons who are so free from personal equation due to heredity, to early associations, to emotions of various kinds, or to temporary disorder of the digestive or nervous machinery that their mental vision is at all time achromatic and not astigmatic, is very small indeed. (Billings,³ p. 561)

Concerned by the potential of the ‘personal equation’ to erode scientific objectivity, members of the medical

community used a range of methods to identify its presence among observers and to curtail its detrimental influence. Drawing chiefly from research including nearly every usage of the term in the *New England Journal of Medicine*, *JAMA*, *Lancet*, and the *British Medical Journal*, we provide a schematic categorisation of attempts – both in practice and aspirational – to curtail the effects of the ‘personal equation’ of those making observations in American and British medical communities. These sometimes heralded and sometimes diverged from current approaches to limiting observer bias in medicine.

Observers – Numbers and arrangement

Controlling the number and arrangement of observers was an oft-proposed method of limiting the personal equation, though authors differed about how this could be done. Some argued in favour of limiting observations to those of a single observer. While this may be counter-intuitive to 21st-century readers, many authors claimed that having multiple observers risked mixing multiple ‘personal equations’, which could likewise mix the impacts of variation on observations and thereby make it difficult to extract meaningful knowledge. In a study of the Wassermann test in a maternity hospital, for example, one author tried to reassure his readers of the integrity of his data by stating that ‘eighty-seven per cent of the laboratory work was performed by the same technician, thus largely eliminating the personal equation’ (Belding and Adams,⁴ p. 816).

By contrast, others argued in favour of using multiple observers to limit the impact of individual personal equations. This could take the form of observers of equal skill or status crosschecking their observations and then reaching a consensus or deferring to an authoritative observer. One group of authors, for example, claimed in their study of diphtheria that ‘the personal equation has been eliminated

by three persons making the examinations with checking of results' (Geiger et al.,⁵ p. 645).

Another researcher, who had examined an association between the differences in blood pressure readings between different arms and aortic aneurysm, tried to eliminate 'as far as possible' his 'personal equation' through cross-checking his diagnoses of aneurysm with assessments made by other clinicians (Williamson,⁶ p. 1516). Other researchers used a more hierarchical approach, as when one author sought to bolster his results by stating that 'in order to remove the personal equation, Dr. P. Challis Bartlett, who for three years was superintendent of the Turland State Sanatorium, has kindly gone over the records' (Pratt,⁷ p. 15).

A variation on this theme entailed comparing or combining results gathered independently by different observers. In a study of body posture and body mechanics among first-year students at Harvard, for example, Lloyd Brown noted that, among physicians placing students into one of four graded categories, 'the grading... was remarkably uniform and, while there was undoubtedly individual variation, the factor of personal equation seems to have been very slight' (Brown,⁸ p. 653). Such an approach could extend to a hope that individual variation would be diluted by still more observers. At the end of the 19th century, this ethos underpinned efforts at large-scale, medical society-driven 'collective investigations'.⁹ Along these lines, one contributor had addressed the Colorado State Medical Society in 1889 about collective investigations of the effects of climate on tuberculosis:

To relieve [the investigations] from the element of the personal equation which an individual's writing must always bear, this Society voted last year to entrust a consideration of this question to a 'Committee of Collective Investigation', which should have power to solicit reports from individual members of this Society. (Fisk,¹⁰ p. 173)

Standardisation and emerging technologies

Many medical authors claimed that standardising methods of data acquisition could reduce the effects of personal equations. Such standardisation, reflecting 19th-century aspirations towards a 'mechanical objectivity',¹¹ could cover the sequence and timing of laboratory steps, classification schemes, and procedural rules. Thus, while discussing leucocytosis as an indicator of pneumonia, Richard Cabot noted that 'in order that the influence of the personal equation might be as nearly as possible the same in all cases, an

exactly identical technique [of drawing and preparing the blood and enumerating the cells] was used in all' (Cabot,¹² p. 117). To support the rigor of standardisation, authors could also hold that training and experience in particular methods further limited the effects of the personal equation (Anon,¹³ p. 79).

The advent of new technologies was frequently championed as means to check the personal equation. In an 1881 address, Billings referred to this hope for medical devices when he stated that:

the balance and the galvanometer, the microscope and the pendulum, the camera, the sphygmograph and the thermometer are some of the means by which investigators, at the bedside and in the laboratory, are seeking to obtain records which shall be independent of their own sensations or personal equations; which shall be taken and used as expressing not opinions, but facts. (Billings,¹⁴ p. 270)

Contributors invoking the personal equation and further aspiring to mechanical objectivity welcomed various medical instruments as 'constant,' 'uniform,' or 'automatic' (Herschell,¹⁵ p. 460, Oliver,¹⁶ p. 1542, Oliver,¹⁷ pp. 1702, 1703, 1704, and Anon,¹⁸ p. 1472): they frequently drew sharp distinctions between knowledge derived from mechanical devices and other methods ostensibly more susceptible to the effects of personal equations, characterising the latter as opinion, or as 'founded on sand' (Austin,¹⁹ p. 1465).

