Placebo orthodoxy and the double standard of care in multinational clinical research

Maya J. Goldenberg

Theoretical Medicine and Bioethics
Philosophy of Medical Research and Practice

ISSN 1386-7415
Volume 36
Number 1

Theor Med Bioeth (2015) 36:7-23
DOI 10.1007/s11017-015-9317-9

Springer

Theor Med Bioeth (2015) 36:7–23
DOI 10.1007/s11017-015-9317-9

Placebo orthodoxy and the double standard of care in multinational clinical research

Maya J. Goldenberg

Published online: 21 January 2015

Abstract It has been almost 20 years since the field of bioethics was galvanized by a controversial series of multinational AZT trials employing placebo controls on pregnant HIV-positive women in the developing world even though a standard of care existed in the sponsor countries. The trove of ethical investigations that followed was thoughtful and challenging, yet an important and problematic methodological assumption was left unexplored. In this article, I revisit the famous "double standard of care" case study in order to offer novel consideration of the *placebo orthodoxy* that underlies much of the ethical debate. This majority view found in medical research is that placebo-controlled trials are methodologically superior to comparative trials that use active controls. I challenge this orthodoxy and argue that lives were unnecessarily lost in these trials as a result. Furthermore, current HIV research on vaccines and microbicides is now poised to repeat the error of subscribing to the placebo orthodoxy.

Keywords Placebo · Randomized controlled trials · Multinational clinical trials · Developing world bioethics · HIV/AIDS · Active-controlled trials

Introduction

In this article, I revisit a famous case that drew public attention and galvanized academic bioethics in the late 1990s. The case concerned a series of zidovudine (AZT) trials to reduce mother-to-child transmission of HIV in the developing world. The controversy centered on the ethics of running placebo-controlled trials (PCTs) when an effective treatment, available in the developed world, already existed. This came to be known as the "double standard of care" debate. I want to bring novel

M. J. Goldenberg (✉)
Department of Philosophy, University of Guelph, Guelph, ON N1G 2W1, Canada
e-mail: mgolden@uoguelph.ca

consideration to this well-examined case study, namely, the *placebo orthodoxy* that underlies much of the ethical debate. This is the majority view found in medical research, i.e., that PCTs are methodologically superior to trials that employ active agents as their controls. I will challenge this orthodoxy and argue that lives were unnecessarily lost in these trials as a result. While there may be circumstances in which placebo trials are scientifically advantageous over active-controlled trials in second-generation research, the orthodoxy problematically makes the case for its methodological superiority *in general*. I further demonstrate that future research is now poised to repeat the error of subscribing to the placebo orthodoxy.

I will begin by reviewing the case, which enlisted over 10,000 pregnant women in 11 countries throughout the developing world. I will then, in the second section, characterize the bioethics community's responses to this case, which, I will argue, problematically left the pivotal methodological question unexplored. In the third section, I will turn to the methodological debate regarding placebo versus active controls in clinical trials. This review will allow me to reject the placebo orthodoxy, at which point I can return, in the fourth section, to the case study in order to assess whether there were grounds for using active controls instead of placebo controls *in this context* (as there is no general argument to be made for active controls in all circumstances). Finally, in the fifth section, I will review where we are now in terms of attitudes toward placebo use in global HIV clinical research and in research ethics scholarship.

Perinatal HIV transmission trials in developing countries

In an April 1997 open letter to then-secretary of health and human resources, Donna Shalala, the Public Citizen Health Research Group raised a high profile alarm about what they saw as serious ethics violations taking place in government sponsored clinical trials. They wrote:

> Unless you act now, as many as 1,002 newborn infants in Africa, Asia and the Caribbean will die from unnecessary HIV infections they will contract from their HIV-infected mothers in nine unethical research experiments funded by your Department through either the National Institutes of Health (NIH) or the Centers for Disease Control and Prevention (CDC). [1]

The charge was controversial because it was launched against what was seen by American agencies, international governments, and health research agencies sponsoring similar trials as a highly ethical effort to translate important recent HIV research findings into simpler and cheaper regimens suited for resource-poor regions of the developing world that were being crushed by the impact of HIV/AIDS. Three years earlier, an NIH-funded PCT of an AZT regimen conducted in the United States and France had been stopped early because of its impressive showing of a two-thirds reduction of vertical transmission of HIV from pregnant women to their offspring (23 % in placebo group vs. 8 % in AZT group) [2]. The women in the placebo group were promptly placed on AZT therapy. The regimen, which came to be known by its grant number "076," was endorsed by the Public Health Service

and established as the standard of care. But it was quickly realized that 076 would not be viable in the areas of the world that needed this intervention most, as the regimen was costly and required significant health care infrastructure including early prenatal care, intravenous use during delivery, and neonatal care. The World Health Organization (WHO) thereby requested that UNAIDS coordinate international research efforts to develop simpler and less costly interventions for use in the developing world [3]. In the United States, the NIH and CDC had joined that effort and sponsored the trials now in question.

Public Citizen's objection was not regarding the studies' goal to establish pragmatic regimens suitable for resource-poor environments. Public Citizen and numerous others agreed that health research could be a catalyst for health equity [4] and that multinational health research could be morally praiseworthy for building knowledge *beneficial to the host community*. These mother-to-child transmission studies fulfilled those minimum criteria.

Instead, Public Citizen objected to the design of those trials, as they all enlisted placebo controls. They charged that using placebos instead of the reliable standard of care contravened the Declaration of Helsinki (Article II.3) [5] and set a dangerous and exploitative double standard for trials conducted in the developing world. They argued compellingly that these trials would never be approved in a Western setting, where research participants would be guaranteed access to *either* the experimental regimen *or* the standard of care, Protocol 076. Why, they asked, was the same imperative to avoid preventable deaths among research subjects not applicable in the developing world? They urged Secretary Shalala to launch an investigation into how those trials had gained ethics approval.

