

# The (Harry) Gold standard: angina, suggestion and the path to the ‘double blind’ test and clinical pharmacology. Part 2: the path to the ‘double blind’

**Scott H Podolsky**

Harvard Medical School, Boston, MA, 02115, USA

**Corresponding author:** Scott H Podolsky. Email: Scott\_Podolsky@hms.harvard.edu

As detailed in Part 1, the complexities of anginal pain and the apparent role of ‘suggestion’ as a source of pain relief among those receiving interventions served to catalyse researchers in Great Britain and the United States in the 1930s and 1940s to utilise patient blinding as a means to offset such sources of biased outcomes. As Part 2 relates, Harry Gold and his colleagues at Cornell, concerned about the ‘subconscious’ or ‘unconscious’ bias of researchers, would extend such attempts at rigour in their antianginal studies to include the blinding of researchers as well, resulting in the development – and by the 1950s, formal articulation – of the ‘double-blind’ method of clinical investigation.

## Harry Gold and researcher blinding

Harry Gold (as related in Part 1) favoured limits in controlling for patient context in clinical studies, focusing on pain in the daily life of the patient, rather than performance on an office exercise treadmill test, as the key clinical endpoint to be measured in anti-anginal studies. Nevertheless, he and his colleagues attempted to extend the rigour of clinical evaluation by focusing on a second component of the patient:clinical researcher dyad – researcher enthusiasm and bias in administering interventions and especially in recording results. As a gauge of prevailing sentiment in this respect, in George LeRoy’s 1941 evaluation of papaverine, while LeRoy was proud of being the consistent, dedicated ‘one observer’ making all the interventions and evaluations in his placebo-controlled study, he admitted that he knew which patients were receiving which remedy. Though he ‘attempted to be as non-committal as possible’, LeRoy acknowledged that ‘discerning patients may well have seen my enthusiasm for these particular xanthine drugs’ (LeRoy,<sup>1</sup> p. 924). Likewise, pointing to the possibility for

both patient *and* clinician bias, Joseph Riseman would remind his own readers in 1943 that ‘the clinical evaluation of the benefit of therapy is, in fact, the physician’s impression of the patient’s opinion of the response to treatment’ (Riseman,<sup>2</sup> p. 672). For Riseman (as described in Part 1), the solution to such potential bias was the ‘objective’ nature of the exercise-tolerance test. For Harry Gold and colleagues, it would be the blinding of the researcher.

Gold had first employed researcher blinding in a study published in 1935 of varying formulations of ether.<sup>3</sup> As he would later recount, in the context of the Great Depression, the question of whether (less expensive) ether out of multi-usage large drums was as effective as ether from small cans held important economic consequences. Conventional opinion was mixed, so Gold proposed in 1933 to put the question to the ‘blind test’ (quotes in original), with the administering anaesthesiologists ‘unaware of the source of the specimens except in terms of code numbers or letters’,<sup>4</sup> and those (presumably Gold and his colleagues) assessing the outcomes kept similarly in the dark.<sup>5</sup> As Shapiro and Shapiro have written, it is unclear whether Gold, by invoking the ‘blind test’, was referring back to renowned pharmacologist Torald Sollman’s own earlier usage of the term, to the contemporary Old Gold cigarette ‘Take the Blindfold Test’ marketing campaign (as Gold’s colleague, Nathaniel Kwit, suggested decades later), or some other source.<sup>6–9</sup> Regardless, Gold found no difference in efficacy between the two sources of ether and would become the world’s leading proponent of researcher blinding as ‘a simple expedient which insures a record free of subconscious bias’ (Gold,<sup>10</sup> p. 8).

For the ‘blind test’ of the efficacy of xanthines in angina, the 1937 published paper reported that in eliciting symptoms from patients, ‘to eliminate the possibility of bias, the questioner usually refrained from informing himself as to the agent that had

been issued until after the patient's appraisal... had been obtained' (Gold,<sup>11</sup> p. 2175). Yet, 'usually' belies a more complicated evolution of the study. As Arthur Shapiro uncovered through interviews with Gold and Kwit, the study had begun in 1932 (a year before the ether study would be planned) with only patients blinded. However, within a few years of running the extensive study, Gold and his colleagues determined that the informed researchers were asking 'leading questions' in determining the efficacy of the remedies. Thus, by the end of the study, it appears that the investigators had changed course to ensure that the evaluating clinicians were in the dark regarding what a given patient received.<sup>6</sup> Moreover, in the midst of the xanthine study, yet, after the ether study, Gold and colleagues likewise applied researcher blinding to an animal study of the impact of aminophylline on experimental coronary infarcts, showing no benefit when the intervention 'was unassisted by the intangible something which is apt to be added by an observer's unconscious bias'.<sup>12,13</sup> Double-blinding had been brought to angina evaluation, if not yet named as such.

