

# Personal reflections on lessons learned from randomized trials involving newborn infants from 1951 to 1967

William A. Silverman



Bill Silverman, Greenbrae, California, 2003.

## Editor's Introduction

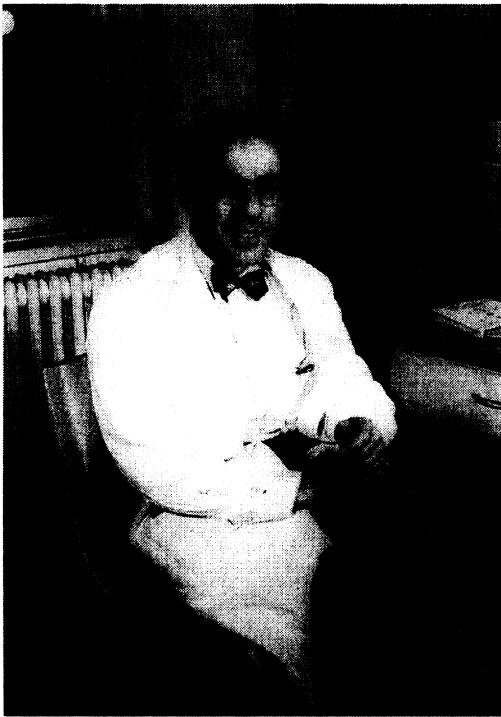
The essay that follows speaks for itself, but both its author and source may require a little introduction. First, the source. We have reprinted this essay from the James Lind Library ([www.jameslindlibrary.org](http://www.jameslindlibrary.org)), a wonderful online cornucopia of historical information created to introduce people to the characteristics of "fair tests" of medical treatments. It includes explanatory essays and over a hundred records containing key passages illustrating the evolution of fair tests over the centuries.

This extraordinary resource is the "retirement" project of the redoubtable Sir Iain Chalmers, that retirement being from his last creative endeavor, the Cochrane Collaboration. He leads this project with a number of other scholars, who are justly proud of their recent Scientific American 2003 Sci/Tech Web award, recognizing the James Lind Library as one of the finest websites on medical topics. Introducing Clinical Trials readers to that site is one purpose of this reprinting.

The second purpose of presenting this essay is to acquaint, or reacquaint, our readership with the life, times, and eloquence of Bill Silverman. Bill, a neonatologist, became a trialist the hard way; seeing his tiny patients suffer and die because he and his colleagues either didn't know what to do, or thought they knew more than they did, a situation probably no less true today than 50 years ago. As is apparent from this essay, Bill became an indefatigable advocate of "evidence-based medicine" almost half a century before the term was coined, and decades before the principles were widely appreciated in medicine. This tale of his early loss of scientific innocence reflects lessons that Bill continued to teach for the subsequent half century, and should be required reading for both students and veterans of clinical trials. It reminds us that clinical trials require from treating physicians not just an understanding of statistical principles, but also humility and courage.

Bill was on a short list of notable trialists we wanted to profile on the pages of this journal in its first year, but he beat us to the punch by interviewing himself. We yielded to the temptation of having his profile written by an interviewer as erudite as his subject. Bill Silverman's professional bona fides are as follows. He graduated from the University of California Medical School in 1942, trained as an intern and resident in pediatrics at UC Hospital in San Francisco between 1942 and 1944, and was resident at The Babies Hospital in New York City between 1944 and 1945. He rose to the academic rank of Professor of Pediatrics, Columbia University; he was elected to the Society for Pediatric Research, the American Pediatric Society, and became an Honorary

Note: This article was first published by the James Lind Library. <http://www.jameslindlibrary.org/essays/cautionary/silverman/html>



Bill Silverman at the Columbia Babies Hospital premature infant station, 1953.