The double standard of care was additionally dubious, Public Citizen argued, because it wrongly presumed the methodological superiority of PCTs over active-controlled trials that used the standard of care as its comparator. The promise of better science presumably justified the use of placebos, as the host community would benefit more fully from this research *in the long run.* In a follow-up letter to then-president Bill Clinton, Public Citizen argued against WHO and CDC recommendations that 076-translation trials are best designed using placebo controls [6]. They countered that active-controlled equivalency trials (ACETs)—trials employing active controls to determine whether a new intervention is no worse or no better than the control[1]—are methodologically preferable [7]. They pressed President Clinton to order those studies to be redesigned as equivalence studies.

Public Citizen subsequently developed their position on ACETs in a widely read *New England Journal of Medicine* (NEJM) article authored by Peter Lurie and Sidney Wolfe [8]. The journal commissioned a response piece by the heads of the NIH and the CDC, Harold Varmus and David Satcher, which offered a defense of the scientific preferability of PCTs and, hence, their ethical acceptability in the face of a deadly epidemic [3]. The exchange highlighted a lively debate—the terms already known in philosophy of science circles—over placebo versus active controls in clinical trials.

[1] PCTs, by contrast, are typically *superiority* trials, which determine the difference between two interventions.

Bioethics response

Public Citizen's letters were followed by a highly effective media campaign—generating considerable press [9–12], public attention, a congressional hearing on bioethics, and support from thought leaders, like *NEJM* editor Marcia Angell, who provocatively likened the perinatal transmission trials to the notorious Tuskegee syphilis studies [13].

The publicity surrounding those studies galvanized academic bioethics. Two prominent bioethics journals— *Bioethics* and the *Hastings Center Report*—devoted whole issues to the debate. The case also spawned the new sub-discipline of "developing world bioethics." The specifics of those trials became a springboard for numerous important ethical inquiries. They included:

(1) dissecting the vagaries of the current codes governing research using human subjects [14, 15];
(2) developing governing principles of "reasonable availability" and "fair benefits" for outsourced clinical trials [16–20];
(3) considering whether ethical norms are universal or contextual [20, 21];
(4) considering how informed consent could be properly obtained among desperately poor individuals with very limited health care options [22, 23]; and
(5) raising questions about social justice and global inequalities [4, 24, 25].

My interest is in the feature of the debate that was *not* taken up: the methodological question about the presumed superiority of PCTs. This omission occurred despite the ample attention given to this matter in the *NEJM* exchange between Public Citizen and the NIH-CDC; those papers were reproduced in numerous bioethics anthologies.

Instead, many bioethicists accepted the majority view of the superiority of placebo controls over active controls without argument. Carol Levine, for instance, affirmed the PCT's gold standard status by asserting that while other methodologies like equivalence studies, case controlled studies, and historical controls, are available, "these generally provide less conclusive data or answer more narrowly constructed questions" [26, p. 44]. The temptation to accept arguments by authority—even the majority view—should have been resisted.

Accepting the placebo orthodoxy frames the ethics of placebo debate as a *trade-off between science and ethics*. Bioethicists need to worry whether this is the correct framing of the debate. Placebo trials are taken to be "scientifically sound but (often) ethically unacceptable" for investigating second generation treatments, whereas active-controlled studies are "ethically sound but scientifically not reliable" [27]. Some commentators accepted ethics as a necessary constraint on scientific practice. Others rejected the ethics-science conflict by challenging the legitimacy of the ethical guidelines. Robert Levine challenged the Declaration of Helsinki's relevancy because of the scientific limits it placed on clinical research by delimiting the acceptable use of placebos [28]. He was influential in changing the Declaration to be more permissive regarding PCTs. Similarly, Frank Miller and Howard Brody

rejected clinical equipoise as a guiding principle for clinical research on the same grounds [29].

Other bioethicists avoided the methodological question by acknowledging the debate with no effort at remediation. Ruth Macklin, for instance, devotes only one paragraph of her book-length monograph on the double standard of care in multinational research ethics to the methodological debate over placebo controls. She writes:

> the broad consensus regarding the need for research in developing countries is a thin reed compared to the deep disagreement about acceptable research designs. Articles continue to appear in the scientific literature, not only about the ethics of placebo-controlled studies, but also about their necessity and scientific merit. Two epidemiologists who have opposed placebo-controlled trials contend that "most of the scientific arguments are either wrong or distorted." Of course opponents disagree. [30, p. 30]

She offers no further discussion of the issue in the 230 pages that follow this mention in chapter one of the book.

An alternate response came from bioethicists who refused to address the methodological debate for reasons ethical and political. George Annas and Michael Grodin argued that the question of how a study should be conducted does not answer the question of whether they should be done at all. Additionally, scientific knowledge does not enable actual intervention into complex global health epidemics [31].

Yet, this parsing of the problem as an "ethics" problem merely pushes back the methodological question that requires attention. Both sides agreed that placebos should be avoided when possible, especially when deaths will result. Placebos were defended in this context because the severity of the AIDS epidemic called for exceptionalism in order to generate the best scientific data. The Declaration of Helsinki and the CIOMS guidelines [32] have since been revised in light of this belief. Those documents no longer forbid placebo use for second generation research. The 2008 revision of the Declaration of Helsinki reads: "Where for compelling and scientifically sound methodological reasons the use of placebo is necessary to determine the efficacy or safety of an intervention" [33, par. 33]. One needs to know what those reasons might be and when they might arise. This language, and some of the bioethics commentary on this subject, seems to take the methodological question to be straightforward and/or easily determined by clinical researchers. This position is incorrect—as was seen in the response to results from the first of the AZT translational trials, a PCT evaluating short-course AZT in Thailand. A dramatic 51 % reduction in vertical transmission was observed between the experimental arm and the control group [34]. Both proponents and opponents of placebo use in these trials claimed their views had been vindicated. The former maintained "that such significant results could only have been achieved with placebo arms; the opponents saw the results as evidence that placebos were never needed in the first place" [26, p. 43].

