Stewart Justman
Missoula, MT USA

"Superiority to Placebo" Prior to the 1962 Drug Amendments

Before the meeting of the Gastrointestinal Drugs Advisory Committee of the U. S. Food and Drug Administration on August 28, 2012, FDA reviewers questioned whether a drug for ulcerative colitis that showed a superiority to placebo of less than 10% really warranted use in place of existing therapies. In two clinical trials, 16.5% and 18.5% of patients treated with adalimumab (Humira) had shown clinical remission after 8 weeks, compared with 9.2% and 9.3% on placebo.[1] Apparently the reviewers had misgivings about a drug which, though superior to placebo, hovered all too close to the line of non-efficacy for the given indication. This essay seeks to establish that for many of those who pioneered the sort of controlled trials now required by the FDA, such a narrow margin of superiority to placebo did not suffice to distinguish a drug *from* placebo.

The story behind the requirement that a drug demonstrate efficacy in clinical trials traces back to the pharmaceutical revolution of the 1950's, when over 500 new drugs or drug combinations, many of them poorly differentiated from placebo, poured into American pharmacies each year.[2] In response to this sudden and complete transformation of the pharmaceutical marketplace, Sen. Estes Kefauver launched hearings in 1959 to call attention to the drug industry's excesses and build support for price and patent reform. The hearings bogged down, however, and by all indications the Drug Amendments of 1962, which remain in

effect to this day, would not have passed but for the thalidomide disaster. It was in the shock of that moment, and perhaps with some degree of confusion, that Congress voted unanimously for a bill that seemed destined for defeat not long before. A crisis of safety came to the rescue of the Drug Amendments' efficacy provision.

The Drug Amendments require new drugs to show "substantial evidence" that they have the effect they are purported to have, with "substantial evidence" consisting of "adequate and well-controlled investigations." The sanctity acquired over the decades by this phrasing has obscured the oddity of using a means of obtaining evidence (rigorous trials) to describe the result itself.[3] Emphasizing method as it does, the term "adequate and well-controlled investigations" implicitly refers to the mid-century movement to reform the very procedures of clinical investigation by introducing stringent safeguards against bias: blinding (and ideally double-blinding), placebo controls and randomization. Hence the rising presence of methodologically demanding trials, as opposed to impressionistic reports, in the medical literature around the time of the Drug Amendments themselves.

The technique of randomization seems to have made its appearance in a 1926 trial of Sanocrysin (a reputed treatment of tuberculosis), wherein a coin toss determined which subjects would receive the drug and which a placebo.[4] As this suggests, the guards against bias in clinical trials complement one another, and in their evolution it is placebo controls that came to bear most directly on the matter of efficacy. One after another, drugs whose potency seemed impressive to the naked eye proved difficult to distinguish from placebo under controlled testing. But if placebo is necessary to establish efficacy (for without running a drug against placebo we might not be able to tell whether it was a placebo itself), where should the

standard of efficacy be set? The only standard offering a clear line of demarcation for regulatory purposes would be superiority to placebo: a threshold which indeed has become the standard for drugs evaluated in placebo-controlled trials in the United States.[5]

Yet if a drug need only show some degree of superiority to placebo, this opens the possibility of drugs trivially better than placebo entering the marketplace under the color of efficacy, quite as if the sort of placebic drug once depreciated by reformers of trial methodology had passed through the ordeals of regulation itself.[6] Consider the anti-migraine drug ubrogepant (Ubrelvy), much advertised at this hour.

In a recent clinical trial of this drug, "Pain freedom at 2 hours was reported by 101 of 464 participants (21.8%) in the ubrogepant 50-mg group, 90 of 435 (20.7%) in the ubrogepant 25-mg group, and 65 of 456 (14.3%) in the placebo group (absolute difference for 50 mg vs placebo, 7.5% . . . 25 mg vs. placebo, 6.4%)."[7] With six or seven points separating it from placebo, ubrogepant falls into the category of drugs whose superiority to that baseline is more statistical than substantive. When the trialists claim that "acute treatment of migraine pain with ubrogepant compared with placebo led to significantly greater rates of pain freedom at 2 hours with both the 50-mg and 25-mg doses", the term "significantly" reads like a play on words. This is but one example of an approved treatment with an advantage over placebo of "near-zero."[8]

In the end, the Gastrointestinal Drugs Advisory Committee recommended approval of adalimumab for ulcerative colitis, and the FDA did so.[9]

The Inactive Placebo

If placebo is worthless, then a drug that performs marginally better can only be worthless also.

In the early days of placebo-controlled trials, before the concept of "the powerful placebo" caught on, the placebo was used to keep subjects untreated so that the test drug's effects could be gauged by comparison. In other words, the placebo was a mask for nontreatment—worthless indeed. This was the case not only in the Sanocrysin trial but two notable trials of the 1930s where placebo allowed the investigators to control for the course of the condition under study.[10]

In an early instance of a carefully controlled investigation, Harold Diehl in 1933 reported a randomized, double-blind study in which various treatments of the common cold were tested against each other and placebo. Explaining his decision to use placebo, Diehl wrote:

> It seemed essential at the beginning of this study to have reports from a series of patients who thought they were taking some presumably effective medication but who in reality received no medication whatever. For this purpose tablets and capsules of lactose were employed. The proportion of good results (35 per cent) reported after lactose is indicative of the spontaneous improvement in acute colds for which any medication that happens to be taken is given credit.[11]

Placebo, then, is a device that allows Diehl to estimate the effect of tested treatments by accounting for the course of a self-limited illness. There is no suggestion that it is an agent in its

own right, let alone a potent one. Quite the contrary. Placebo did not produce improvement in 35% of cases; time did. And if placebo has no effect, then a treatment that proves only marginally better is necessarily of negligible value. Notably, Diehl judges the figure of 42% of subjects who improved on acetylsalicylic acid "not significantly greater" than the 35% who improved on, but not because of, lactose. In a related study, Diehl et al. found differences between the test treatment (in this case a vaccine for the common cold) and placebo that were statistically significant but "too small to be of practical importance."[12]

Similar in its understanding of placebo and its estimate of marginal superiority is a controlled study of angina treatments by Evans and Hoyle also published in 1933, but in Britain. Here too placebo has none of the mystique of power it was later to acquire; rather, it is a spurious semblance of treatment that enables the investigators to control for the erratic course of an illness marked by spontaneous variation. As they note,

> No facts seem to be available on variations in the severity of symptoms during the course of angina of effort over weeks or months. This knowledge is essential if we are to have control of therapeutic investigations. A contribution to this problem is furnished by our control observations.[13]

That is, the authors, like Diehl, use placebo to account ("control") for the course of the condition under study. And because placebo treatment is nontreatment, a test drug that proves minimally or ambiguously superior to placebo is of too little value to be employed in clinical practice in any but unusual cases. Papaverine, for example, is dismissed as follows: "A

greater number of patients showed moderate improvement with papaverine compared with a placebo, although this was less than with morphine. The difference, however, is slight, and does not support the view that papaverine has any real value in continuous treatment." Like Diehl, Evans and Hoyle use a placebo baseline to rule out treatments of no or dubious value, not to rule in treatment technically superior to zero.

