The Hepatitis Experiments at the Willowbrook State School

The hepatitis experiments performed at the Willowbrook State School are routinely cited as one of the most serious breaches of research ethics of the post–World War II period. This determination is principally due to the inclusion of the experiments in Henry K. Beecher's 1966 article "Ethics and Clinical Research" in the New England Journal of Medicine. Beecher's criticism set off a decade of debate about the ethics of clinical research at Willowbrook, with sharply differing opinions from leaders in the field. Beecher extended his critique of the experiments at Willowbrook in his book Research and the Individual in 1970.

Willowbrook was an institution for the mentally retarded operated in Staten Island, New York, from 1947 to 1987. For many, Willowbrook is seen today as a symbol of both the improper institutionalization of the retarded and the successful use of the legal system to force state governments to improve the conditions for retarded citizens under their care. For the research ethics community, Willowbrook has become a potent symbol of unethical research. The experiments are often referred to in the same litany as the Jewish Chronic Disease Hospital case and the Tuskegee syphilis experiments (see Chapters 6 and 8). Indeed, Willowbrook is seen by many as the "pediatric Tuskegee," and the principal scientist involved in the studies, Saul Krugman, is routinely vilified.

The reality of the experiments at Willowbrook is more complicated. What really happened at Willowbrook? What are the real lessons of Willowbrook for contemporary research ethics?

Hepatitis Before Willowbrook

Krugman began his work at Willowbrook in 1954. At the time, the causative agent for hepatitis was thought to be a virus and the disease was characterized by two related clinical patterns. The first pattern was infectious hepatitis, thought to be transmitted by the ingestion of infectious material from feces. Transmission of infectious hepatitis by food workers through inadequate sanitation facilities, or by person-to-person contact without good handwashing, had been documented. The second pattern was serum hepatitis, in which the infection was transmitted through inadequately sterilized needles or blood transfusions.

The diagnosis of hepatitis was made by observation of a clinical pattern of vomiting, anorexia, jaundice, and liver tenderness. Blood enzyme assays to detect liver damage were just being introduced. Reliance on the clinical symptoms alone for diagnosis meant that the infection might go undetected or be misdiagnosed. In the mid-1950s, it was unclear whether these "subclinical" cases of hepatitis could still lead to the spread of the infection.

Previous research by Joseph Stokes at the University of Pennsylvania had demonstrated that injections of gamma globulin, an antibody-rich distillate of human serum, could modulate the clinical course of hepatitis by means of "passive" immunity. Stokes theorized that if hepatitis infection occurred during the period of passive immunity produced by gamma globulin, the clinical disease would be mild and long-lasting immunity to future infection might result. He called this theory "passive-active" immunity.

The Initial Studies at Willowbrook

Krugman came to the Willowbrook State School as a consultant in infectious disease from New York University and Bellevue Hospital. He described his intentions at Willowbrook in the New England Journal of Medicine in February of 1958:
The present report is concerned with an attempt to control the high prevalence of infectious hepatitis in an institution for mentally defective patients. Its purpose is threefold: to describe the circumstances under which the disease occurred, and the effect of gamma globulin in reducing its occurrence; an attempt to induce "passive-active immunity" by feeding virus to persons protected by gamma globulin; and to describe the excretion of virus during the incubation period of the disease.12

The investigations, funded in part by the Armed Forces Epidemiology section of the U.S. Surgeon General's Office, began with an epidemiologic survey of hepatitis at the school. Krugman demonstrated that the majority of hepatitis cases were acquired while at the institution, rather than as the result of infection prior to admission. By surveying the sewer and water systems, the growth and preparation of food, and the clinical histories of those who prepared and served the food, he also demonstrated that the source of hepatitis at the school was contact among infected students rather than infection from the food supply.

The Willowbrook strain of hepatitis was mild compared with other reported cases. Indeed, there were no deaths from hepatitis either in the patient population or in the attendants from 1953 to 1957. Krugman documented the rate of clinically apparent hepatitis among children and attendants at the school. The rate of acquisition of hepatitis among children at the school was to become a source of much contention, but Krugman's estimate at the time was that 40 to 50 patients per 1,000 per year contracted hepatitis.