Methodological debate: placebo vs. active-controlled trials

Among the many commentators weighing in on the notorious AZT trials, Public Citizen was alone in denying that PCTs provided the quickest access to the most reliable information. Even those who shared their conviction that the trials were unethical presumed a trade-off between science and ethics, with ethical considerations trumping the dictates of science. To make a stronger case that there is *no* trade-off when active controls are employed, one needs to turn to the philosophy of science and epidemiology literature on placebo versus active controls in clinical trials.

I will now review the arguments that effectively undermine the majority view on placebos—the "placebo orthodoxy." With no grounds for thinking that PCTs are superior to ACETs *in general,* one can then turn back to the case study to consider whether PCTs were methodological preferable *in that context*, or whether Public Citizen was correct to propose (on methodological grounds) that these studies should have been redesigned as equivalence studies to compare shorter AZT regimens to the 076 standard of care.

Lurie and Wolfe [8] were not the first to challenge the presumed methodological superiority of PCTs. Three years earlier, epidemiologists Kenneth Rothman and Karin Michels [35] outlined a similar argument. Their target was not multinational clinical research, but the research and regulatory norm in general. They presented an ethical argument against PCTs that hinged largely on denying its purported methodological superiority. A flurry of response letters similarly defended the ethical appropriateness of placebo controls on methodological grounds—that they provided the most reliable data, which are more beneficial to patients than unreliable data, and were thereby ethical.

Rothman and Michels had hit a nerve by challenging what was later called "placebo mania" by Rothman [36] and the "placebo orthodoxy" by Benjamin Freedman et al. [37]. In my review of the literature on the subject, "placebo orthodoxy" upholds the methodological superiority of PCTs over ACETs on these four grounds:

(1) placebo controls are necessary to establish a baseline measure or null point for measuring the treatment effect meaningfully;
(2) PCTs are simpler, quicker, more efficient than ACETs [3];
(3) PCTs require fewer participants than ACETs to establish statistical significance [38];
(4) ACETs lack assay sensitivity [39].

Regarding (1), the "baseline" argument problematically presumes that the placebo effect is stable. It is well known, however, that the placebo effect varies widely from study to study [40, 41]. Placebos can even be manipulated as an experimental variable. For example, different colored placebos often have different effects for the same condition. Knowing this, which colored placebo (and subsequent placebo effect) should one count as the baseline against which one

may discern a drug's biological effectiveness? How does one generalize a study's findings? [37, 42]. While a positive finding in a well conducted PCT can demonstrate that the experimental intervention was superior to *that* placebo, it is possible that the drug would not prove superior to other matched placebos.

Further, placebos could only act as a stable reference in clinical trials *if* they were really inert substances or "nothing." Yet, placebos, just like pharmaceuticals, may possess pharmacological profiles, which include peak times, carry-over and cumulative effects, toxicities, and side effects [42]. This is especially the case with "active" placebo controls, placebos that have been designed to mimic the side effects of the experimental agent in order to ensure proper blinding in the study. When one considers all this variability in the placebo effect, it is difficult to see how a placebo can be taken as a null point permitting direct measure of meaningful treatment effects.

Regarding (2), PCTs are widely endorsed on the grounds that they are simpler and can be conducted more quickly than studies using active controls. The "efficiency argument" was strongly promoted by the CDC and NIH [3, 43]. Yet this position only holds at face value. A study with a stable and inert comparator *would* be easier to design and run than a study using a complex regimen as the control. But the previous comments about placebos deny that antecedent.

Furthermore, a PCT with this hypothetically simple placebo would likely suffer from serious methodological difficulty: it could easily become unblinded as the placebo would not create an effect that is qualitatively similar to the active agent under investigation [44]. Data from unblinded PCTs are distorted by subjects' and researchers' expectations, and subjects who suspect that they are receiving placebo often drop out of trials.

While researchers can insist that their studies are "simpler," the scientific advantage of "simple studies" is illusory once blinding is compromised. The data suggests that this happens frequently [45, 46]. There have been earnest efforts to confront and reduce the problem of unblinding in PCTs through methodological innovation (a) by adding new arms to the study— for example, "in arm three, subjects would be given placebo and told it is placebo; in arm four, subjects would be given the drug and told it is placebo; in arm five, subjects would be given no treatment," and so on [37]—or (b) by creating "active" placebo controls that mimic the side effects of the experimental agent [37, 47]. These corrective efforts are promising, but they take away from the original merit ascribed to PCTs. These studies are no longer clearly simpler, quicker, and smaller than ACETs.

Regarding (3), Federal Drug Administration (FDA) scientist Robert Temple defends PCTs because they require fewer participants than ACETs in order to demonstrate a statistically significant effect [38]. This claim is correct. Statistical significance can be obtained in small studies if the effect estimate is strong enough. Placebo controls can make the effect of a new drug appear larger, thereby permitting statistical significance to be achieved with fewer participants in the trial than a measure of presumably smaller effect when comparing two active treatments. The FDA relies heavily on statistical significance in judging the efficacy of new drugs [38], however this practice has been criticized on the grounds that statistical

significance is a poor measure of treatment efficacy [48–52]. The significance of an association depends on the strength of the association and its statistical variability. Yet, the critics argue, equal consideration of those two characteristics can mislead: "A weak effect can be 'significant' if there is little statistical variability in its measurement, whereas a strong effect may not be 'significant' if there is substantial variability in its measurement" [35, p. 396]. Instead, "only the strength of the effect should be fundamental to the decision about approval of the drug" [35, p. 396] because this is what patients and providers want to know. Statistical variability should ideally be reduced nearly to zero when assessing the magnitude of a drug effect so that random error does not influence the assessment. The best way to reduce statistical variability, however, is to conduct large studies. In the end, ACETs are not substantially larger than *reasonably designed* PCTs, i.e., those designed to provide meaningful information about the efficacy of a new drug rather than merely establish statistical significance.