Glancing back at the Evans and Hoyle study some years later, J. H. Gaddum situated it as the first of a series of investigations, all of which found that "drugs tested for the continuous treatment of angina pectoris were practically the same as the effects of dummy treatments."[14] As it happens, in one of these investigations 37% of patients treated for angina pain with vitamin E and 27% treated with placebo reported improvement, a difference the authors deem inconsequential. "As we have indicated", they write, "the response to alpha tocopherol was essentially the same as that to the placebo."[15] By exposing the fallacy of uncontrolled observation so convincingly, the thread of studies to which this belongs helped establish the methods of the clinical trial as we now know it. Both careful in their procedures and consistent in their findings, the series of angina trials added up to a strong argument in favor of what the Drug Amendments of 1962 would call "adequate and well-controlled investigations."

In the United States, the movement to reform the theory and practice of clinical investigation is particularly indebted to the work of Harry Gold, considered by many on his side of the Atlantic the founder of clinical pharmacology, and leader of a study of certain angina treatments then in favor that ran contemporaneously with Evans and Hoyle's.[16] Given that Gold et al. list "spontaneous variations in the course of the pain" as the first of many confounders liable to muddle evaluation of such treatments unless taken into account, and

given that they undertook long courses of treatment with both placebo and drug in the same patient, it seems that they too used placebo to control for the natural history of a notably variable illness. Placebo had no value (other than as a placeholder) and was not a treatment.

Over the five years' duration of this meticulous study, Gold and his colleagues revised their procedures, quite as if the poor state of existing practice left them no choice but to invent new protocols of investigation as they went.[17] Only after the fact did they learn that a methodologically exceptional study of angina treatments had run concurrently with their own in London. Citing the Evans and Hoyle study as an honourable exception to the rule of uncontrolled and therefore pointless investigations, they note that "Our study, carried out under substantially similarly controlled conditions, has yielded similar results, and our conclusion is in accord with theirs; namely, that the xanthines exert no specific action that is useful in the routine treatment of cardiac pain." This verdict against the xanthines rests not so much on quantitative data (which are not tabulated in a manner comparable to the British study) as on probative evidence, specifically that "every type of change in pain observed during the use of a xanthine was reproduced in the same individual by a period in which a placebo was used." If nontreatment reliably yields the same response in the same person as a test treatment, then the latter is assuredly worthless.

As the inert placebo grew into the powerful placebo over the years to come—as a control for the course of disease took on something of the aspect of a treatment in its own right—the wary investigator had one more thing to be wary of. Gold was especially concerned with the placebo's ability to confuse estimates of dubious drugs, and in the spirit of bringing into the open the rather unmentionable topic of the use of placebos in clinical practice,[18] he

organised an historic symposium on placebos at Cornell in 1946, an account of which takes us about as close to the headwaters of the voluminous literature on the placebo effect as we are likely to get.[19] Participants at this first of the Cornell Conferences on Therapy discussed the universality of the placebo effect in clinical medicine, its surprising strength, and its potential to confound judgments of therapeutic efficacy. It was in connection with the last topic that Gold himself made a strong case against the use of placebic drugs in medical practice.

The stenographic summary of the proceedings of the 1946 conference closes on a point underscored by Gold: unless a doctor uses a "pure" placebo such as lactose, the patient's reaction to placebo is likely to be misconstrued as a response to a pharmacological agent, and so the doctor becomes the dupe of the placebo effect. "When we use an agent of questionable pharmacological activity . . . there is the danger of deceiving two people, both the patient and the physician. The doctor may come to think that the agent has potency when, in fact, it has none."[20] In a striking move that others were to duplicate, with variations, in the decade and a half between the inaugural Cornell conference and the Drug Amendments of 1962, Gold equates drugs of dubious or irrelevant potency with substances of no potency at all.

In the notes of the conference Gold summarises the proceedings, ending with his own argument that while a doctor has every right to prescribe a placebo, the placebo itself "gains no validity by the inclusion of materials of doubtful indication, of equivocal actions, or in ineffectual amounts. It was urged that the inclusion of such materials be discouraged for they frequently deceive the physician into believing that the particular agent possesses other than psychotherapeutic properties." In other words, the physician who decides to use a placebo would be well advised to choose an unambiguously inert one rather than an "impure" placebo

(perhaps a vitamin or subtherapeutic dose of medication) which offers a semblance of medical propriety but whose actual value does not materially exceed that of lactose. As the procedural reforms espoused by Gold and others began to make their mark on clinical investigation, the distinction between pure and impure placebos lent itself to the interpretation of trial findings. In effect, when a drug was tested against placebo and found marginally superior (as with vitamin E for angina), the control was a pure placebo and the drug an impure one. As paradoxical as it may seem, the drug was marginally superior to placebo and yet a placebo itself.

Two different understandings of marginal superiority to placebo have emerged so far. In the first, because a placebo is an inert substance used to control for the course of illness, it has no treatment value whatever, and therefore a test drug that does slightly better (like papaverine in the Evans and Hoyle study) is too nearly useless to merit a place in medical practice, except perhaps in rare cases. In the second, placebo has a power of its own and drugs of little merit free-ride on it, thereby disguising their own worthlessness. As the reputed power of the placebo grew in parallel with the explosive increase in the American pharmacopoeia in the years following the Cornell conference, the latter understanding of drugs trivially superior to placebo predominated.