Over the next decade, Gold would continue to modify the name of such double-blinding, all the while promoting its utility. In a 1943 talk on the 'Treatment of Cardiac Pain', in somewhat revisionist fashion, he noted of the xanthine study that 'our experiments were made with both eyes blindfolded; the patient didn't know what he was getting, and the doctor didn't know what he was giving'.<sup>14</sup> Critiquing studies of androgens as vasodilators during the same talk, he reported: 'I am not in the slightest degree impressed with these results. The studies were not made with the double-eye blind test'. By January of 1947, in the notes for a talk at the New York Academy of Medicine on 'Recent Advances in Therapeutics', we see what appears to be Gold's first invocation of 'the double-blind test', to critique the absence of such rigour in evaluations of antihistamines.<sup>15</sup> This was followed, in a withering March 1948 critique of papaverine for cerebral vasospasm, by a slight modification of the term – now 'double blind-test' – and continued scepticism.<sup>16</sup>

By that year, while the name of the methodology was still in flux, Gold and his colleagues had further refined their technique. In a favourable study of intravenous aminophylline for exertional angina, they reverted to prior nomenclature in reporting that 'the study was conducted by the 'blind' method' (While Gold was not an author on the article, an acknowledgement noted the Beth Israel Hospital (New York) authors' 'appreciation to Dr. Harry Gold, Chief of the Cardiovascular Research Unit, for his advice and help during the course of this study'. The study was 'aided'

by a grant from the Council on Pharmacy and Chemistry of the American Medical Association.).<sup>17</sup> In particular,

the materials for injection, 10 c.c. of a 2.4 per cent aminophylline solution and an identical quantity of physiologic saline, were prepared by a nurse, for each day, in identical syringes marked only with code numbers so that the contents were unknown to the observer as well as to the subject. (Bakst et al.,<sup>17</sup> p. 529)

As further evidence of the efforts taken to ensure such blinding, they continued:

A method was devised for varying the order of the trials [no injection versus saline versus aminophylline] so that all possible combinations of the three tests were used in different sequence. The code numbers on the syringes indicated the order in which the injections were to be given. The contents of the syringes and the corresponding code numbers were noted on cards, sealed in envelopes, and kept sealed until the entire study was completed. (Bakst et al.,<sup>17</sup> p. 529)

The authors also noted: 'Whether saline injection alone produces any increase in the capacity for effort, through suggestion or other means, was not determined because the number of suitable tests for such a comparison was too small' (Bakst et al.,<sup>17</sup> p. 533).

### Approximating the rigour of the laboratory

Two years later, Gold's team published their cross-over evaluation of the vasodilator khellin (dime-thoxy-methyl-furano chromone).<sup>18</sup> Gold would later characterise this as the first true 'clinical pharmacology' study, approximating the rigour of laboratory investigation, with objective measurement of symptoms recorded in a seemingly novel 'daily report card' system, and rigorous blinding of both patients and researchers.<sup>19</sup> Such blinding required a large team, with different members playing their coordinated but independent roles. From one end, as they noted:

One person received the 'daily report cards,' decided on changes in dosage and dispensed the supply of tablets with directions for their use. He knew what the patient had been taking but this knowledge played no part in the record of the results, for his function was neither to question patients regarding the effect of the tablets nor to record judgment;

he merely assembled and filed 'daily report cards'.  
(Greiner et al.,<sup>18</sup> pp. 144–146)

At the other end were the examiners, questioning patients 'under conditions of the "double blind test" in which neither the physician nor the patient knew at the time whether the evaluation related to the placebo or khellin' (Greiner et al.,<sup>18</sup> p. 146). This August 1950 appearance seems to have been the first published invocation of the term 'double blind' (Note that in the notes for a talk on the khellin study at the annual meeting of the American Society for Pharmacology and Experimental Therapeutics delivered in November of 1949, Theodore Greiner had noted its 'double blind' nature. But the term did not appear in the published abstract for the talk. In their published February 1950 study of alpha tocopherol on angina, featuring five of the co-authors from the khellin study, the researchers noted the 'doubly blind conditions' under which the alpha tocopherol study was conducted).<sup>20–22</sup>