Krugman and his co-investigators set out to explore the protective effects of gamma globulin on the children at Willowbrook. After an initial trial with what was shown to be an inadequate dose, a second trial compared hepatitis rates between two groups of recently admitted students, only one of which was given gamma globulin injections. The results were startling. The children given gamma globulin appeared to be protected against clinical hepatitis for 39 weeks. The duration of the protection against infection was unexpected, because in the work by Stokes and others the protective effects of gamma globulin had lasted only 6 weeks. In order to explain the difference, Krugman asked whether the prolonged protection against hepatitis in persons injected with gamma globulin might be due to Stokes' passive-active immunity: "If so, it might be induced artificially by feeding virus to patients protected by an injection of gamma globulin."12

This hypothesis is the essential aspect of Krugman's experimental program, namely, that infection of children with a mild form of hepatitis could be an effective strategy to confer long-lasting immunity. In a report in 1957, Krugman wondered,

Would gamma-globulin prevent the spread of hepatitis, and if prevention occurred, would the effect be transitory or would it be prolonged in such a way as to suggest "passive-active" immunity (Stokes)? Could "passive-active" immunity be induced experimentally in small isolated groups by injecting gamma-globulin and then feeding hepatitis virus?13

The idea that infection with a mild form of a viral agent could induce immunity was well established by the time of Krugman's work, and in 1957 Krugman directly refers to his research as "immunization."13 Much of the work on infectious diseases of childhood focused on just this approach. The polio trials14 are perhaps the most famous example, but the work to induce immunity to measles also followed a similar pattern at precisely the same time, the mid-1950s15 (see Chapter 5).

**Ethical Issues Considered Before Beginning the Research**

In outlining their intention to initiate the research, Krugman and colleagues wrote that "[t]he decision to feed hepatitis virus to patients at Willowbrook was not undertaken lightly."12 The depth of planning for the trial and the lengthy list of ethical considerations prior to beginning the research are clearly enumerated in the 1958 *New England Journal of Medicine* article:

It is well recognized that infectious hepatitis is a much milder disease in young children. Hepatitis was especially mild at Willowbrook; it was even benign in adults and there were no deaths. . . . Only the local strain or strains of virus already disseminated at Willowbrook would be used. . . . Since the annual attack rates of jaundice were high, for example 20 to 25 per 1000, and since in all probability cases of hepatitis without jaundice were occurring with the frequency equal to overt forms, it was apparent that most of the patients at Willowbrook were naturally exposed to hepatitis virus. . . . The advantages were considered of inducing the infection under the most favorable circumstances such as special isolation quarters with special medical and nursing personnel to provide close observation and extra care. . . . The study was planned so as to begin with very small and obviously ineffective doses of virus and to increase the dosage level gradually, in accordance with the results obtained. . . . The study group would contain only patients whose parents gave consent. . . . A serious uncontrolled endemic situation existed in the institution, and knowledge obtained from a series of suitable studies could lead to its control. . . . These factors were instrumental in the decision to proceed with the plan for titrating virus and inducing so-called passive active immunity.
The plan was sanctioned by the authorities of the New York State Department of Mental Hygiene, by the Armed Forces Epidemiologic Board of the Office of Surgeon General. From today's perspective, this list of considerations mimics those presented in protocol applications to an institutional review board. Krugman designed an experiment that presented the least risk possible to those enrolled. He began with a low dose to observe side effects, created a specialized system for monitoring the children, and used an agent known to produce a mild form of the disease. He took into account the risks that the children faced in the absence of participating in the research. He considered the benefit to those enrolled as well as to other children facing the same circumstances. He obtained consent from the parents of every child who participated. And he obtained an independent review of the study design from experts in the field.

One result of the research program at Willowbrook was a reduction in the incidence of hepatitis among patients and employees by 80 to 85 percent. Yet a beneficial outcome does not justify unethical research.

**Criticsisms of the Willowbrook Studies**

Criticism of the Willowbrook experiments was first published in the *New England Journal of Medicine* in 1966 by Beecher, who continued his attack in 1970 in his *Research and the Individual*. Beecher set the tone for subsequent condemnations of the experiments, and the legacy of his errors can be seen not only in the literature but also in a brief unsuccessful attempt to outlaw all pediatric research in New York. Beecher and later critics have made seven interlocking charges against the experiment:

1. Research that is done not for the benefit of the children involved in the study, but for others, is unacceptable. One of Beecher's primary concerns in writing the 1966 article was to criticize experimentation on one group of individuals solely to benefit another group. He cites the World Medical Association's draft code on ethics—which was to become known as the Declaration of Helsinki—and concludes, "[t]here is no right to risk injury to one person for the benefit of others."

Beecher's criticism misses the mark at Willowbrook. Krugman had been clear in each report of the Willowbrook research that the goal of the research was to induce immunity in the children participating in the research so as to afford them protection against future infection. Krugman was a problem at Willowbrook. Were Krugman to have performed the experiments on children who were not in an institution, and therefore not at an increased risk of acquiring hepatitis, then a case could be made that the experiment would place the children at risk only to benefit other children or adults. In the modern parlance, there was a "prospect of a direct benefit" to the children participating in the study, although this wording was unavailable to either Beecher or Krugman.