The Potent Placebo

The 1926 trial of Sanocrysin was randomized and placebo-controlled but not double-blinded; the xanthine trial reported in 1937 was placebo-controlled but not randomized, and

only in the middle of it did the investigators come to appreciate the importance of double-blinding; the historic streptomycin trial reported in 1948 was randomized but not (for good reason) placebo-controlled.[21] It was in the 1950's that the elements of the clinical trial coalesced into a single distinctive package of procedural controls on bias. The same years witnessed an exponential increase of the American pharmacopoeia.

In the 1950's, drugs that had never seen a controlled test in human subjects flowed into American pharmacies at such a rate that Gold and colleagues conducted a study specifically to validate a short-cut method a practicing doctor might employ to assess a drug without succumbing to the usual biases of evaluations performed in the office itself. For the purposes of demonstration, the group tested a preparation representing the most popular class of drugs in the United States—laxatives—against a known standard. In the course of the trial, "minor differences between the bran [that is, the test preparation] and placebo" emerged.[22] As the investigators sharpened the trial to probe this difference, it disappeared, their final verdict being that the bran preparation and placebo could not be distinguished. In essence, the bran was an impure placebo. An unvoiced implication of the study seems to be that many drugs then coming into use would also be found wanting if duly tested.

The stocked pharmacy and its "sea of confusion"[23] lie in the background of an article that speaks for the concerns of the reformers of clinical investigation, Henry Beecher's "The Powerful Placebo" (1955), which has since been cited over 2500 times.[24] At once a sort of meta-analysis before the fact and an editorial, "The Powerful Placebo" argues, in the tradition of Gold, that the placebo effect is potent enough to play havoc with the evaluation of therapies unless carefully controlled. Tabulating the findings of a handpicked assortment of trials, many

of them his own, Beecher concludes that on average 35% (+/- 2%) of patients with subjective complaints, including pain, will improve on placebo; and in the manner of an editorialist, and in keeping with the position he takes elsewhere, he maintains that only a trial with strict procedural and statistical safeguards can possibly disentangle drug effects from placebo in such cases.

Decades after it appeared, Beecher's influential article was sharply challenged by two critics on many grounds, the first of which is that it misattributes spontaneous improvement to the power of the placebo.[25] The critics have a point, all the more in that Beecher cites both Diehl's study of treatments of the common cold and the Evans and Hoyle study of angina treatments as evidence of "the powerful placebo," when the fact is that both employ placebo to control for a fluid course of illness. In effect, even while Beecher espouses the more recent concept of the worthlessness of placebic drugs, he incorporates by reference the original notion that a drug minimally superior to no treatment is unworthy of medical use.

While Beecher was concerned that the placebo effect can overwhelm the signal of a valuable drug, by the end of "The Powerful Placebo" the stress falls on the reverse problem: the artificial enhancement of the performance of ineffective drugs by the power of the placebo. "Many 'effective' drugs have power only a little greater than that of a placebo. . . . Many a drug has been extolled on the basis of clinical impression when the only power it had was that of a placebo." Significantly, these two statements occur almost side by side even though the first refers to drugs marginally better than placebo and the second to drugs no better than placebo at all. Just as Gold judged impure placebos (that is, placebos with a semblance of pharmacological activity) as no better, but certainly more confusing, than pure ones, so

Beecher regards drugs "a little" superior to placebo as nothing but placebos that have secured a place in the pharmacopoeia. In his précis of the 1946 conference on the placebo effect, Beecher highlights Gold's comments on placebic drugs.

We can well understand why Beecher would dismiss drugs scarcely distinguishable from a sham treatment employed as a prop in a medical ritual. Ever since he observed a kind of spontaneous anesthesia in men wounded in the Anzio landing in 1944, he had taken a strong interest in the treatment of pain, and it is in this connection that "The Powerful Placebo" notes at one point that "75% of a group in severe postoperative pain are satisfactorily relieved by a large dose of morphine . . . but 35% are relieved by the placebo."[26] On this showing, placebo may indeed be powerful, but its power does not approach morphine. In a study of his own tabulated in "The Powerful Placebo", Beecher found that by a "conservative" estimate, pentobarbital relieved half of patients in postoperative pain, as compared to 20% on placebo, and concluded, "it is obvious that the helping rate of both 60 mgm. and 90 mgm. of pentobarbital is consistently and significantly better than saline."[27] Here, then, we have a gauge to Beecher's understanding of a "significant" gap between drug and placebo—one that is non-trivial and undoubtedly meaningful.

We know Beecher would not consider mere superiority to placebo a valid standard, as he concludes in "The Powerful Placebo," "Clearly, arbitrary criteria of effectiveness of a drug must be set up," and superiority to baseline is non-arbitrary. Presumably he thought the line should be drawn considerably higher. As noted, before the enactment of the Drug Amendments of 1962, modest superiority was not enough to distinguish a drug from placebo in the eyes of many. In the report of a Scottish study of experimental cough cited by Beecher,

codeine outperforms placebo as a suppressant on most measures, yet the investigator finds no evidence "that the action of codeine is any greater than can be explained by the factor of suggestion."[28] He appears to regard codeine as a placebic treatment. In the table of a study of angina treatments from which Beecher draws a 38% rate of placebo efficacy, the test drug (khellin) shows an efficacy of 44%, a difference to which the authors attach no importance.[29]

Minor differences between drug and placebo may mean even less than they appear to. As Beecher observes in another paper published in 1955, "drug-wise" subjects are only too good at detecting test medications by their side effects, thereby defeating the blind on which the trial depends.[30] While he believed only "experienced" subjects are capable of this feat of discernment, we often hear of both investigators and subjects able to guess correctly between drug and placebo at a rate well beyond chance,[31] and some 55 years after "The Powerful Placebo" Irving Kirsch caused consternation with the argument that subjects' ability to identify the telltale side effects of antidepressants accounts for their margin of superiority over placebo in clinical trials. Recognizing nausea and all the rest as signs of a drug, subjects take a paradoxical encouragement from their own discomfort and form the self-fulfilling expectation that they will see a benefit from the drug because it *is* a drug, not a placebo. In essence, they give the drug a placebo bonus.[32] (After all, encouragement is a tonic for the discouragement of depression.) In the face of a phenomenon like this—first glimpsed in the 1950's[33]—a difference of a few points between drug and placebo loses whatever little meaning it originally had.