The khellin study became an extended opportunity to emphasise the necessity of observer blinding. Holding up LeRoy as a cautionary example, Gold and colleagues publicly noted that an 'evaluation of the physician's enthusiasm (positive suggestion) on the angina of effort would be a study of some interest in itself but it seems self-evident that the physician's enthusiasm is inadmissible in a scientific experiment' (Greiner et al.,<sup>18</sup> p. 153). Privately, and perhaps still stinging from LeRoy's intimation that earlier, negative studies of xanthines may have stemmed from hospital personnel with variable 'diagnostic acuity', Gold noted to himself: 'He knew what the drugs were and he apparently communicated that knowledge to his patients. It was not a true blind test'.<sup>23</sup> The khellin paper, featuring such then- and future stars of clinical pharmacology as Gold, Nathaniel Kwit, McKeen Cattell, Janet Travell, Theodore Greiner and Walter Modell, likewise became an opportunity to discuss and show, by example, the broader requirements for methodological rigour. For instance, whether a given patient would first receive placebo versus khellin was determined by a 'randomized' process, in which half received khellin first, the other half placebos first (Greiner et al.,<sup>18</sup> p. 146). Concerns over the pitfalls of premature analysis were revealed in self-reflective fashion, with an admission of their own eagerness for an early answer, and their initial consideration of khellin as seemingly effective from this partial analysis, a conclusion that would be overturned by their full study (Greiner et al.,<sup>18</sup> p. 146, Gold,<sup>24</sup>). Extending their gaze outward, the comments section was taken up with an extended critical discussion of prevailing

forms of studies, including a pointed critique of what was seemingly 'still the most prevalent, namely, the one in which the patient receives the drug and returns after a week or two with a verbal report on impressions' (Greiner et al.,<sup>18</sup> p. 152). As they concluded: 'This is probably evaluation at its worst'.

### Stabilising the 'double-blind' method

The khellin crossover angina study would be paralleled and followed by two additional double-blind, placebo-controlled studies, of alpha tocopherol and heparin, respectively.<sup>25,26</sup> For the first time among Gold and his team's angina studies, the crossover approach would be replaced by matched-pair randomisation (in which patients were first 'matched' to one another by pre-established criteria, and then randomised within pairs to either the active or placebo control group). With the insufficiency of the crossover approach predicated on concern for the long-term storage of alpha tocopherol in the body, and for the impact of changing 'environmental' and 'seasonal' factors in the heparin study, the two studies represented some of the earliest uses of such matched-pair randomisation in the medical literature.<sup>27</sup> While Janet Travell (who would later garner additional fame as the first female Presidential physician, serving as physician to John F. Kennedy), during the planning and early implementation phases of the heparin study, referred to the 'double blind technic' and 'double blind method' to be employed, she and her co-authors referred in the published paper in April 1953 to the 'double blind-fold method'.<sup>26,28,29</sup>

Terminology was clearly still being stabilised, and Gold, Travell and their colleagues would display some ambiguity concerning what exactly constituted a 'double-blind' study. In the 1950 khellin paper, this referred not to the clinician administering the treatment (though such clinicians would not be involved in the evaluation of patients), but solely to the patients and evaluators. By the time of the 1953 heparin paper, however, Travell and colleagues would note that the

double blindfold method...meant the study was conducted by a team, and that not merely the patients but also the physicians who questioned and examined them, injected the solutions and later assessed the data were unaware of the nature of the coded solution given to any particular patient. (Rinzler et al.,<sup>26</sup> p. 439)

Such enduring ambiguity in the usage of 'double blind' has persisted into the 21st century.<sup>30</sup>

Nevertheless, it would indeed be ‘double blind’ (or ‘double-blind’) that would stick as a term itself (outside the Cornell group, Harvard’s Henry Beecher referred to blinding as the “‘unknowns” technique’, which Gold considered to himself ‘a substitute for our double-blind term’, while across the Atlantic, John Gaddum preferred the term ‘dummy’ to ‘placebo’, though acknowledging in 1954 that such controlling of both subject and research bias was ‘known in America as a double blind test’.) (Beecher,<sup>31</sup> Gold,<sup>32</sup> Gaddum,<sup>33</sup> p. 197). By 1954, Gold would use a Cornell Conference on Therapy to declare:

The whole history of therapeutics, especially that having to do with the action of drugs on subjective symptoms, demonstrates that the verdict of one study is frequently reversed by another unless one takes measures to rule out the psychic effect of a medication on the patient and the unconscious bias of the doctor. The double-blind insures this. (Cornell Conferences on Therapy,<sup>34</sup> p. 724)