This is, of course, not to say that only the children at Willowbrook would benefit from the experiment; if Krugman were correct, then the induction of "passive-active" immunity might provide a boon to others who lived in crowded conditions with an increased potential for acquiring hepatitis. It is likely that the prospect of effective immunization against hepatitis that might be used with military recruits was the reason for the funding provided for the experiments. But the prospect of benefiting others does not exclude the prospect of benefit to the children at Willowbrook.

2. Deliberate infection of a person with an infectious agent as a part of research is unacceptable. Beecher's argument is that the intentional induction of an infectious disease is an unacceptable practice as part of research, regardless of the reason or the potential benefits of the research. Although he does not elaborate his concern, it appears that he has a principled objection to making someone sick when they are part of an experiment.

Beecher's objection is not very persuasive. There is no ethical weight that should be attached to the use of an infectious agent in a study independent of the effect that the infectious agent has on the study's risk. Beecher's rhetoric of "infection" carries with it undertones of dirt or pestilence when none is reasonably present. Beecher's argument appears to rest on a view of the human body as being irrevocably damaged by contact with infectious agents, and this is simply not the case, as the history of immunization programs amply demonstrates. The ethical issue is the harm done by the infection, not the mere fact of infection itself.

3. The parents who consented were unaware of the risks of participation. Beecher's claim is not that parents did not consent, but that there was inadequate disclosure of the details of the trial to the parents. His argument is that the research was so risky that no reasonably informed parent ought to have consented, and he takes the fact that the parents did consent as evidence that the consent process must have been inadequate.

Not much is known about the specific information provided to parents of children approached to participate in the Willowbrook experiments. In 1967, Joan Giles, Krugman's longtime collaborator in the hepatitis studies, described the consent process in the following way:

I explain that there is no vaccine against infectious hepatitis, that the disease is always present here, and that their child is quite likely to come in contact with it by the intestinal-oral route common to a close quartered group of this type. I also tell them that we can modify the disease with gamma globulin but we can't provide lasting immunity without letting them get the disease. I explain that we use blood serum taken from Willowbrook patients who had hepatitis and that experience has shown a minimum dosage that can induce the disease in a form even less severe than occurs naturally in patients outside the hepatitis unit.

In *Research and the Individual* Beecher responds to Giles' comments by arguing that "it was not clear whether any or all of the parents were told that hepatitis sometimes progresses to fatal liver destruction or that there is a possibility that cirrhosis developing later in life may have had its origin in earlier hepatitis." Beecher's criticism boils down to a concern that there was a failure to focus on the serious but small risk of death due to hepatitis with liver failure. His criticism ignores that this complication had not been seen during the survey of hepatitis carried out at Willowbrook before the studies began: "Hepatitis was especially mild at Willowbrook; it was even benign in adults and there were no deaths." In considering the overall quality of the consent process described by Giles, and acknowledging that she may have been explaining it in the best possible light considering Beecher's criticism, it is hard to argue convincingly that the parental consent was so insufficiently informed as to make the entire process unethical and the consents invalid.
4. Parents were coerced into enrolling their children in the research by the lack of available space at the school. Beecher’s criticism is based on events that were reported in 1967 but that occurred in 1964. Admissions to Willowbrook were halted due to overcrowding, yet space remained for additional children in the separate hepatitis research building. At that time, letters were sent by Dr. Jack Hammond, the medical director of Willowbrook and coauthor on several reports of the hepatitis experiments, to the parents of children who were on the waiting list informing them that there was space in the research building. Beecher’s conclusion was that the investigators could not ethically be allowed to benefit, in the form of new children in their trial, from the lack of space at the school, and that enrollment should have ceased once parents had only the option of enrolling their children in the study or of not placing their children in Willowbrook at all.

The grounds for Beecher calling this letter unacceptably coercive are unclear. Parents clearly did want to admit their children in the school before they heard of the hepatitis experiments, and there is no evidence that the clinical standards for admission to the school were manipulated for those parents willing to enroll their children in the experiments. Parents were offered a set of options, neither of which was by itself unethical. There was no evidence of monetary or other incentives that induced the parents to choose enrollment in the studies. It is not prima facie unacceptable to require consent to research participation as a prerequisite for entry into a specialized care facility. Under such a reading of coercion, one might conclude that all institutions such as the NIH Clinical Center, where patients are admitted by agreeing to participate in a research program, systematically engage in unacceptable coercion. Such a reading abuses the meaning of the term coercion.

5. Infection with hepatitis was not “inevitable” for children admitted to Willowbrook as Krugman had argued. The rate of hepatitis infection among the children at Willowbrook has been the subject of enduring debate. Krugman and others argued that if infection with hepatitis were “inevitable” for children admitted to Willowbrook, then it would be acceptable to infect them under controlled conditions.