Microscopic Margins

In addition to the several studies cited in "The Powerful Placebo" are a number actually discussed, among them Wolf and Pinsky's recent trial of mephenesin, the chemical progenitor of meprobamate. Now known mainly for his work with the famous but anonymous "Tom" whose gastric fistula permitted direct observation of the gastric mucous membrane, Stewart Wolf was asked by E. R. Squibb & Sons in the early 1950's to conduct a "critical evaluation of reports in the literature that mephenesin (Tolserol) exerted a specific effect on subjective anxiety and tension and on their objective manifestations."[34] The unnamed reports were presumably clinical impressions of mephenesin administered without placebo controls or any other guard against bias and error: the sort of testimonial that can make a drug look brilliant until the mirage vanishes upon more careful testing. And this is just what happened.

Correcting the laxity of uncontrolled observations, Wolf and Pinsky conducted a double-blind, placebo-controlled trial of mephenesin with 31 outpatients of New York Hospital, only to discover that mephenesin and placebo performed virtually, but not altogether, identically. (It should be noted that the authors do not mention exactly how ratings of the objective and subjective manifestations of anxiety were assigned.) According to a rudimentary graph, about 60% of patients given either mephenesin or placebo were unchanged, while perhaps 35% of those on mephenesin improved, as compared to about 30% of those on placebo. Judging this an unconvincing showing, the authors suggest that mephenesin belongs to the same class as those agents that at first seem promising but are later found to have "no appreciably specific therapeutic action." In short, not having been asked to determine if mephenesin was superior to placebo but to evaluate its merits, Wolf and Pinsky did just that. In the tradition of Gold (and Wolf and Gold overlapped at Cornell after World War II), they find mephenesin tantamount to

placebo despite a nominal superiority to placebo. Mephenesin, it appears, is a placebo with some anxiolytic activity. In the précis of this study in "The Powerful Placebo," Beecher assesses the drug as "therapeutically ineffective."

Reviewing this study occupying little more than two pages in a 1954 issue of the *Journal of the American Medical Association*, the 21st-century reader is struck by the modesty of its procedures and, in particular, by the comparative absence of statistics. Importantly, the authors do not claim that the tenuous superiority of mephenesin to placebo is of no significance because it could have been owing to chance. That the performance of mephenesin virtually mirrored placebo in all three categories of ratings—Better, Same, Worse—for both symptoms and signs told the authors all they needed to know. A similar pattern of findings emerged from the Gold et al. xanthine trial, wherein "most of the patients reported no change, a small number reported that the pain was worse, and about one fourth of the patients in each group reported improvement."[35] Both Gold and the authors of the mephenesin study conclude that the tested treatments are worthless, even though mephenesin showed some advantage over placebo and the angina treatments did not.[36]

Another drug judged a placebo despite scoring slightly higher in a well-conducted trial is the aforementioned meprobamate. Amongst a slew of questionable trials of this drug, the authors of a 1958 review[37] single out one study for its procedural merit and report its relevant findings as follows: 17 subjects rate meprobamate as poor or very poor, 9 as having no effect, and 25 as good or very good; the comparable figures for placebo being 21, 8, and 22. Even though drug outperforms placebo on both ends of the scale, the authors of the review, like

those of the original trial,[38] conclude unequivocally that meprobamate does not differ from placebo.

The question of meprobamate's efficacy was more than academic, in that the drug, under the folksy name of Miltown (later Equanil), quickly became the most prescribed drug in the United States upon its release in 1955.[39] As Mesmerism swept pre-revolutionary Paris, so meprobamate swept the nation, catching the zeitgeist as its predecessor mephenesin did not. By the mid-1950's, the introduction of chlorpromazine, the first effective anti-psychotic drug, had raised the hope that tranquilizers suitable for the general public would relieve neuroses (such as anxiety) just as effectually. Compatible with everyday life and offering benefits theoretically akin to those of drugs too powerful for use by the populace, meprobamate appeared on the scene to answer this hope.

So meteoric was the rise of Miltown/Equanil that trialists were thrown into the curiously futile position of attempting to determine its actual merits after the fact. A comparatively benign interpretation of the craze was offered by Lasagna in 1962:

> It is still unclear as to why the sales of meprobamate skyrocketed so brilliantly. Its trade name of Miltown was, to be sure, unusual, but hardly likely to explain the fantastic spurt. More likely the explanation lies in the felicitous combination of a sedative effect of real comfort to many tense, anxious patients, the lack of serious toxicity in most patients, shrewd advertising, free publicity from radio and TV comics, newspapers, and national magazines—all superimposed on the public's propensity for ready acceptance

> because of the potent and widely heralded tranquilizers introduced a short time before meprobamate but not found to be particularly useful in milder mental illness.[40]

On this reading, which seems just, the drug's intrinsic properties are but one element responsible for its fame and fortune. Borne aloft by its own folklore, Miltown was a placebo with a sedative component. A review of the literature in 1971 cites many studies in which meprobamate showed equivocal efficacy, a pattern suggestive of a placebic drug.[41]

At the dawn of the Miltown phenomenon, Lasagna, discussing the intricacies of the clinical trial, dismissed findings whose only significance is statistical.

> Let us say that as a result of hundreds of observations, one drug is found to be a few per cent more effective than a placebo in relieving cough. The p value for this drug may be < 0.05, but what real meaning does this have? How important is it to know that a drug is "microscopically" better than no drug at all?[42]

A year before, the author himself detected such minimal margins in a well-controlled study. There he found methylparafynol "not significantly different" from placebo in inducing sleep in less than one hour in pre-operative patients, with the drug working in 69% of cases (as judged by observers) and placebo in 66%: a statistically significant but otherwise negligible distinction.[43] In a 1996 interview Lasagna mentioned that 65% of the placebo group and "about 65%" of the controls in this trial fell asleep within an hour, "so we were just studying the relative ease with which people admitted for elective surgery fell asleep."[44]

Clearly Lasagna, like Beecher, saw no point in distinguishing a drug only technically superior to placebo from a placebo. As a pharmacologist he well knew that there are also dosages at which a drug may be tantamount to placebo, and he soon encountered an example in a trial whose uniquely grim significance was not apparent at the time and has gone largely unnoticed to this day. This story too involves the population-wide acceptance of a mild sedative, although the drug in question had just begun to make inroads into the United States; its real center of operations was West Germany, where it was sold not only as a sleeping pill but in combination with aspirin and in other seemingly innocuous preparations, virtually in the manner of a panacea. There was even a liquid formulation for children.[45]

Before it became known that many women who took thalidomide in pregnancy gave birth to severely deformed infants, Lasagna conducted a study in which patients in the Johns Hopkins Hospital received either 100 or 200 mg of the drug or placebo as a sleep aid—the first placebo-controlled study of a drug before its release to the market.[46] At the lower dosage thalidomide yielded 6.2 hours of sleep, a figure Lasagna quite rightly considers "not significantly different" from 6.1 hours with placebo.[47] (However, a Canadian study published in 1961 found 100 mg of thalidomide "more active" than placebo, in that subjects on the drug fell asleep five minutes faster and slept four minutes longer.)[48] Do we then have evidence of patients taking 100 mg of thalidomide—a virtual placebo—as a sleep aid?