Regarding (4), Temple and Ellenberg argue that ACETs lack *assay sensitivity* while PCTs do not [39]. This difference with respect to assay sensitivity, a trial's ability to distinguish effective treatment from ineffective treatment, stems from the presumably different *inferential approaches* to interpreting results in the two trial designs. Temple and Ellenberg argue that a positive finding in a PCT demonstrates the efficacy of the experimental treatment. One is justified to conclude that the experimental agent works. The trial has assay sensitivity—no further information is required to justify this claim. A finding of equivalence in an ACET, on the other hand, can mean two things—either both are effective or neither are effective. Equivalence studies lack assay sensitivity to justify one interpretation as correct. One needs to rely on external information—past trials and mechanistic reasoning—to justify the claim that the active control is effective and infer that the experimental agent is therefore also effective [39]. But, Temple and Ellenberg argue, one may err in one's inference. Furthermore, that an active control has a good past record of performance in prior PCTs does not guarantee that the active control worked in *this* trial. To have that reassurance, one would need to add a placebo arm.

Both Anderson [53] and Howick [47] undermine the assay sensitivity argument by demonstrating that placebo trials suffer from the same assay sensitivity problem as equivalence studies. More precisely, both authors argue that the inferential structure for establishing treatment efficacy in PCTs is not so different from ACETs. When one gets a positive finding in a PCT, it can mean two things: either the experimental agent is efficacious *or* the placebo effect is very weak (viz. weaker than an experimental agent's weak effect). Variability in placebo effects (discussed earlier) means that one cannot draw reliable conclusions from a single study. Instead, one needs to rely on external evidence, like other similar studies, to generate a legitimate inference regarding the efficacy of an experimental treatment. This is why medical research commonly relies on meta-analysis of multiple comparable studies to draw conclusions and make practice or policy recommendations. For both ACETs and PCTs, one needs to draw from the totality of evidence and "do the best we can to justify our assumptions and suspend judgment

otherwise" [53].[2] This complex exercise of inferential reasoning cannot be bypassed by the PCT. The same cautious and critical approach is needed for legitimately interpreting study results.

These challenges undermine the placebo orthodoxy's privileging of PCTs over ACETs as a general pronouncement on clinical trial methodology. They do not, however, support the converse that ACETs are generally scientifically advantageous for second generation research. Instead, they allow for the particular research context to determine the most suitable methodology.

Inference from external data in multinational AZT trials

With the placebo orthodoxy undermined, no case can be made for the preferability of PCTs *in general*. I now return to my case study to assess whether *the specifics of this case* justified the use of PCTs *in this context.* The answer depends on two considerations: (i) asking the right research question; (ii) reviewing the totality of evidence.

Supporters of ACETs often argue that this design has the merit of answering the question that patients and providers really want to know—how does the experimental treatment compare to the established/available treatment? Austin Bradford Hill's famous claim is often invoked:

> Is it ethical to use a placebo? The answer to this question will depend, I suggest, upon whether there is already available an orthodox treatment of proved or accepted value. If there is such an orthodox treatment, the question will hardly arise, for the doctor will wish to know whether a new treatment is more, or less, effective than the old, not that it is more effective than nothing. [55, p. 1043]

But Varmus and Satcher [3] defended the appropriateness of placebos similarly. With the standard of care in these developing countries typically being *nothing*, PCTs in fact *answered the question that everyone wanted to know*: how does the treatment compare to nothing? Now, the force of Public Citizen's methodological argument in favor of ACETs came from their insistence that it was already known—by inference from historical data—that short courses of AZT were better than nothing. If that claim was correct, then the CDC and the NIH were pursuing the wrong research question, and there were no epistemic grounds for conducting PCTs. One can assess this claim by reviewing the *totality of evidence.*

Lurie and Wolfe [8] charged the WHO, CDC, and NIH with inadequately incorporating the data from the 076 trial and other sources into their reasoning for endorsing PCT design. The data, Lurie and Wolfe argued, established the reasonable

[2] Spencer P. Hey and Charles Weijer [54] similarly called for totality of evidence in properly inferring both claims of efficacy and efficiency in clinical trials. They proposed that "what is needed … is a series of trials, whose various designs are perturbed to optimally contribute toward demonstrating a robust pattern of evidence" [54, p. 3]. This convergence of conclusions is not surprising, as the "totality of evidence argument" follows from Anderson's observation [53] that Temple and Ellenberg's [39] concern that ACETs rely on external information is just an instance of Duhemian under-determination.

conclusion that short course AZT is better than nothing, and so PCTs were not needed. They found it "disturbing" that the rich data available from a well conducted randomized trial was being ignored. I will now review the external information compiled by Public Citizen in order to consider whether legitimate inference could have been made to justify an equivalence study.

The key piece of missed information was a subgroup analysis of the 076 data regarding the duration of prepartum AZT therapy. Public Citizen argued that it was known in early 1994—prior to the completion of 076—that a shorter course of AZT therapy might work as well as longer treatment and that the shorter course was almost certainly more efficacious than a placebo [56]. 076 had been stopped in late 1993 in order for women in the placebo arm to be placed on AZT therapy. Some of those women were quite advanced in pregnancy when they began treatment. In the 076 trial publication, the authors stated that the efficacy of AZT was observed in all subgroups regardless of length of prenatal treatment [56].

The data supporting that claim were not published, but they were presented to the NIH Data Safety Monitoring Board in February 1994. The data suggested that among those women getting 12 or fewer weeks of AZT (7 week average), the shorter courses were significantly better than placebo, with 7.7 % infected infants in the AZT group versus 22.9 % in the placebo cohort (see Fig. 1). This tentatively (given the small sample size and post hoc nature of the analysis) answered the very question being asked by the translational PCTs [56].

Since the publication of the 076 results, further data had accumulated suggesting the equivalency of shorter and longer regimens of AZT, and that shorter regimens were better than nothing in both developed and developing world settings [56]. Public Citizen charged that this data precluded the need for placebo comparators.