*Founder Fellow of Britain's Royal College of Paediatrics and Child Health. He is the author of Dunham's Premature Infants. Third Edition (Hoeber, 1961); Retrolental Fibroplasia: A Modern Parable (Grune & Stratton, 1980, available online at [www.neonatology.org/classics/parable/default.html](http://www.neonatology.org/classics/parable/default.html)); Human Experimentation: A Guided Step Into the Unknown (Oxford University Press, 1985); Where's the Evidence? (Oxford University Press, 1998); and he has written the column "From our correspondent" in the journal Paediatric and Perinatal Epidemiology since its inception in 1987. He lives in Greenbrae, California with his wife Ruth, with whom he has three children, Dan, Jen and David.*

In 1945, when I began the practice of general pediatrics in New York City, I was appointed to the teaching staff at Columbia University as an instructor at The Babies Hospital. Fortunately for me, as it turned out, Richard Day had just returned to the hospital (he spent the second world war at the US Army's Environmental Research Laboratory working on an improved hand-glove for use in arctic climates). I quickly became one of his most devoted followers. He was the first teacher I ever met who replied to most of my questions with an unapologetic, "I don't know!"

When Dick came across Bradford Hill's book (*Principles of Medical Statistics*) around 1947, he immediately recognized the importance of controlling biases and of statistical arguments in clinical research. He decided to spread the word at our hospital in a series of talks to the staff. But the word "statistics" was off-putting and attendance was very poor – I was the only one in the audience at his last lecture! Like Dick, I was completely sold on the numerical approach; soon we were making nuisances of ourselves by criticizing the subjective "in-my-experience" reasoning of our coworkers. For example, when we first tried to evaluate x-ray diagnoses of hyaline membrane disease in premature infants, we approached John Caffey, the pioneering pediatric radiologist at our hospital (behind his back we called him "Cactus Jack"). We asked him to make "blind" readings of x-rays (to read unmarked films, without clinical information, and again after a one month lapse). He refused to cooperate and was unmoved by our pleas for the need to determine rates of false positive and false negative diagnoses of hyaline membrane disease: "I wouldn't believe your statistical arguments," he told us, "even if you proved them to me." (It took me years to realize that Jack was certainly right about the nature of ultimate proof in medicine!)

In 1949, a large federally funded premature infant station was opened in The Babies Hospital; it was part of a newly established city-wide program for the specialized care of the smallest newborn infants. They were transported from small maternity hospitals in a specially outfitted ambulance, operated by the City of New York, to stations in the large university hospitals. Routine ophthalmoscopic examinations of each infant at weekly intervals had just been initiated in our new unit to assess the prevalence of retinopathy of prematurity (originally called retrolental fibroplasia, but I'll use ROP, the present-day term, here). This mysterious disease was first described in Boston in 1942, but early stages had never been diagnosed in New York up to this time. A few weeks after our new unit opened, Fred Blodi, the young ophthalmologist-in-training who carried out the eye examinations, found the first example of the initial retinal vascular signs of the strange blinding disorder. (The baby, my patient, was the child of the professor of biochemistry. Prior to this pregnancy his wife had six miscarriages. The current 1200 gram premature infant was the first of her offspring to live.)

We panicked at the thought that this "precious child," who was now thriving, would go blind. Something must be done immediately, we felt, to prevent a tragedy. Fred Blodi suggested we try using the then new anti-inflammatory agent, adrenocorticotropic hormone (ACTH). In desperation, we grasped at this "straw" and administered the

powerful drug. As it had never been used before in newborn infants, we made a crude guess about dosage based on reports of use in experimental animals. The vascular proliferation seemed to subside after treatment began, but the systemic effects of the drug (ravenous hunger, growth arrest, Cushing syndrome-like manifestations) were frightening. We reduced the dose but the eye changes worsened, so the dose was increased and, again, the changes improved. After about a week on treatment, we cautiously reduced the ACTH. This time there was no flare up and we stopped treatment without incident. The infant gained weight and her eyes were almost normal when she was sent home.

Following this dramatic experience, 31 infants with early changes of ROP were treated with ACTH: 25 left the hospital with normal eyes, two became blind, two lost all vision in one eye, and two had useful vision with only minor retinal scars. Our results were impressive when compared with seven untreated infants with early retinal disease who had been looked after at Lincoln Hospital: six of these had become totally blind. Although ACTH seemed to be the cure for ROP, we were brought up short by the observations in two infants; the one infant at Lincoln who had escaped blindness without treatment, and another baby, followed by Blodi at our hospital, in whom early abnormalities had subsided without treatment.