A Sept. 1961 letter to the *British Medical Journal* concerning 13 cases of neuropathy after use of Distaval as a hypnotic notes that most of the patients took the "usual" dose of 100 mg.[49] A Dec. 1960 letter to the same journal by a GP in Aberdeenshire reports that four patients complaining of paraesthesia took 100 mg. of Distaval for sleep and that he has found it

"a most effective hypnotic."[50] These cases "may correspond to the experience of other practitioners," he adds. Fourteen months later, a letter from two doctors in Plymouth reported that a woman who gave birth to a deformed infant (who died within minutes) took 100 mg. of thalidomide "nightly" for the first five months of her pregnancy.[51] Investigating a cluster of ten cases in Stirlingshire, Speirs found that in four the mother took 100 mg. at night and in a fifth, three 25 mg. tablets per day; in two additional cases, the mother took between 100 and 200 mg.[52] In a Canadian case reported in 1962, a woman hospitalized during pregnancy was given 100 mg. of thalidomide each evening for three days and 200 mg. on the fourth. She too gave birth to a deformed infant.[53] Finally, in a 1963 study that identified ten thalidomide infants born in British Columbia, in the three instances where the mother's bedtime dose is specified, two of the doses are 100mg. and the third 50 mg.[54]

Thalidomide was branded as Contergan in West Germany, Distaval in Australia, and Kevadon in the United States. In the United States, the drug was used for "investigational" purposes; that is, it was handed out by doctors who then observed its presumed effects—precisely the sort of uncontrolled trial whose worthlessness had already been exposed in the medical literature. In all, an indeterminate number of pregnant women who were prescribed thalidomide for sleep incurred the grave risks of a teratogen in order to reap the benefits of a placebo.

A single 100 mg tablet of thalidomide "appears to be enough to cause severe phocomelia": the flipper-like deformity of the limbs characteristic of the thalidomide syndrome.[55]

Progress vs. Puffery

By his account, Lasagna along with Walter Modell (a former student of Harry Gold) was responsible for the critical inclusion of an efficacy clause in the Drug Amendments of 1962.[56] A generation younger than Gold and Beecher, Lasagna began his work as an investigator just as the sea of confusion was engulfing the pharmaceutical marketplace like an ironically realised vision of plenty.

By 1962 Lasagna was an accomplished trialist and an authority on the placebo effect, with its power to mimic the operation and confuse the evaluation of drugs. In an article that year, he pointed out that a study at Johns Hopkins (where he founded the first division of clinical pharmacology in the United States) confirmed the conventional wisdom that neurotic and hyperkinetic children tend to respond to tranquilizers and mentally defective and antisocial children do not, except that the percentages of those helped were "almost identical" when the subjects were given placebo.[57] Reviewing this well-controlled study, we find that at its pre-specified finish line, one of the test drugs—meprobamate again—proved minimally superior to placebo: 1.3 vs. 1.2 on the study's improvement scale across all groups.[58] Lasagna advocated an efficacy requirement for new drugs, and by efficacy he certainly did not mean a performance like this.[59] Like others before him, he was concerned to decertify drugs (or drug uses) that failed to distance themselves convincingly from placebo, not to certify anything with a hint of efficacy.

At one point in his testimony before the Kefauver committee, Lasagna again brought up approximate placebos. On 19 July 1961, asked by Sen. Roman Hruska to give an example of a useless drug, he replied:

> There are, oh, perhaps a dozen or more oral or buccal enzymes on the market which are supposed to be useful in the treatment of all kinds of diseases, everything from phlebitis to inflammation of all sorts. These products are generally recognized by people who are experts in those fields to be without activity when given in this way. . . . In the few trials that have been done where the medications have been compared with inert agents, both agents have worked to just about the same degree.[60]

Confronted by a pharmacopoeia riddled with untested products, many of which were likely to be approximate placebos, reformers of the theory and practice of drug evaluation attempted as best they could to sort out the valuable from the worthless, the significant from the insignificant. In the Kefauver hearings themselves, Lasagna argued strongly that new drugs similar to existing ones should offer clearly significant, that is, clinically meaningful, improvements.

It was after Lasagna found thalidomide effective at 200 mg for inducing sleep but noted that the "safety margin of the drug in man" is unknown,[61] that its American manufacturer, William S. Merrell, sent samples to more than 1200 doctors for "testing." Presumably Merrell was looking for the rave reviews that so often greet a new drug. The reformers of clinical investigation rejected this notion of a test, arguing that all too many drugs deemed remarkably

effective by doctors and patients alike prove in more unbiased evaluation to be little more than placebo. Unlike those who consider only "positive" trials worthy of publication,[62] these architects of the institution of the well-controlled trial attached importance to the finding that a test drug was *not* appreciably better than placebo. The determination of so many first-rate investigators in the 1950's in particular to reckon honestly with the open secret of the placebo effect belies the caricature of that decade as an era of hypocrisy and repression, the 20th century's Victorian period.