Temple and Ellenberg were correct to suggest that inference from historical data must be done carefully. One must consider whether there is any reason to think that the 076 data would *not* provide a legitimate comparator. The inference from the 076 data to the multinational trials in question may be illegitimate due to salient differences between the trials. They included a potential difference in (1) the administration of AZT during labor and delivery. The use of oral instead of

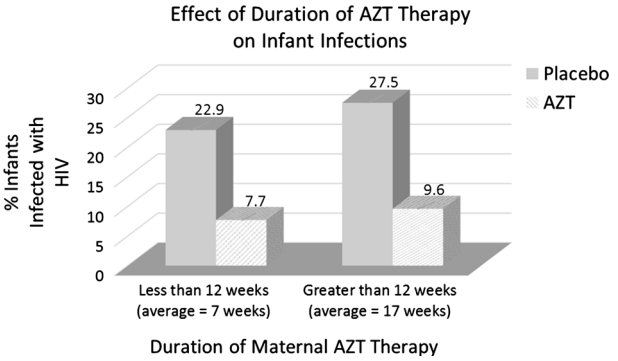

Fig. 1 Effect of duration of AZT therapy on infant infections (recreated with permission [56])

intravenous administration in the short-course regimens could make the 076 comparator irrelevant. But a CDC-sponsored study using both modes of administration found their subjects to have similar AZT levels in their blood [56]. This finding confirmed previous pharmacokinetic modelling data [8].

A second salient difference potentially lay in (2) the differences between subject populations. The 076 trial drew from American and French subject pools that, it was argued, may have different vertical transmission rates and different responses to AZT. Disease prevalence is known to vary among populations for a variety of genetic, environmental, and socioeconomic reasons. People living in extreme poverty in the developing world may respond differently to pharmaceuticals than those in the developed world due to limited or no prior exposure to pharmacologic agents ("drug naivety"), poor nutrition, and anemia. However, there was ample evidence to allay that (legitimate) worry, all of which was available in 1997 when this debate was underway. Presenters at the 1997 Global Strategies for the Prevention of HIV Transmission from Mothers to Infants conference shared trial findings suggesting that HIV transmission rates among women provided the 076 regimen in developing countries rivaled those observed in the treatment arm of the 076 trial. While 076 had transmission rates of 7.6 versus 22 %, women administered similar AZT regimens had transmission rates of 7.9 % in Thailand, 7.7 % in Poland, and 11.9 % in the Bahamas compared to placebo group rates of roughly 22 % [56]. There were no contrary data to support the notion that the 076 regimen would be ineffective when administered in developing countries [56].

A third potentially salient difference involved (3) the breastfeeding practices among trial subjects in the developed versus developing world. The 076 regimen stipulates that HIV-infected women should refrain from breastfeeding. Women in the developing world, some felt, were unlikely to follow that specification, which could make the 076 comparison inappropriate. HIV can be transmitted to the infant through breast milk, however, breastfeeding also protects infants against other infectious diseases. Furthermore, infant formula may not be available and bottle feeding might stigmatize women by revealing their HIV status to their family and community. Public Citizen, however, argued that this commonly cited concern ignored estimates of the probability of HIV transmission from breastfeeding published prior to the completion of 076. Breastfeeding was estimated to represent a 14 % risk of transmission from mother to infant [56], so subjects still stood to benefit from the 076 regimen even if the benefit were to be reduced.

The totality of this external information, it was argued, precluded the need for PCTs. Lurie and Wolfe concluded that "on the basis of the ACTG 076 data, knowledge about the timing of perinatal transmission, and pharmacokinetic data, the researchers should have had every reason to believe that well-designed shorter regimens would be more effective than placebo" [8, p. 854]. This, of course, was exactly what the CDC maintained to be the research question of interest. Lurie and Wolfe could not answer the question definitively,[3] of course, but they provided good

[3] While this patchwork of inferential information made a sufficiently strong case against the need for PCTs, it was not strong enough to preclude the need for further experimentation all together. Before implementing large-scale health interventions that impact the lives of many and utilize considerable resources, one needs more secure data than reasonable sounding inferences. Direct experimentation is

reason to accept this hypothesis, thereby destabilizing the genuine scientific uncertainty (clinical equipoise) that is supposed to ethically justify the use of placebo controls.

One year later, a more definitive answer was provided by the completion of the first multinational trial, a CDC funded PCT in Thailand that confirmed short-course AZT was better than placebo in preventing perinatal HIV transmission [34]. The Public Citizen Health Research Group commissioned a news release stating: "The Centers for Disease Control and Prevention (CDC) today released results of a study in Thailand confirming what the US government knew four years ago: Short courses of the drug AZT are highly effective in reducing transmission of HIV from pregnant women to infants" [57]. The results of the CDC study were similar to the subgroup analysis of the 076 data (Fig. 1). This finding caused Wolfe to lament that "had they only paid attention to the results of their own earlier studies, this disgraceful loss of life would have been avoided" [57].

The placebo orthodoxy and multinational research ethics today

In the 18 years that have passed since the "double standard" debate first transpired, the multinational HIV-research landscape has predictably changed with respect to the chemical agents and interventions under review, but other features, namely, the placebo orthodoxy, remain the same.

In anticipation of the completion of the first multinational PCT, UNAIDS and the NIH had promised that a positive finding in favor of short-course AZT over placebo would prompt the elimination of the placebo arms in the other trials underway [57]. Indeed, one of the multinational studies comparing AZT to a placebo was redesigned as an ACET [58]. The others were further along in data collection and, instead, stopped enrollment at that point so that women in the placebo arms could receive treatment [59, 60]. Yet, other studies investigating non-AZT interventions (vitamin therapy, immunoglobulin therapy, and vaginal washes) retained their placebo arms, and subsequent studies still employed placebos. Perhaps because public scrutiny had subsided, the placebo orthodoxy resumed.