Our exciting post hoc results did not satisfy the rules of evidence we had been touting loudly at our hospital for more than two years! There was no way to avoid the issue: inspired by the then recently reported MRC randomized trial of streptomycin for pulmonary tuberculosis in the UK, we concluded that a randomized controlled trial involving the two New York hospitals was the only way to test the efficacy question quickly and rigorously. After much soul searching, Dick Day and I approached our chairman, Rustin McIntosh, and explained our agonizing problem. Although we felt sure ACTH prevented blindness, we were worried about broadcasting this provocative claim based on the weak evidence in a consecutive series of our patients without a concurrent comparison group (and we could not ignore Blodi's experience with infants who had improved spontaneously). We asked for permission to conduct what was for us a completely unprecedented clinical exercise: a randomized controlled trial involving newborn infants had never been carried out in our hospital, or for that matter, anywhere else. After a few minutes of silence at the end of our impassioned presentation, Rusty said, "You must do it!" I will always remember his courageous decision. This was years before research ethics committees reviewed such proposals; the entire responsibility for this daring action rested on his shoulders.

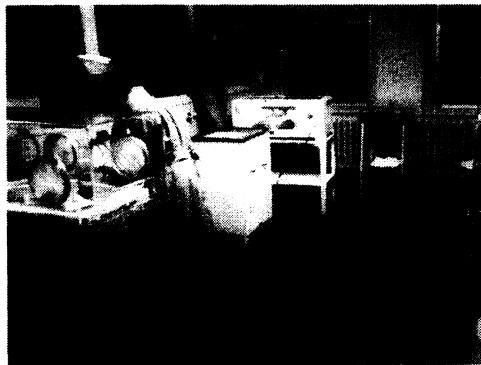
It took only a few months to accumulate the required experience in the two hospitals [1]. Allocation to ACTH or no ACTH was decided by drawing marbles from a jar containing an equal number of white and blue marbles: one morning, when a new infant became eligible for enrollment, I noticed that our head nurse shook the jar vigorously, turned her head away, pulled a marble out (just as she had been instructed); but because she didn't like the "assignment," she put the marble back, shook the jar again, and pulled out the color that agreed with her bias! The importance of Bradford Hill's precaution in Britain's famous streptomycin trial to conceal the order of assignment in sealed envelopes was immediately obvious!

Approximately one-third of the infants treated with ACTH became blind (in one or both eyes), but only one-fifth of those who had not received the drug. Furthermore, there were more fatal infections in the ACTH-treated group. When the entire two year experience in both hospitals was examined, the frequency of scarring ROP was not startlingly different, but the mortality rates were quite disparate: the *untreated* group had fared better.

Two years later, ophthalmologists at Johns Hopkins reported that three-quarters of infants with early vascular changes of ROP showed spontaneous regression to normal. This was exactly our experience at The Babies Hospital during the ACTH treatment episode. Following this disheartening (but very educational!) experience, a formal controlled trial was undertaken to test the possibility that exposure to light was responsible for blinding these babies. Once more the results were negative [2]. Needless to say, these dramatic early experiences transformed my life – I became a highly vocal advocate of randomized controlled trial methodology!

By the early 1950s, after numerous false leads, most physicians immersed in the study of ROP were understandably skeptical of all new claims of cause and cure. The disease was now responsible for blinding about 10000 premature infants throughout the world. In the spring of 1953, a group of concerned pediatricians and ophthalmologists, convened in Bethesda by the US Public Health Service, discussed the increasingly urgent claims that liberal use of supplemental oxygen was responsible for the huge epidemic (*In A cautionary tale about supplemental oxygen*, I have given an account of the oxygen story [3]). The above noted experience at The Babies Hospital explains why Dick Day and I fought so vigorously for the randomized controlled trial format when the 1953–54 national trial of supplemental oxygen was planned.

Following the belated realization that a seemingly benign intervention like oxygen – a time-honored life-saving "drug" – could have such unexpected,



Babies Hospital premature infant station, 1951. Babies are housed in Gordon-Armstrong incubators in preparation for discharge.

unrecognized and devastating consequences, we realized that almost *everything* we were doing to care for premature infants was untested. At mid-century, before the arrival of ventilators and microchemistry, care of marginally viable newborn infants was essentially "pastoral." Like the approach taken by farmers caring for newborn piglets, conditions considered ideal for survival were provided, and it was assumed that those who were "meant" to survive would do so. But none of these purportedly "ideal conditions" had ever been subjected to formal parallel treatment trials.