Shortly before Beecher reported that the placebo effect ran at a rate of about 35% across a broad band of clinical trials, Gold observed that a placebo effect "can be demonstrated in about 30 or 40 per cent of all patients with all sorts off disorders. This fact is responsible for the vast literature on drugs that have come into therapeutics with high promise and have left the scene with little loss."[63] In the light of this sorry history, reformers like Gold and Beecher could not be expected to invest much credence in drugs weakly if at all differentiated from placebo. If their aim was "not to prevent progress but to provide a reliable basis for distinguishing it from puffery,"[64] they never suggested that the standard of superiority to placebo provided such a basis. Perhaps because they *were* reformers, not regulators, those who gave life and meaning to the Drug Amendments' key term, "adequate and well-controlled investigations," did not propose or endorse a single yardstick of efficacy for use across the board. A scale of efficacy specific to angina treatments appears in a double-blind 1958 trial, with a "slight response" defined as an increase of exercise tolerance of 10% to 19% over placebo—a gain described as devoid of clinical significance because unlikely to be durable or even reproducible.[65]

The goal of distinguishing progress from puffery seems to have animated the work of the evaluators commissioned by the post-1962 FDA to review the existing pharmacopoeia as directed by law. The outcome of such an exercise had been dimly presaged at the Cornell conference on placebos when a participant mentioned that he once undertook to examine the New York Hospital Formulary of twenty years before. "I remember we gave it a pretty careful searching in the committee, and I think that one third to one sixth of those drugs were inert. What is going to happen to our present hospital formulary when someone goes over it a hundred years from now?"[66] It did not take a hundred years.

As a result of the implementation of the Drug Efficacy Study launched by the FDA in 1966, some 35% of the 3443 drugs reviewed by expert panels were withdrawn from the market.[67] Perhaps Beecher would have experienced déjà vu in seeing his postulated rate of placebo efficacy return as the rate of drugs decertified because they were found lacking in efficacy itself. But the rejection of entire swathes of drugs as a result of the DES is not an allusion to Beecher but the culmination of what was by then a tradition of making informed qualitative distinctions between valid and useless treatments. Recall Evans and Hoyle: "The difference, however, is slight, and does not support the view that papaverine has any real value in continuous treatment."[68] Or Gold et al.: "every *type* of change in pain observed during the use of a xanthine was reproduced in the same individual by a period in which a placebo was used [my emphasis]."[69] Making a similar frankly qualitative judgement, the authors of the vitamin E study observe that "the patients on placebo medication who said they were much better made exactly the same kind of statements regarding the relief of pain as those on alpha tocopherol."[70] For that matter, one day after Lasagna's cited testimony, Walter Modell argued

before the Kefauver committee that "We can make the useful drugs both less dangerous and more efficient by weeding out the useless, the ineffective, and the duplicates."[71] The very categories used by the DES reviewers—Effective, Probably Effective, Possibly Effective, Ineffective—were necessarily qualitative.

While the case has been made that the post-1962 culling of the nation's pharmacopoeia represents "a veritable slaughter of the innocents,"[72] there is no returning to the days when clinical impression was enough to establish a treatment. Clinical impression convinced many that the xanthines were effective treatments of cardiac pain even after their effect vanished upon testing.[73] In the very year of the Drug Amendments, Lasagna noted that a certain anabolic steroid appeared to keep many psychotic patients from soiling their bed, yet when the wonder drug was properly tested, it fell short of placebo.[74]

The difference between clinical impression and clinical trial is illustrated by the case of the first antidepressant, imipramine, whose discoverer was so impressed by its performance that he all but declared that it could cure homosexuality.[75] In the report of a trial of imipramine three years later, Lasagna and colleagues cautioned that "The placebo success rate appears sufficiently high to require anti-depressant drug trials of considerable size in order to provide convincing evidence of drug differences [sic]."[76] In germ in this comment lie untold numbers of future antidepressant trials, all struggling with the placebo effect and many clustering around the line of non-efficacy.[77] Of course, if it takes a trial of considerable size to tease out a difference between drug and placebo, the difference is probably not great. As Lasagna wrote in 1964, "One must not lose track of the fact that we are interested in biologically significant

differences, not merely statistically significant ones." He then quoted with approval a statistician's comment, "A difference is not a difference at all unless it really matters."[78]

Interviewed some thirty years later, Lasagna made the same distinction between two kinds of significance,[79] as if his point had not been taken. Indeed it had not. Over the intervening decades, many a clinical trial had come to resemble an elaborate mechanism for eliciting a statistical distinction between drug and placebo[80]—the sort of distinction without a difference that reformers of clinical investigation from the 1930's on judged meaningless.

---

[1] Robert Lowes, *Medscape*, 28 Sept. 2012; https://www.medscape.com/viewarticle/771795.

[2] Walter Modell and Raymond Houde, "Factors Influencing Clinical Evaluation of Drugs," *JAMA* 167 (1958): 2190-99.

[3] On "substantial evidence", see Stewart Justman, "The Weight of a Term: "Substantial Evidence" and Buried Data," *Perspectives in Biology and Medicine* 61 (2018): 201-14.

[4] Joseph Gabriel, "The Testing of Sanocrysin: Science, Profit, and Innovation in Clinical Trial Design, 1926-31," *Journal of the History of Medicine and Allied Sciences* 69 (2013): 604-32.

[5] Daniel Carpenter, *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA* (Princeton: Princeton University Press, 2010), 474n. On the low standard of efficacy established by law, see Jonathan Darrow, "Pharmaceutical Efficacy: The Illusory Legal Standard," *Washington and Lee Law Review* 70 (2013): 2073-2136.

[6] On the idolatrous worship of statistical significance, see Stephen Ziliak and Deirdre McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives* (Ann Arbor: University of Michigan Press, 2011).

[7] Richard Lipton, David Dodick, Jessica Allani et al., "Effect of Ubrogepant vs Placebo on Pain and Most Bothersome Associated Symptom in the Acute Treatment of Migraine: The Achieve II Randomized Clinical Trial," *JAMA* 322 (2019): 1887-98.

[8] Darrow, *op. cit.* (note 5), 2076.

[9] This is not to suggest that the FDA does not care about low efficacy. In the singular case of a drug for the prevention of prostate cancer (finasteride), FDA reviewers noted with concern that it lowered the incidence of detected disease by only 14% (compared to placebo) when certain clerical errors in the classification of biopsies were corrected. The original, published figure was 24.8%. Set against this diminished gain was an apparent increase in the incidence of high-grade cancer associated with the drug in question. See the Briefing Document for the ODAC (Oncologic Drugs Advisory Committee) meeting of December 1, 2010.

[10] Placebo is still sometimes used to control for the natural history of disease. No one proposes that in the Prostate Cancer Prevention Trial, the placebo group saw any benefit from placebo. Ian Thompson, Phyllis Goodman, Catherine Tangen et al., "The Influence of Finasteride on the Development of Prostate Cancer," *New England Journal of Medicine* 349 (2003): 215-24.