The orthodoxy's stronghold continued even after the October 2000 publication of the one ACET conducted alongside the numerous PCTs in developing counties [61]. Results from this NIH-funded study supported Public Citizen's claim that ACETs could provide reliable information *and* that they could be scientifically advantageous by offering more relevant data than PCTs.

Public Citizen had earlier challenged the rationale of the NIH insisting on the scientific necessity of placebo controls yet funding this one ACET nonetheless [7]. The primary investigator had been strongly pressured by the agency to redesign the study to include a placebo arm, but he resisted and was supported by his university in doing so. The NIH eventually relented.

Footnote 3 continued
generally more trustworthy. The inferential data, however, was compelling enough to help determine what kind of direct experimentation should be undertaken.

Table 1 Results from an active-controlled trial comparing three shortened AZT regimens against a control arm similar to Protocol 076 (data from [61])

Arm	Mother	Infant	Transmission rate	
1	Long[i]	Short	4.7 %	[i]Women in all groups received oral AZT during labor
2	Short[i]	Long	8.6 %	
3	Short[i]	Short	10.5[ii] %	[ii]Terminated at interim analysis
4[iii]	Long[i]	Long	6.5 %	[iii]Similar to 076 regimen

Public Citizen held up the study as an exemplar of an ethical *and* scientifically advantageous study—with no trade-off. The study investigated short-course regimens by comparing three shorter AZT regimens against a regimen somewhat similar to 076. 1,437 HIV-positive pregnant women were randomly assigned to the four arms, which varied in duration of AZT prophylaxis received by both mother (long or short) and infant (long or short) (see Table 1).

The HIV transmission rate in the short-short arm was higher than the others so that arm was terminated upon interim analysis. Even so, the short-short arm was so much better than the placebo arms in all of the other studies, as well as better than transmission rates in previous non-treatment studies conducted before AZT was proven effective for this purpose, that it was not difficult to conclude that short-short was better than nothing. This counters Varmus and Satcher's concern that with any findings ascertained without a placebo comparator, it would "still be unclear whether the affordable intervention is better than nothing and worth the investment of scarce resources" [3]. The results also showed the transmission rates of the long-long (6.5 %), long-short (4.7 %), and short-long (8.6 %) to be statistically indistinguishable. This demonstrated that pragmatic regimens are worth pursuing, and it did so without the use of a placebo arm.

The study also offered something that the PCTs did not: it provided clinicians with much better guidance as to which part of the AZT regimen is critical to efficacy. The rate of *in utero* transmission was found to be significantly higher in the two regimens with shorter maternal treatment (5.1 %) compared with the two groups with longer maternal treatment (1.6 %) [61]. The benefit of long prenatal treatment was established because the long–short regimen was equivalent to the long–long regimen. It also showed that *if* long maternal treatment is provided, only short treatment of infants is needed [61].

But what if this or other ACETs had revealed non-equivalence between the experimental agent and the active control? Indeed, one might reasonably expect to find the short-course regimens to be less effective than the intensive comparator. One might object that this finding would amount to useless data, and therefore, an ACET design would run the risk of wasting precious time and resources. Avoiding such waste could arguably justify employing placebo controlled trials.

However, a finding of non-equivalence is not useless. The comparative look at different pragmatic regimens just described would have provided valuable information regarding where limited resources should be directed *even if* equivalence had not been ascertained. The shorter course AZT regimens were

already known to be better than nothing, but how it compared to the (impractical) standard of care was not yet known. This new knowledge not only informs the normative questions of whether these regimes warrant the investment of scarce resources, but also what features of the treatment regimen (i.e., prenatal vs. neonatal prophylaxis, duration of drug therapy) produce the most significant health benefits.

In bioethics, research attention continues to be paid to the ethics of multinational trials, but the focus is on new HIV research priorities, namely, HIV-preventative vaccines and microbicides. Both vaccine and microbicide trials are first generation research, which makes their use of placebo appropriate. Yet, in a recent paper on the ethics of microbicide research, Macklin foreshadows the placebo question in future trials once a microbicide becomes established [62]. In response to the strong position that placebos are never acceptable for second generation research, she rehearses the placebo orthodoxy that previous analysis should have undermined. She writes: "This guidance point is controversial from a methodological perspective. A clinical trial comparing an experimental preventive method with a proven method takes longer to complete and is more costly, and the results may be much more difficult to interpret than a placebo-controlled trial" [62, pp. 200–201]. These developments in developing world HIV trials and in multinational research ethics scholarship suggest that the placebo orthodoxy is going strong despite the challenges launched by public reaction to those infamous mother-to-child HIV-transmission studies.

In conclusion

A review of the infamous perinatal HIV-transmission trials reveals widespread acceptance of the placebo orthodoxy among researchers, regulators,[4] and bioethicists working on the ethics of multinational research. Methodological analysis suggests the placebo orthodoxy is unfounded, as are, by extension, the ethical arguments supportive of PCTs. I regret that philosophers of science did not involve themselves in the intense public and academic debates over these trials, as they would have had something valuable to add to those important discussions. I support current calls for "socially relevant philosophy of science" [63, 64] in light of this case analysis.

I argued that there was sufficient "external information" or "historical data" available to preclude the need for placebo controls in those trials. Short-course AZT was known to be better than nothing. There were epistemic grounds for legitimate inference from 076 and other data to justify 076 as the comparator in ACETs. Many lives could have been saved by acting on this information. Future discussion about

[4] One might object that the NIH and CDC could have been genuinely motivated by cost considerations rather than placebo orthodoxy in their promotion of PCTs, given that my arguments regarding the limits of statistical significance testing only make the case that PCTs *should not* be cheaper that ACETs if the former are conducted well. Instead, PCTs *are* cheaper when their evidentiary limits are accepted or ignored. It is not known what knowledge the NIH and CDC had regarding those design limits at the time of this controversy, and how that knowledge had impacted institutional practice. I thank an anonymous reviewer for bringing this point to my attention.

the ethics of HIV research in the developing world should learn from these findings and not subscribe to the placebo orthodoxy.