Clearly, Gold and his colleagues had been interested in using rigorous clinical studies to shift practices concerning angina. They may have been motivated by experiences like that at another of the Cornell Conferences on Therapy, in 1946, when, after asking legendary cardiologist Harold Pardee whether the suggested impact of a xanthine was not ‘simply [that of] a placebo’, Pardee responded that despite placebo-controlled studies, he had ‘seen things happen which made me think that the drugs are really active’ (Cornell Conferences on Therapy,<sup>35</sup> p. 298). And as late as 1950, Gold and his colleagues could lament that despite the efforts of critical investigators from Evans and Hoyle onward, the ‘survival qualities’ of seemingly ineffective remedies like the xanthines was impressive, a tribute not only to ‘the urgent need of patients for relief and the want of effective measures with which to supply it’, but to the fact ‘that experience indicating beneficial effects has on its side the force of suggestion, and that the methods employed in the investigation of these agents may not have been sufficiently free from defects to carry complete conviction’ (Greiner et al.,<sup>18</sup> p. 151). By the end of the 1950s, Gold would add the inertia of ‘long years of habitual prescribing based on early and authoritative impressions’ as a factor promoting the ‘survival’ of such remedies despite scientific evidence to the contrary (Gold,<sup>36</sup> p. 44).

By this time, and after nearly a quarter of a century of effort to shift the evaluation of anti-anginal therapeutics, Gold and his colleagues could seemingly claim victory, if narrowly defined. As they noted, ‘nowadays it is a rare study of coronary vasodilators

that does not specify control with placebo and “double blind” evaluation of cardiac pain’ (Greiner et al.,<sup>37</sup> p. 244). Yet, as they realised, this was only a partial victory: ‘In other areas of therapeutics, the majority of studies are so poorly designed that their data contain no indication as to the correctness of the conclusion’ (Greiner et al.,<sup>37</sup> p. 244). Gold had his sights set beyond anti-anginals, and Part 3 of this series will examine the relationship between the ‘double-blind’ method and the advent of ‘clinical pharmacology’ more generally from the 1950s onward.

### Declarations

**Competing Interests:** None declared.

**Funding:** None declared.

**Ethics approval:** Not applicable.

**Guarantor:** SHP.

**Contributorship:** Sole author.

**Acknowledgements:** I have benefitted greatly in conceiving and writing this series of articles from the insights and feedback provided by Iain Chalmers, David Jones, Ted Kaptchuk and Nick Rasmussen. I am also grateful to Nicole Milano and Tali Han at the Medical Center Archives of New York-Presbyterian/Weill Cornell for their efforts in making the Harry Gold papers available for examination.

**Provenance:** Invited Article from the James Lind Library.

### References

1. LeRoy GV. The effectiveness of the xanthine drugs in the treatment of angina pectoris. *JAMA* 1941; 116: 921–925.
2. Riseman JEF. The treatment of angina pectoris. *N Engl J Med* 1943; 229: 670–680.
3. Hediger EM and Gold H. U.S.P. ether from large drums and ether from small cans labeled ‘for anesthesia.’ *JAMA* 1935; 104: 2244–2248.
4. Gold H. Plan of Clinical Investigation on Ether Deterioration. 2 October 1933. Box 16, ff 35, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell (inferred though unsigned, 1933).
5. Gold H and Gold D. The stability of U.S.P. ether after the metal container has been opened: with preliminary results of a clinical comparison of U.S.P. ether in large drums with ether in small cans labeled “for anesthesia.” *Anesth Analg* 1935; 14: 92–95.
6. Shapiro AK and Shapiro E. *The Powerful Placebo: From Ancient Priest to Modern Physician*. Baltimore: Johns Hopkins University Press, 1997.
7. Sollman T. The crucial test of therapeutic evidence. *JAMA* 1917; 69: 198–199.
8. Sollman T. The evaluation of therapeutic remedies in the hospital. *JAMA* 1930; 94: 1279–1281.