[11] Harold Diehl, "Medicinal Treatment of the Common Cold," *Journal of the American Medical Association* 101 (1933): 2042-49.

[12] H. Diehl, A. Baker and D. Cowan, "Cold Vaccines: An Evaluation Based on a Controlled Study," *Journal of the American Medical Association* 111 (1938), 1168-73. Without referring to Diehl, Gaddum later found that in the typical trial on the common cold, the test drug yields good results and the dummy "almost equally good results". J. Gaddum, "Walter Ernest Dixon Memorial Lecture," *Proceedings of the Royall Society of Medicine* 47 (1954): 195-204.

[13] William Evans and Clifford Hoyle, "The Comparative Value of Drugs Used in the Treatment of Angina Pectoris," *QJM: An International Journal of Medicine* 2 (1933): 311-38.

[14] Gaddum, *op. cit.* (note 12).

[15] Seymour Rinzler, Hyman Bakst, Zachary Benjamin et al., "Failure of Alpha Tocopherol to Influence Chest Pain in Patients with Heart Disease," *Circulation* 1 (1950): 288-93.

[16] Harry Gold, Nathaniel Kwit and Harold Otto, "The Xanthines (Theobromine and Aminophylline) in the Treatment of Cardiac Pain," *Journal of the American Medical Association* 108 (1937): 2173-79.

[17] Arthur Shapiro and Elaine Shapiro, *The Powerful Placebo: From Ancient Priest to Modern Physician* (Baltimore: Johns Hopkins University Press, 1997): 143-44; 153-54.

[18] Suzanne White, "Medicine's Humble Humbug: Four Period in the Understanding of the Placebo," *Pharmacy in History* 27 (1985), 51-60. Reflecting in a 1968 interview on the state of things around the time of the xanthine study, Gold remarked that "The reputation of a placebo was so bad that doctors in general had no use for its name." See Shapiro and Shapiro, *op. cit.* (note 15), 141.

[19] Harold Wolff et al., "Conferences on Therapy," *New York State Journal of Medicine* 46 (1946): 1718-27.

[20] *Ibid.*, 1726.

[21] Medical Research Council, "Streptomycin Treatment of Pulmonary Tuberculosis," *British Medical Journal*, 30 Oct. 1948: 769-82. The MRC"s 1944 trial of patulin for the common cold did employ placebo, however.

[22] Theodore Greiner, Irwin Bross and Harry Gold, "A Method for Evaluation of Laxative Habits in Human Subjects," *Journal of Chronic Diseases* 6 (1957): 244-55.

[23] Walter Modell, *The Relief of Symptoms* (Philadelphia: W. B. Saunders, 1956), 160.

[24] Henry Beecher, "The Powerful Placebo," *JAMA* 159 (1955),: 1606-09.

[25] Gunver Kienle, and Helmut Kiene, "The Powerful Placebo Effect: Fact or Fiction?", *Journal of Clinical Epidemiology* 50 (1997): 1311-18.

[26] Cf. Henry Beecher, *Measurement of Subjective Responses* (New York: Oxford University Press, 1959), 65.

[27] Arthur Keats and Henry Beecher, "Pain Relief with Hypnotic Doses of Barbiturates and a Hypothesis," *Journal of Pharmacology and Experimental Therapeutics* 100 (1950): 1-13.

[28] B. R. Hillis, "The Assessment of Cough-Suppressing Drugs," *Lancet*, 21 June 1952: 1230-35. Cf. Beecher's assessment of codeine as an approximate placebo in *Measurement of Subjective Responses* (note 26), e.g. 423.

[29] Theodore Greiner, Harry Gold, McKeen Cattell 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)," *American Journal of Medicine* 9 (1950): 143-55.

[30] Henry Beecher, "Appraisal of Drugs Intended to Alter Subjective Responses," *Journal of the American Medical Association* 158 (1955): 399-401.

[31] See e.g. Irving Kirsch, "The Use of Placebos in Clinical Trials and Clinical Practice," *Canadian Journal of Psychiatry* 56 (2011): 191-92.

[32] Irving Kirsch, *The Emperor's New Drugs: Exploding the Antidepressant Myth* (New York: Basic, 2010).

[33] Louis Lasagna, "The Controlled Clinical Trial: Theory and Practice," *Journal of Chronic Diseases* 1 (1955), 353-67; A. May, "The Use of Tranquillizers," *Postgraduate Medical Journal* 33 (1957),

505-10; Stewart Wolf, "The Pharmacology of Placebos," *Pharmacological Reviews* 11 (1959), 689-704; F. Letemendia and A. Harris, "The Influence of Side-Effects on the Reporting of Symptoms," *Psychopharmacologia* 1 (1959): 39-47.

[34] Stewart Wolf and Ruth Pinsky, "Effects of Placebo Administration and Occurrence of Toxic Reactions," *JAMA* 155 (1954): 339-41.

[35] Gold, Kwit and Otto, *op. cit.* (note 16): 2176.

[36] A run-in to the xanthine study did show a treatment to be nominally superior to placebo. Some 82% of patients given glyceryl trinitrate tablets and 78% given placebo reported relief, a patently trivial difference.

[37] Victor Laties and Bernard Weiss, "A Critical Review of the Efficacy of Meprobamate (Miltown, Equanil) in the Treatment of Anxiety," *Journal of Chronic Diseases* 7 (1958): 500-19.

[38] M. J. Raymond, C. J. Lucas, M. L. Beesley, B. A. O'Connell, "A Trial of Five Tranquillizing Drugs in Psychoneurosis," *British Medical Journal*, 13 July 1957: 63-66.

[39] Laties and Weiss, *op. cit*. (note 37): 500.

[40] Louis Lasagna, *The Doctors" Dilemmas* (New York: Harper & Brothers, 1962), 228.

[41] David Greenblatt and Richard Shader, "Meprobamate: A Study of Irrational Drug Use," *American Journal of Psychiatry* 111 (1971): 1297-1303. The authors call attention to the drug's toxic potential, however. The successor drug, diazepam (Valium), showed a similar profile of equivocal efficacy. See Arthur Shapiro, Elmer Struening, Elaine Shapiro and Barry Milcarek, "Diazepam: How Much Better Than Placebo?", *Journal of Psychiatric Research* 17 (1983): 51-73.

[42] Lasagna, "The Controlled Clinical Trial: Theory and Practice" (note 33): 364.