References

1. Lurie, Peter, Sidney M. Wolfe, Wilbert Jordan, George J. Annas, Michael Grodin, and George Silver. 1997. Letter to the Department of HHS concerning their funding of unethical trials which administer placebos to HIV-infected pregnant women through NIH and the Centers for Disease Control. HRG Publication #1415. April 22, 1997. http://www.citizen.org/publications/publicationredirect.cfm?ID=6612. Accessed Sept. 25, 2014.
2. O'Connor, Edward M., Rhoda S. Sperling, Richard Gelber, et al. 1994. Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with Zidovudine treatment. *New England Journal of Medicine* 221(18): 1173–1180.
3. Varmus, Harold, and David Satcher. 1997. Ethical complexities of conducting research in developing countries. *New England Journal of Medicine* 337(14): 1003–1005.
4. Pratt, Bridget, and Bebe Loff. 2013. Linking international research to global health equity: The limited role of bioethics. *Bioethics* 27(4): 208–214.
5. World Medical Association. 1996. *Declaration of Helsinki*. 48th WMA General Assembly, Somerset West, South Africa.
6. World Health Organization. 1994. *Recommendations from the Meeting on mother-to-infant transmission of HIV by use of antiretrovirals*. Geneva: WHO, 1994. http://www.columbia.edu/cu/musher/AIDS_case/geneva_conference.htm. Accessed Nov. 12, 2013.
7. Wolfe, Sidney M., and Peter Lurie. 1997. Letter to President Clinton urging him to stop sponsoring unethical biomedical research on mother-to-infant HIV transmission in developing countries. HRG Publication #1419. June 11, 1997. http://www.citizen.org/Page.aspx?pid=4070. Accessed Oct. 7, 2013.
8. Lurie, Peter, and Sydney M. Wolfe. 1997. Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. *New England Journal of Medicine* 337(12): 853–856.
9. French, Howard W. 1997. AIDS research in Africa: Juggling risks and hopes. *New York Times*, October 9.
10. Cohen, Jon. 1997. Ethics of AZT studies in poorer countries attacked. *Science* 276: 1022.
11. Bagenda, Danstan, Philippa Musoke-Mudido, Susan Okie, and Marcia Angell. 1997. A look at ethics and AIDS—we're trying to help our sickest people, not exploit them: In the researcher's code of conduct, contradictions abound. *Washington Post*, September 28: C3.
12. Stolberg, Sheryl Gay. 1998. Placebo use halted in studies of AIDS: Placebo use is suspended in overseas AIDS trials. *New York Times,* February 19: A1, A16.
13. Angell, Marcia. 1997. The ethics of clinical research in the third world. *New England Journal of Medicine* 227(12): 847–849.
14. Macklin, Ruth. 2001. After Helsinki: Unresolved issues in international research. *Kennedy Institute of Ethics Journal* 11(1): 17–36.
15. Lie, Reidar K., Ezekiel Emanuel, Christine Grady, and David Wendler. 2004. The standard of care debate: the Declaration of Helsinki versus the international consensus opinion. *Journal of Medical Ethics* 30(2): 190–193.
16. Glantz, Leonard H., George J. Annas, Michael A. Grodin, and Wendy K. Mariner. 1998. Research in developing countries: Taking "benefit" seriously. *Hastings Center Report* 28(6): 38–42.
17. Crouch, Robert A., and John D. Arras. 1998. AZT trials and tribulations. *Hastings Center Report* 28(6): 26–34.
18. Arras, John D. 2004. Fair benefits in international medical research. *Hasting Center Report* 34(3): 3.
19. Participants in the 2001 Conference on Ethical Aspects of Research in Developing Countries. 2004. Moral standards for research in developing countries: From "reasonable availability" to "fair benefits." *Hastings Center Report* 34(3): 17-26.
20. Harris, John. 2008. Global norms, informed consent, and hypocrisy in bioethics. In *Global bioethics: Issues of conscience for the twenty-first century*, ed. Ronald M. Green, Aine Donovan, and Steven A. Jauss, 297–322. New York: Oxford University Press.