A detergent mist was at that time the most recent and widely acclaimed treatment for respiratory distress syndrome – the new name for hyaline membrane disease, then the leading life-threatening disorder seen in premature infants. Since oxygen was used to propel the detergent into the incubators through a nebulizer, it was important to find out whether the effect of this treatment was sufficiently favorable to balance an increased risk of ROP (additionally, the detergent was seeping into the electric motors of the incubators and ruining them). We began a fixed sample size randomized controlled trial in 1953 to address these uncertainties [4]. Having decided the size of what we would deem an "important difference" in outcome, we calculated the number of our patients that we would need to enroll. To avoid the temptation of "stopping the race when our horse was ahead", we pledged not to "peek" at the results until the required number of infants had been enrolled in this first formal comparison of detergent mist versus plain water mist. We were unable to detect any differences in the mortality rates over the five days following enrollment and the findings at post-mortem examination (autopsy) in infants who died in each of the compared groups.

In 1954, we followed this trial with another comparing nebulized water mist versus high

humidity (80–90% relative humidity) and, once more, were unable to detect any apparent difference in first five day mortality, obstructive respiratory signs (measured by a "retraction score" we devised for this purpose), and in findings at autopsy [5].

Finally, at the end of 1954, we undertook the third (and what we thought would be the last) in the series of fixed sample size trials of atmospheric conditions. This time we sought to compare high humidity (80–90% relative humidity) versus moderate humidity (30–60% relative humidity). The latter condition had been maintained in American incubators for almost two decades before mist treatment had been introduced.

In the 1954–55 "humidity" trial we used a fixed sample size, factorial design (similar to formats used for many years in agricultural field trials), as suggested by John Fertig, Professor of Biostatistics in the School of Public Health at Columbia University. This plan was adopted to allow us to carry out a concurrent comparison of two regimens of antibacterial prophylaxis: penicillin/sulfisoxazole (widely used in the US and prescribed routinely in our unit with no apparent problems for more than a year) versus tetracycline (a newly available agent which promised to be easier to administer). The results of the study turned out to be deeply disturbing [6]. First five day mortality was strikingly higher among infants allotted to the arm of the trial treated with the widely accepted agents for prophylaxis (penicillin/sulfisoxazole) compared to the group receiving the proposed replacement (tetracycline). A previously unknown and subsequently demonstrated effect of sulfisoxazole was responsible for the startling and completely unexpected result: the drug displaced albumin-bound bilirubin in the serum of jaundiced neonates with the result that they sustained fatal brain damage from a condition known as kernicterus. Moreover, we realized immediately that the trial would have been stopped sooner if we had "peeked" at the results during the course of what we regarded as a very mundane exercise, instead of waiting until all infants had been enrolled.

This horrific experience convinced us of the need to find a method of "controlled peeking" at the accumulating results in randomized trials that were still recruiting. John Fertig suggested that we consult the statistical group at Columbia who had devised an interesting sampling scheme during the second world war. The US Navy needed an efficient method to determine the "dud" rate in batches of torpedoes manufactured by various munitions companies. The military approached Columbia's famous statistician, Abraham Wald, who worked out a sequential design for sampling, which minimized the number of torpedoes that needed to be exploded to obtain a reliable estimate of the proportion of

defective bombs in each batch. (Wald died in a tragic plane crash a few years before we consulted his coworkers.) We adopted a version of the sequential scheme that allowed continuous, but controlled, oversight of results in all subsequent randomized controlled trials conducted at The Babies Hospital.

In the meantime, we examined the results of the concomitant "humidity" section of the factorial randomized controlled trial and found another startling result [7]. We did not expect to detect any difference in first five day mortality, but found instead that it was *lower* in infants allotted to the "high humidity" arm of the trial! We were puzzled by this outcome because the respiratory retraction scores, incidence of infection and findings at post-mortem were virtually the same in both groups. In "dredging" through the records, we found a small but consistent decrease in body temperature among infants reared in "low humidity" (30–60% relative humidity).