[43] Louis Lasagna, "A Comparison of Hypnotic Agents," *Journal of Pharmacology and Experimental Therapeutics* 111 (1954): 9-20.

[44] *The Psychopharmacologists II: Interviews by Dr. David Healy* (London: Chapman & Hall, 1998), 138.

[45] Helen Taussig, "The Thalidomide Syndrome," *Scientific American,* Aug. 1962: 29-35.

[46] David Healy, Derelie Mangin and Kalman Applbaum, "The Shipwreck of the Singular," *Social Studies of Science* 44 (2014): 518-23. Lasagna's study appeared in the June 1960 issue of *Journal of Chronic Diseases* (see note 48), and William S. Merrell applied for FDA authorization in September. See Peter Temin, *Taking Your Medicine: Drug Regulation in the United States* (Cambridge, Mass.: Harvard University Press 1980), 123.

[47] Louis Lasagna, "Thalidomide—A New Nonbarbiturate Sleep-Inducing Drug," *Journal of Chronic Diseases* 11 (1960): 627-31.

[48] T. A. Ban and L. Schwartz, "Comparative Clinical Study of Three Hypnotic Drugs," *Canadian Medical Association Journal* 84 (1961): 1259.

[49] Pamela Fullerton and Michael Kremer, "Neuropathy After Intake of Thalidomide (Distaval), *British Medical Journal* 2.5256 (30 Sept. 1961); 855-58.

[50] A. Leslie Florence, "Is Thalidomide to Blame?" *British Medical Journal* 2.5217 (31 Dec. 1960): 1954.

[51] A. Willman and J. Dumoulin, "Distaval (Thalidomide) and Foetal Abnormalities," *British Medical Journal* 1.5276 (17 Feb. 1962): 444.

[52] A. Speirs, "Thalidomide and Congenital Abnormalities," *Lancet* 1962;279.7224: 303-05.

[53] A. Rodin, L. Koller, J. Taylor, "Association of Thalidomide (Kevadon) With Congenital Anomalies," *Canadian Medical Association Journal* 86 (1962): 744-46.

[54] Anthony Larsen, Leonard Hole and Barbara Mackenzie, "Congenital Deformities Associated with the Use of Thalidomide During Pregnancy, British Columbia, 1961-72," *Canadian Journal of Public Health* 54 (1963): 505-08.

[55] Taussig, *op. cit.* (note 45), 33.

[56] *The Psychopharmacologists II*, *op. cit.* (note 44), 140.

[57] Louis Lasagna, "Clinical Trials: Nuisance or Necessity?", *Methods of Information in Medicine* 1 (1962): 79-82.

[58] Leon Cytryn, Anita Gilbert and Leon Eisenberg, "The Effectiveness of Tranquilizing Drugs Plus Supportive Psychotherapy in Treating Behavior Disorders of Children: A Double-Blind Study of Eighty Outpatients," *American Journal of Orthopsychiatry* 30 (1960): 113-29.

[59] For Lasagna's comments on the hearings and associated issues see Louis Lasagna, "Congress, the FDA, and New Drug Development: Before and After 1962," *Perspectives in Biology and Medicine* 32 (1989): 322-43.

[60] *Hearings Before the Subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, United States Senate, Eighty-seventh Congress, First Session, Pursuant to S. Res. 52 on S. 1552, July 5, 6, 18, 19, 20, 21 and 25, 1961*, 285. See

https://www.google.com/books/edition/Drug_Industry_Antitrust_Act/h6MTAAAAIAAJ?hl=en&gbpv=1.

[61] Lasagna, "Thalidomide" (note 47): 630.

[62] Stuart Ritchie, *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth* (New York: Metropolitan, 2020).

[63] Harry Gold, "Conference on Therapy: How to Evaluate a New Drug," *American Journal of Medicine* 17 (1954): 722-27.

[64] Harry Marks, *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900-1990* (Cambridge: Cambridge University Press, 1997), 150.

[65] Joseph Riseman, George Altman and Sidney Koretsky, "Nitroglycerin and Other Nitrites in the Treatments of Angina Pectoris," *Circulation* 17 (1954): 22-39.

[66] Wolff et al., Cornell Conference on Therapy (note 19): 1719.

[67] Edward Shorter, "Looking Backward: A Possible New Path for Drug Discovery in Psychopharmacology," *Nature Reviews | Drug Discovery* 1 (2002): 1003-06.

[68] Evans and Hoyle, *op. cit.* (note 13): 327.

[69] Gold, Kwit and Otto, *op. cit*. (note 16): 2178.

[70] Rinzler, Bakst, Benjamin et. al., *op. cit.* (note 15): 291.

[71] Prepared statement submitted to the Kefauver committee on July 20, 1961. See *Hearings Before the Subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, United States Senate, Eighty-seventh Congress, First Session* (note 56), 1099.

[72] Shorter, *op. cit.* (note 63): 1006.

[73] Greiner, Gold, Cattell et al., *op. cit.* (note 29).

[74] Lasagna, "Controlled Trials: Nuisance or Necessity" (note 53).

[75] Roland Kuhn, "The Treatment of Depressive States with D-22355 (Imipramine Hydrochloride)," *American Journal of Psychiatry* 115 (1958): 459-64.

[76] Rudolf Höhn, Gertrude Gross, Martin Gross and Louis Lasagna, "A Double-Blind Comparison of Placebo and Imipramine in the Treatment of Depressed Patients in a State Hospital," *Journal of Psychiatric Research* 1 (1961): 76-91.

[77] See e.g. Arif Khan, Shirin Khan and Walter Brown, "Are Placebo Controls Necessary to Test New Antidepressants and Anxiolytics?", *International Journal of Neuropsychopharmacology* 5 (2002): 193-97.

[78] Louis Lasagna, "On Evaluating Drug Therapy" in *Drugs in Our Society*, ed. Paul Talalay (Baltimore: Johns Hopkins University Press, 1964), 99.

[79] "If you have a trivial difference and you study enough patients, you come up with biologically insignificant but statistically significant figures." See *The Psychopharmacologists II* (note 44), 165. On the confusion of statistical for practical significance, see Lance Waller, "A Note on Harold S. Diehl, Randomization, and Clinical Trials," *Controlled Clinical Trials* 18 (1997): 180-83.

[80] David Healy, *Pharmageddon* (Berkeley: University of California Press, 2012).