21. Baylis, Francois. 2008. Global norms in bioethics: Problems and prospects. In *Global bioethics: Issues of conscience for the twenty-first century*, ed. Ronald M. Green, Aine Donovan, and Steven A. Jauss, 323–339. New York: Oxford University Press.
22. Schuklenk, Udo. 1998. Unethical HIV perinatal trials establish bad precedent. *Bioethics* 12(4): 312–319.
23. Macklin, Ruth. 2003. Bioethics, vulnerability and protection. *Bioethics* 17: 473–486.
24. Benatar, Solomon R., and Peter A. Singer. 2000. A new look at international research ethics. *British Medical Journal* 321: 324–326.
25. London, Alex J. 2005. Justice and the human development approach to international research. *Hastings Center Report* 35(1): 24–37.
26. Levine, Carole. 1998. Placebos and HIV: Lessons learned. *Hastings Center Report* 28(6): 43–48.
27. Spławiński, Jacek, and Jerzy Kuźniar. 2004. Clinical trials: Active control vs placebo—what is ethical? *Science and Engineering Ethics* 10(1): 73–79.
28. Levine, Robert J. 2000. Some recent developments in the international guidelines on the ethics of research involving human subjects. *Annals of the New York Academy of Science* 918: 170–178.
29. Miller, Franklin G., and Howard Brody. 2003. A critique of clinical equipoise: Therapeutic misconception in the ethics of clinical trials. *Hastings Center Report* 33(1): 19–28.
30. Macklin, Ruth. 2004. *Double standards in medical research in developing countries*. New York: Cambridge University Press.
31. Annas, George, and Michael Grodin. 1998. Human rights and maternal-fetal HIV transmission prevention trials in Africa. *American Journal of Public Health* 88(4): 560–563.
32. Council for International Organizations of Medical Sciences (CIOMS). 2002. *International ethical guidelines for biomedical research involving human subjects*. Geneva: CIOMS.
33. World Medical Association. 2008. *Declaration of Helsinki*. 59th WMA General Assembly, Seoul, Korea.
34. Shaffer, Nathan, Rutt Chuachoowong, Philip A. Mock, et al. 1999. Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: A randomised controlled trial. *Lancet* 353: 773–780.
35. Rothman, Kenneth J., and Karin B. Michels. 1994. The continued unethical use of placebo controls. *New England Journal of Medicine* 331: 394–398.
36. Rothman, Kenneth J. 1996. Placebo mania. *British Medical Journal* 313: 3–4.
37. Freedman, Benjamin, Charles Weijer, and Kathleen Cranley Glass. 1996. Placebo orthodoxy in clinical research I: Empirical and methodological myths. *Journal of Law, Medicine and Ethics* 24: 243–251.
38. Temple, Robert. 1982. Government viewpoint of clinical trials. *Drug Information Journal* 16: 10–17.
39. Temple, Robert, and Susan S. Ellenberg. 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Annals of Internal Medicine* 133(6): 455–463.
40. Moerman, Daniel E. 1983. General medical effectiveness and human biology: Placebo effects in the treatment of ulcer disease. *Medical Anthropology Quarterly* 14(4): 3–13–16.
41. Moerman, Daniel E. 2000. Cultural variation in the placebo effect. Ulcers, anxiety, and blood pressure. *Medical Anthropology Quarterly* 14(1): 51–72.
42. Freedman, Benjamin. 1990. Placebo-controlled trials and the logic of clinical purpose. *IRB* 12(6): 1–6.
43. Center for Disease Control and Prevention. 1997. CDC studies of AZT to prevent mother-to-child HIV transmission in developing countries. *AIDSinfo.* http://aidsinfo.nih.gov/news/363/cdc-studies-of-azt-to-prevent-mother-to-child-hiv-transmission-in-developing-countries. Accessed Nov. 29, 2013.
44. Weijer, Charles. 1999. Placebo-controlled trials in schizophrenia: Are they ethical? Are they necessary? *Schizophrenia Research* 35: 211–218.
45. Hróbjartsson, A., E. Forfang, M.T. Haahr, B. Als-Nielsen, and S. Brorson. 2007. Blinded trials taken to the test: An analysis of randomized clinical trials that report tests for the success of blinding. *International Journal of Epidemiology* 36(3): 654–663.
46. Shapiro, Stan, Dean Fergusson, and Glass Kathleen Cranley. 2010. Substituting placebo for established, effective therapy: Why not? *Canadian Medical Association Journal* 182(16): 1749–1753.
47. Howick, Jeremy. 2009. Examining the alleged methodological superiority of placebo over 'active' controls. *American Journal of Bioethics-Neuroscience* 9(9): 34–48.
48. Salsburg, David S. 1985. The religion of statistics as practiced in med journals. *The American Statistician* 39(3): 220–223.

49. Gardner, Martin J., and Douglas G. Altman. 1986. Confidence intervals rather than P values: Estimation rather than hypothesis testing. *British Medical Journal* 292: 746–750.
50. Rothman, Kenneth J. 1986. Significance questing. *Annals of Internal Medicine* 105(3): 445–447.
51. Savitz, David A. 1993. Is statistical significance testing useful in interpreting data? *Reproductive Toxicology* 7(2): 95–100.
52. Pandis, Nikolaos. 2013. Confidence intervals rather than *p* values. *American Journal of Orthodontics and Dentofacial Orthopedics* 143(2): 293–294.
53. Anderson, James A. 2006. The ethics and science of placebo-controlled trials: Assay sensitivity and the Duhem-Quine thesis. *Journal of Medicine and Philosophy* 31(1): 65–81.
54. Hey, Spencer Phillips, and Charles Weijer. 2013. Assay sensitivity and the epistemic contexts of clinical trials. *Perspectives in Biology and Medicine* 56(1): 1–17.
55. Hill, Austin Bradford. 1963. Medical ethics and controlled trials. *British Medical Journal* 1: 1043–1049.
56. Wolfe, Sydney M., and Peter Lurie. 1997. Letter concerning unethical placebo-controlled studies on HIV-positive pregnant women in developing countries. HRG Publication #1430. October 23. http://www.citizen.org/Page.aspx?pid=2411. Accessed Jan. 2, 2015.
57. Public Citizen. 1998. Dozens of Thai infants needlessly infected with HIV in unethical U.S. government-funded study. [Press release]. HRG Publication #1433. http://www.citizen.org/pressroom/pressroomredirect.cfm?ID=238. Accessed Jan. 2, 2015.
58. Guay, Laura A., Philippa Musoke, Thomas Fleming, et al. 1999. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet* 354: 795–802.
59. Dabis, Francois, Philippe Msellati, Nicolas Meda, et al. 1999. 6-month efficacy, tolerance, and acceptability of a short regimen of oral zidovudine to reduce vertical transmission of HIV in breastfed children in Côte d'Ivoire and Burkina Faso: A double-blind placebo-controlled multicentre trial. *Lancet* 353: 786–792.
60. Wiktor, Stefan Z., Ehounou Ekpini, John M. Karon, et al. 1999. Short-course oral zidovudine for prevention of mother-to-child transmission of HIV-1 in Abidjan, Côte d'Ivoire: A randomised trial. *Lancet* 353: 781–785.
61. Lallemant, Marc, Gonzague Jourdain, Sophie Le Coeur, et al. 2000. A trial of shortened zidovudine regimens to prevent mother-to-child transmission of human immunodeficiency virus type 1. *New England Journal of Medicine* 343: 982–991.
62. Macklin, Ruth. 2010. Intertwining biomedical research and public health in HIV microbicide research. *Public Health Ethics* 3(3): 199–209.
63. Kourany, Janet. 2010. *Philosophy of science after feminism.* Oxford: Oxford University Press.
64. Tuana, Nancy. 2010. Leading with ethics, aiming for policy: New opportunities for philosophy of science. *Synthese* 177(3): 471–492.