Nevertheless, many also recognised that interpretation of the outputs of medical devices, ranging from sphygmomanometers to X-rays to electrocardiograms, were not immune to the influence of the personal equation. As late as 1947, a *JAMA* editorialist commenting on inter-individual and intra-individual variation in the reading of chest X-rays continued to point to the importance of the "'personal equation" in the interpretation of a chest roentgenogram.' In line with the implementation of blinded chest X-ray assessments in the MRC trial of streptomycin at the same time,²⁰ he warned that 'there has been a tendency to assume that roentgenology is an exact science and that the objectivity of the medium defied error. Complacency has been a consequence of such assumption' (Anon,²¹ pp. 399–400).

Blinding

Seemingly independent of one another yet each invoking the personal equation, several authors on both sides of the Atlantic turned to a range of methods that would later come to be termed 'blinding' (sometimes 'masking'). In attempting to offset

suggestion and bias, they carried forward variants of a methodology that had been periodically invoked for centuries.^{22,23} Some researchers invoking the 'personal equation' blinded themselves to patient identifiers or conditions. In 1911, for example, authors seeking to assess the different forms of leukocytes in pulmonary tuberculosis attempted 'to eliminate the personal equation as much as possible' by requiring that the 'one who examined the blood knew nothing about the patients, or what they were getting, or how they were affected, or when they began or ended treatment' (Solis-Cohen and Strickler,²⁴ pp. 564–565). Analogously, blinding was also proposed within medical education. In France, a new policy was implemented whereby 'the examiner [would be made] ignorant of the identity of the examinee' and thus limit the effects of the personal equation during grading (Anon,²⁵ p. 809).

Researchers used several measures in attempts to blind themselves to influences on the measurements they were making in real-time. Investigators examining the diurnal variation in the haemoglobin content of blood used a Duboscq colorimeter because:

it leaves the observer in absolute ignorance of the numerical reading until he has finally matched the colour [to the comparison solution], and therefore eliminates the personal equation, a factor of the greatest importance where minute changes have to be ascertained. (Dreyer et al.,²⁶ p. 589)

Another researcher examining tobacco amblyopia devised a method to blind himself to his current and previous measurements of patients' visual fields.²⁷

Others would similarly blind themselves and their patients to the results of previous measurements. In a study assessing the frequency of diseases in different populations, the tabulator took 'great pains... to avoid errors due to the personal equation,' by remaining blinded to the project's results until all of the data had been collected. It was thus

impossible to form any estimate of how [the tabulated results] were coming out until the research was finished and the totals were added up. It was thus impossible for the observer to push or bend the figures in the direction of any theory of his own. (Cabot,¹² p. 117)

Another study, mapping cutaneous hyperalgesia using pin pricking, devised a procedure such that both observers and patients would avert their eyes to avoid being swayed by prior mappings of the same area (Anon,²⁸ p. 33). Observers invoking the

personal equation even defended themselves against the bias future information could have on their observations, arguing that observations should be recorded and so fixed at the moment they are made instead of after the consideration of additional datapoints that may distort their interpretation or documentation. Discussing his own physical examination practices, for example, a clinician argued that a physician 'should record his observations at the time he makes them', before his 'opinion can be influenced by additional and possibly contradictory evidence' (Pratt,²⁹ p. 523).

Researchers also used blinding methods to remove the personal equation from attempts to settle academic disputes. In an assessment of the accuracy of percussion of the heart as a measurement of the Nauheim (bath) treatment of heart disease, a critical author encouraged his reader to demonstrate to himself that the personal equation affected heart percussion, instructing him to 'blindfold himself and make out upon a given case the upper limit of relative cardiac dulness [sic], marking it upon the surface of the chest with an aniline pencil' and then repeat the process, upon which he would find that 'the result is a series of lines at short distances from each other upon the chest, some of them intersecting others' (Herschell,³⁰ pp. 413–414).