It seemed very unlikely to us that slightly low body temperature was responsible for an increase in mortality: for more than 20 years, incubator temperatures in America had been intentionally set to maintain relatively low but steady body temperature. This widely accepted practice was based on the findings in a prolonged observational study in Boston in the 1920s (reported in 1933) of the influence of various conditions of the physical environment on the well being of premature infants. Our unpredicted difference in mortality among babies cared for in the different humidities compared seemed to be a fluke.

Nonetheless, we were now very confident we had a powerful tool to test the "temperature hypothesis" suggested by the associations turned up in the 1953–54 trial. In March 1956, we began a trial comparing first five day mortality among infants housed in incubators maintained at two contrasting levels of ambient temperature (31–32°C versus 28–29°C) and one level of humidity (80–90% RH) [8]. A matched pairs sequential plan (devised by John Fertig and Agnes Berger) was used to allow a running analysis of outcome. In February 1957, a predetermined "decision-line" was crossed, indicating that lower mortality was associated with the warmer incubators.

Six years later these results were confirmed independently in trials conducted in Baltimore and in Pittsburgh. Three separate replications confirming the surprising findings in our 1953–54 trial – small differences in body temperature were associated with measurable differences in mortality – and these findings settled a very old score. Seventeen years before our trial, Dick Day had made some painstaking physiological measurements of thermoregulation in premature infants. His findings – that these babies were truly home-

othermic – challenged the widespread practice of caring for newborn infants in slightly cool incubators. But the authorities of the time dismissed his suggestions out of hand, and the everyday custom of maintaining relatively low body temperature in newborn infants continued unchanged for years.

Following these revealing randomized trials, we conducted several more tests of physical environments to tie up some loose ends: a randomized controlled trial comparing two levels of humidity at one body temperature, maintained by a servo-controlled radiant warmer constructed specifically for this purpose [9]; a trial examining the influence of the thermal environment on acid-base homeostasis in the first hours of life of normal neonates [10]; a randomized controlled trial examining the effect of the thermal environment on growth and on cold resistance of small infants after the first week of life [11]; and, finally, in 1967, a trial of the effects of thermal environment and caloric intake on growth after the first week of life [12].

I found these early exploitations of the power of randomized controlled trials very exciting, but I was increasingly aware that the statistical approach was anathema to free-wheeling doctors who resented any doubts being expressed about the effectiveness of their untested treatments. (A well known anesthesiologist at Columbia, tired of our criticism of her use of immersion of newborn babies in ice-water to resuscitate them if they were asphyxiated, finally enrolled in a course on statistics. When she finished, she told Dick Day and me, "Now I know what you guys have been talking about, but I still don't believe it.") Most of my colleagues disparaged the numerical approach to clinical problems and their disdain rubbed off on house staff. These doctors-to-be resented being obliged to follow the terms of a study protocol and the discipline of deciding treatment by opening opaque envelopes. Over and over again, in our trials, I saw them hold the sealed envelopes up to a light to try to read the treatment assignments.

I felt it was important for trainees to learn about the new methods, and when the 1953–54 factorial trial was planned I asked for volunteers to help, so they would get hands-on experience. Not one stepped forward. When I wrote up the incredible results of the antibacterial prophylaxis trial, I sent a copy of the draft to Columbia's world famous professor of neuropathology (he had read all the slides of brain tissue "masked" as to treatment). He refused to allow his name to appear on the paper as a co-author: "Mere statistics!" he sniffed. When I was invited by a prominent medical school to give a talk, suggesting *The Randomized Trial* as a title, I received a telegram asking "A randomized trial of what?" My hosts found it inconceivable that I would talk for an hour "only about methodology" to assess the effects

of our treatments. At research society meetings, randomized controlled trials were also belittled; one of America's brightest up-coming researchers was very frank, "Bill," she said, "your trials are so boring!"

In 1969, I saw the destructive results of these condescending attitudes (the mind-set should be called "reductionist snobbery") at a meeting to plan a cooperative study of the proper level of oxygen in the treatment of premature infants. (In *The Unresolved Oxygen-Level Issue...* I have described how the anti-trialists delayed a rigorous trial of this important question for 35 years [13].) The inflated reverence for the reductionist approach to highly complex problems (the reductionist, when faced with Newton's problem to discover the source of gravity, cuts open the apple and looks inside) appeared again in 1977, at a meeting of the prestigious Institute of Medicine. I made a plea for a large scale multicenter randomized controlled trial to settle the long standing question whether vitamin E prophylaxis could reduce the risk of ROP blindness. But the Director of the National Institute of Child Health and Human Development won the debate. He told the panel, "Randomized trials have never proved anything; all we need to do is study the chemistry of tocopherol." The efficacy of this form of prophylaxis remains uncertain to the present day.