9. Advertisement. Presenting... Charlie Chaplin in the blindfold cigarette test. *New Yorker*, 23 June 1928, p. 45.
10. Gold H. Some recent developments in drug therapy. *North End Clinic Quarterly* 1941; 2: 5–17.
11. Gold H, Kwit NT and Otto H. The xanthines (theobromine and aminophylline) in the treatment of cardiac pain. *JAMA* 1937; 108: 2173–2179.
12. Gold H. Concerning Therapeutics [Delivered Wednesday, April 6, 1938, at Queens County Medical Building, Forest Hills, Long Island]. Box 7, ff 9, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1938.
13. Gold H, Travell J and Modell W. The effect of theophylline with ethylenediamine (aminophylline) on the course of cardiac infarction following experimental coronary occlusion. *Am Heart J* 1937; 14: 284–296.
14. Gold H. Treatment of Cardiac Pain [Symposium, Post Graduate Hospital, May 4, 1943]. Box 8, ff 26, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1943.
15. Gold H. Recent Advances in Therapeutics [Friday Afternoon Lecture Series, The New York Academy of Medicine, January 17, 1947]. Box 7, ff 40, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1947.
16. Gold H. Discussion of the Paper [“Papaverine in the Treatment of Hypertensive Encephalopathy”] by Russek and Zolman [March 23, 1948]. Box 17, ff 38, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1948.
17. Bakst H, Kissin M, Leibowitz S and Rinzler S. The effect of intravenous aminophylline on the capacity for effort without pain in patients with angina of effort. *Am Heart J* 1948; 36: 527–534.
18. Greiner T, Gold H, Cattell M, Travell J, Bakst H, Rinzler SH, et al. A method for the evaluation of the effects of drugs on cardiac pain in patients with angina of effort: a study of khellin (visammin). *Am J Med* 1950; 9: 143–155.
19. Gold H. Editorial: The proper study of mankind is man. *Am J Med* 1952; 12: 619–620.
20. Greiner TH. A Method for the Evaluation of the Effect of Drugs on the Cardiac Pain of Angina of Effort: A Study of Khelin. To be Presented to the Pharmacological Society. Box 17, ff 4, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, (undated, though date inferred from Greiner 1950), 1949.
21. Greiner T. A method for the evaluation of the effect of drugs on the cardiac pain of angina of effort: a study of khellin. [abstracts of papers, American Society for Pharmacology and Experimental Therapeutics, Inc. Fall meeting, Indianapolis, Indiana, 17–19 November 1949]. *J Pharmacol Exp Ther* 1950; 98: 10.
22. Rinzler SH, Bakst H, Benjamin Z, Bobb AL and Travell J. Failure of alpha tocopherol to influence chest pain in patients with heart disease. *Circulation* 1950; 1: 288–293.
23. Gold H. “Angina Pectoris – LeRoy – 1941.” Box 15, ff 16, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell (inferred though unsigned), undated.
24. Gold H. Clinical Pharmacology of Cardiac Drugs: 23<sup>rd</sup> Series of Friday Afternoon Lectures at the New York Academy of Medicine [March 25, 1949]. Box 7, ff 74, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1949.
25. Travell J, Rinzler SH, Bakst H, Benjamin ZH and Bobb AL. Comparison of effects of alpha-tocopherol and a matching placebo on chest pain in patients with heart disease. *Ann N Y Acad Sci* 1949; 52: 345–353.
26. Rinzler SH, Travell J, Bakst H, Benjamin ZH, Rosenthal RL, Rosenfeld S, et al. Effect of heparin in effort angina. *Am J Med* 1953; 14: 438–447.
27. Welsh BC, Podolsky SH and Zane SN. Pair-matching with random allocation in prospective controlled trials: the evolution of a novel design in criminology and medicine, 1926–2021. *J Exp Criminol* 2022. <https://doi.org/10.1007/s11292-022-09520-2>.
28. Travell J to Gold H. 4 January 1952. Box 5, ff 2, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1952.
29. Travell J. Memo to Harry Gold: Progress Report of Study of Heparin for Angina Pectoris. 11 April 1952. Box 5, ff 2, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell, 1952.
30. Schulz KF, Chalmers I and Altman DG. The landscape and lexicon of blinding in randomized trials. *Ann Intern Med* 2002; 136: 254–259.
31. Beecher HK. Experimental pharmacology and measurement of the subjective response. *Science* 1952; 116: 157–162.
32. Gold H. Notes on “Henry K. Beecher, Experimental Pharmacology and Measurement of the Subjective Response; *Science*, 116; 157, 1952.” Box 17, ff 42, Harry Gold Papers, Medical Center Archives of New York-Presbyterian/Weill Cornell (inferred though unsigned), undated.
33. Gaddum JH. Clinical pharmacology. *Proc R Soc Med* 1954; 47: 195–204.
34. Cornell Conferences on Therapy. How to evaluate a new drug. *Am J Med* 1954; 11: 722–727.
35. Cornell Conferences on Therapy. Treatment of coronary artery disease. *Am J Med* 1946; 1: 291–300.
36. Gold H. Experiences in human pharmacology. In Laurence DR (ed.), *Quantitative Methods in Human Pharmacology and Therapeutics: Proceedings of a Symposium Held in London, on 24th and 25th March, 1958*. London: Pergamon Press, 1959, pp. 40–54.
37. Greiner T, Bross I and Gold H. A method for evaluation of laxative habits in human subjects. *J Chronic Dis* 1957; 6: 244–255.