Authors also considered blinding patients and/or researchers to limit the personal equation in assessments of therapy. Invoking patient blinding, one researcher argued that 'to properly test a drug or method of treatment it is well to give no intimation of the effects expected' (Anon,³¹ p. 86) because patients, with their own subjective personal equations, could be 'very impressionable and amenable to suggestive therapeutics.' Invoking researcher blinding, in 1913, Michigan's AW Hewlett noted specifically with respect to therapeutic trials:

The personal equations of different observers, the tendency to bias, differences in the modes of administration, in the doses employed, and in the cases selected for treatment, all tend to obscure the significance of reported results. In order to obtain trustworthy data, it is necessary that a considerable number of observations on patients should be made under considerations which eliminate personal bias and reduce to minimum the errors inherent in statistics. (Hewlett,³² pp. 319–321)

The American Medical Association's Council on Pharmacy and Chemistry supported Hewlett's controlled investigation of natural versus synthetic sodium salicylate for the treatment of fever, pain, and delirium. This had entailed supplying the

remedies in coded boxes to 82 investigators, keeping them ignorant regarding which remedy each box contained, and ultimately finding that the two remedies were indistinguishable.³²

Control groups and random allocation

As the Hewlett example suggests, in addition to blinding, certain authors suggested or employed methods that separated participants into control groups to limit the personal equation in the rendering of comparisons and assessments of causality or efficacy. In an evaluation of tuberculosis statistics, an author invoking ‘the statistical method’ advocated ‘isolating and recording control cases’ to ‘eliminate to some extent the ‘personal equation’ of the observer’ and so better characterise the course of the disease (Clark,³³ p. 1693). Control groups were also invoked in this sense to assess therapeutic effects. In a discussion of antistreptococcal serum, for example, one investigator critical of the current state of research on the topic and the degree to which the ‘personal equations’ of investigators had gone unchecked, argued that investigators should ‘compare long series of cases with and without the given treatment under otherwise like surroundings’ (Cotton,³⁴ p. 107). Another ‘personal equation’-invoking investigator employed control cases (not alternated, it seems) to assess the effectiveness of several vaccines against post-surgical sepsis (Goadby,³⁵ pp. 589–592).

Control groups could also be created by those referencing the personal equation through the systematic, prospective random or alternate allocation of patients to treatment and non-treatment groups. Investigators invoked the personal equation in the very first line of their report detailing the effects of ‘convalescent serum in the treatment of preparalytic poliomyelitis.’ They designed their study to limit the personal equation by treating alternate patients with the serum; however, because family physicians frequently demanded that serum be used, many more patients were treated than not (Fischer,³⁶ p. 482).

In a discussion of research about the effectiveness of out-patient medical care, another author held that the only way to answer the question scientifically and eliminate the personal equation was to ‘make a definite study of a number of individual patients selected at random’ (Davis,³⁷ p. 916). The Hewlett study cited above offset both the ‘personal equation’ as observer bias and the ‘personal equation’ as ‘cases selected for treatment’ and ‘as modes of administration’ of remedy, which is to say the variability of patients and their treatments. It offset observer bias by blinding clinician-evaluators as to which remedy they had

in fact employed in each case, while the variability of the patients and their treatments was offset by the random allocation of patients to various treatment groups, as each investigator was given five boxes with one of the two remedies being studied, and five with the other. As we have noted previously,² Hewlett’s³² use of the term personal equation and the actions associated with it served as a bridge to 20th-century attempts to add blinding to random allocation as key features of fair comparisons in assessing the effects of treatment.

Conclusion

Methodologies to curtail observer bias and ensure fair comparisons are cornerstones of 21st century medicine. Therapeutic assessments rely upon random allocation to comparison groups and blinded outcome assessment. The 1948 British Medical Research Council’s trial of streptomycin²⁰ is frequently considered a watershed in medical research study design, but as several authors have previously noted, each of the methods the trial employed have histories of their own that predate the landmark study.³⁸ Attempts to limit the distorting effects of the personal equation are an important part of this rich history. Nevertheless, it would be a mistake to understand attempts to curtail the personal equation solely in a teleological fashion in which authors gradually anticipated the methods in the British Medical Research Council’s report and 20th-century medicine practices more broadly. Instead, attempts to limit the personal equation as observer bias were eclectic, both temporally and methodologically. In this way, responses to the personal equation reflect the United States and British medical communities being in flux across the late 19th and early 20th centuries, striving for scientific objectivity but still lacking a consensus about how to reach that goal.

Declarations

Competing Interests: None declared.

Funding: None declared.

Ethics approval: Not applicable.

Guarantor: SHP.

Contributorship: RB and SHP both participated in all parts of study development and writing of the manuscript.

Acknowledgements: The authors thank Andrew Turner for his valuable research and insights during the construction of their prior paper on the history of the personal equation, and thank Sir Iain Chalmers for his ongoing and thoughtful feedback on the paper and the topic.

Provenance: Invited article from the James Lind Library.