The impatient let's-try-it-and-see approach in the burgeoning field of neonatal medicine has resulted in therapeutic disaster after disaster. It is extremely difficult, I've learned over the years, to convince physicians about the importance of a hard won modern lesson (the amazing sulfisoxazole incident was the most indelible and instructive example).

Since knowledge in medicine is *never* complete, the use of concurrent controls in clinical trials of proposed interventions cannot prevent all therapeutic catastrophes. But the precaution can *always* bring about a substantial reduction in the number of patients maimed and killed as the result of inevitable surprises!

Finally, I'm encouraged to see that an editorialist [14] has recently emphasized "the critical importance of randomized clinical trials in evaluating new therapies – even heroic procedures performed in only a small fraction of neonates – before they are adopted as part of standard practice." For example, the willingness of pediatric surgeons to submit a daring fetal intervention [15] to a rigorous parallel-treatment trial is a hopeful sign: perhaps anti-trialist obstruction is, at long last, on the wane.

Note: These informal comments were written in Greenbrae, California (at the request of my friend, Iain Chalmers) in October 2003, on the occasion of my 86th birthday (10/23/03)

## References

1. Reese AB, Blodi FC, Locke JC, Silverman WA, Day RL. Results of the use of corticotropin (ACTH) in treatment of retrorenal fibroplasia. *AMA Arch Ophth* 1952; **47**: 557–55.
2. Locke JC, Reese AB. Retrorenal fibroplasia. The negative role of light mydriatics and the ophthalmoscope examination in its etiology. *Arch Ophth* 1952; **48**: 44–47.
3. Silverman WA. A cautionary tale about supplemental oxygen: the albatross of neonatal medicine. *Pediatrics* 2004; **113**, in press.
4. Silverman WA, Andersen DH. Controlled clinical trial of effects of Alevaire mist on premature infants. *JAMA* 1955; **157**: 1093–96.
5. Silverman WA, Andersen DH. A controlled clinical trial of effects of water mist on obstructive respiratory signs, death rate, and necropsy findings among premature infants. *Pediatrics* 1956; **17**: 1–10.
6. Silverman WA, Andersen DH, Blanc WA, Crozier DN. A difference in mortality rate and incidence of kernicterus among premature infants allotted to two prophylactic regimens. *Pediatrics* 1956; **18**: 614–25.
7. Silverman WA, Blanc WA. The effect of humidity on the survival of newly born infants. *Pediatrics* 1957; **20**: 477–86.
8. Silverman WA, Fertig JW, Berger AP. The influence of the thermal environment upon the survival of newly born premature infants. *Pediatrics* 1958; **22**: 876–85.
9. Silverman WA, Agate FJ, Fertig JW. A sequential trial of the non-thermal effect of atmospheric humidity on survival of the newborn infant of low birthweight. *Pediatrics* 1963; **31**: 719–24.
10. Gandy GM, Adamsons K, Cunningham N, Silverman WA, James SL. Thermal environment and acid-base homeostasis in human infants during the first few hours of life. *J Clin Invest* 1964; **43**: 751–58.
11. Glass L, Silverman WA, Sinclair JC. Effect of the thermal environment on cold resistance and growth of small infants after the first week of life. *Pediatrics* 1968; **41**: 1033–46.
12. Glass L, Silverman WA, Sinclair JC. Relationship of thermal environment and caloric intake to growth and resting metabolism in the late neonatal period. *Biol Neonat* 1969; **14**: 324–40.
13. Silverman WA. Commentary: The unresolved oxygen-level issue – hijacked by anti-trialists. *J Perinat* 2004; in press.
14. Wenstrom KD. Fetal surgery for congenital diaphragmatic hernia. *New Eng J Med* 2003; **349**: 1887–88.
15. Harrison MR, Keller RL, Hawgood SB, et al. A randomized trial of fetal endoscopic tracheal occlusion for severe fetal congenital diaphragmatic hernia. *New Engl J Med* 2003; **349**: 1916–24.