References

1. Canales J. *A tenth of a second: a history*. Chicago: University of Chicago Press, 2009.
2. Brinkmann R, Turner A and Podolsky SH. The rise and fall of the “Personal Equation” in American and British medicine, 1855–1952. *Perspect Biol Med* 2019; 62: 41–71.
3. Billings JS. Scientific men and their duties. *Boston Med Surg J* 1886; 115: 561–565.
4. Belding DL and Adams CB. The Wassermann Test – Wassermann tests in a Boston maternity hospital. *Boston Med Surg J* 1922; 187: 815–821.
5. Geiger JC, Kelly F and Bathgate V. Diphtheria carriers. *JAMA* 1916; 66: 645–646.
6. Williamson OK. The value of blood-pressure determination in the diagnosis of aneurysm of the thoracic aorta. *Lancet* 1907; 170: 1516–1519.
7. Pratt J. Results obtained by the class method of home treatment in pulmonary tuberculosis during a period of ten years. *Boston Med Surg J* 1917; 176: 13–15.
8. Brown LT. Bodily mechanics and medicine. *Boston Med Surg J* 1920; 182: 649–655.
9. Marks H. ‘Until the sun of science... the true Apollo of medicine has arisen’: collective investigation in Britain and America, 1880–1910. *Med Hist* 2006; 50: 147–166.
10. Fisk SA. The effect of the climate of Colorado upon phthisis pulmonalis, as shown by the analysis of one hundred recorded cases. *Boston Med Surg J* 1889; 121: 173–177.
11. Daston L and Galison P. *Objectivity*. New York: Zone Books, 2007.
12. Cabot RC. Leukocytosis as an element in the prognosis of pneumonia. *Boston Med Surg J* 1893; 129: 117–118.
13. Anon. The Association of Clinical Pathologists. *Lancet* 1933; 222: 78–79.
14. Billings JS. Address on our medical literature. *Lancet* 1881; 118: 265–270.
15. Herschell G. Notes on the treatment of heart disease by mechanically-resisted movements. *Lancet* 1896; 148: 460–461.
16. Oliver G. The Croonian Lectures: a contribution to the study of the blood and the circulation. *Lecture I*. *Lancet* 1896; 147: 1541–1547.
17. Oliver G. The Croonian Lectures: a contribution to the study of the blood and the circulation. *Lecture III*. *Lancet* 1896; 147: 1699–1706.
18. Anon. The pulse-rate and arterial tension in the newborn infant. *Lancet* 1913; 181: 1472.
19. Austin AE. Progress in gastroenterology for 1927. *Boston Med Surg J* 1928; 197: 1464–1469.
20. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948; 2: 769–782.
21. Anon. The ‘personal equation’ in the interpretation of a chest roentgenogram. *JAMA* 1947; 133: 399–400.
22. Kaptchuk T. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med* 1998; 72: 389–433.
23. Kaptchuk TJ. A brief history of the evolution of methods to control observer biases in tests of treatments. *JLL Bulletin: commentaries on the history of treatment evaluation*, 2011. See www.jameslindlibrary.org/articles/a-brief-history-of-the-evolution-of-methods-to-control-of-observer-biases-in-tests-of-treatments/ (last checked 21 September 2021).
24. Solis-Cohen SA and Stickler A. The effect produced by some therapeutic measures on the different forms of leucocytes in pulmonary tuberculosis. *Boston Med Surg J* 1911; 165: 563–568.
25. Anon. France. *BMJ* 1932; 17: 809.
26. Dreyer G, Bazett HC and Pierce HF. Diurnal variations in the haemoglobin content of the blood. *Lancet* 1920; 196: 588–591.
27. Harman NB. The visual fields in tobacco amblyopia. *Lancet* 1904; 164: 821–822.
28. Anon. An epitome of current medical literature. *BMJ* 1909; 2: 33.
29. Pratt J. The physical examination in pulmonary tuberculosis. *Boston Med Surg J* 1918; 178: 519–527.
30. Herschell G. Critical remarks upon the nauheim treatment of heart disease. *Lancet* 1896; 147: 413–415.
31. Anon. Reports of societies. *Boston Med Surg J* 1905; 153: 85–86.
32. Hewlett AW. Clinical effects of “natural” and “synthetic” sodium salicylate. *JAMA* 1913; 61: 319–321.
33. Clark H. Tuberculosis statistics. *Lancet* 1913; 182: 1693–1696.
34. Cotton FJ. The present status of the antistreptococcal serum. *Boston Med Surg J* 1899; 140: 105–109.
35. Goadby K. An inquiry into the natural history of septic wounds. *Lancet* 1916; 188: 585–595.
36. Fischer A. Human convalescent serum in the treatment of preparalytic poliomyelitis. *Am J Dis Child* 1934; 48: 481–501.
37. Davis M. Efficiency tests of out-patient work. *Boston Med Surg J* 1912; 166: 915–921.
38. Chalmers I, Dukan E, Podolsky SH and Davey Smith G. The advent of fair treatment allocation schedules in clinical trials during the 19th and early 20th centuries. *J R Soc Med* 2012; 105: 221–227.