It is now clear that Krugman’s rhetoric inflated the risk of infection with hepatitis. He reported in 1958 that the rate of hepatitis with jaundice was 25 per 1,000 per year, and that the rate of infection without jaundice was likely to be twice that, or 50 per 1,000 per year. Yet a recent best estimate using data available to Krugman at the time concludes that between 30 and 53% of the children admitted to Willowbrook would have acquired hepatitis during a childhood spent at the institution. These estimates are below the claim of “inevitability” cited by Krugman and his supporters. Although all children in the experiments would contract hepatitis, only half—using a “generous” estimate—of the children not participating in the trial would contract the disease. There may have been a subgroup of children in whom the risk of infection was greater—perhaps those with a greater degree of disability or those exhibiting specific behaviors—and if so, then there may have been a subset of children for whom infection was “inevitable.” But as these characteristics were not used in selecting children for the trial, the claim that infection was “inevitable” for the children in the general population does not withstand close scrutiny.

How much does this matter to the overall assessment of the experiment? If the goal of the trial were to study the effects of infection per se—or if the goal were, as Beecher suggests, simply to determine the period of infectivity—then the lack of “inevitability” denies the trial, because the risk to the children not enrolled in the trial is less than that to those enrolled. Yet this was not the case, because there was the prospect of direct benefit to the children participating in the experiments.

If we correctly recognize that the experiments were done in an attempt to confer long-lasting immunity, then we can ask at what threshold of risk for an infectious illness in a given population should we begin immunization trials. We can get a sense of the acceptable threshold at the time by comparing Krugman’s work to the other immunization research of his era. Using the 30% figure, the risk of contracting hepatitis as a child at Willowbrook was substantially greater than the risk of contracting polio as a child in the general population. The point is that we ought to use a threshold risk in the population substantially lower than “inevitable” for the comparison of the risks of trial participation. Compared to other trials at the time, a risk of 30% was certainly over the acceptable threshold.

6. The experiments were unacceptable “experiments in nature.” Some have criticized Krugman for participating in a problematic “experiment in nature,” a situation in which something bad is known to be happening to a group of people, and rather than preventing the bad event, a researcher exploits the situation by studying those negatively affected by it. Rather than study hepatitis in children, the argument goes, Krugman had a moral duty to change the institutional conditions that led to the infection.

Calling the research at Willowbrook an “experiment in nature” rests on a mistaken idea that infection of the children was done in a convenient population simply to understand the consequences of infection. As Krugman explained in 1967, “Willowbrook was not chosen because its population is mentally retarded, but because it had endemic infectious hepatitis and a sufficiently open population so that the disease [hepatitis] could never be quieted by exhausting the supply of susceptibles.” Krugman was intervening in an epidemic situation, not simply standing by and observing. More importantly, his goal was to help those afflicted or likely to be afflicted by the illness in the very institution where the study was being done. Krugman’s aim was to remedy the situation he found, not just to use it for an experiment. Again, the criticism that the studies were “experiments in nature” rests on a failure to see them as a program of immunization designed to address the problem of hepatitis in the institution.

7. The researchers should have cleaned up the conditions that led to the increased risk of infection rather than studied how to protect the children via immunization. At Willowbrook, the increased hepatitis risk faced by the children was a consequence of the decision to gather children with mental disabilities and incontinence together in an institution, rather than a consequence of the children’s disabilities per se. It can thus be argued that the conditions leading to the increased risk of hepatitis at Willowbrook were artificially created, because they were a result of a policy of institutionalization, and that by halting the institutionalization of children, the risk of hepatitis would be greatly reduced without the children having to undergo the risk of participation in research. If so, did the investigators have a moral duty to change the policy and to thereby decrease the risk of hepatitis faced by the children?

In order to answer this question, we must first know whether there were steps short of closing the institution (and not involving...
imunization) that might have prevented the risk of hepatitis infection. Preventing the fecal-oral spread of infectious agents among incontinent children in an institution is not a simple matter, even in a resource-rich environment. Control of hepatitis A outbreaks in neonatal intensive care units remains difficult even today. Effective cohorting of children to prevent cross infection takes strict measures, with quarantining of all infectious children. Prior to the work of Krugman and his colleagues, such cohorting within the institution would have proven ineffective, because identification of those who were infectious was not possible. Nor would it have been clear what the duration of quarantine should be. In the context of a long-term residential program, physical measures to prevent infection would likely have meant the end of interactions among the children, with the indelible closing of play groups and other measures thought to be therapeutic. Faced with those options, an attempt to discover an effective means of conferring immunity seems an appropriate means to address the medical risk to the children while preserving their ability to participate in the life of the institution.

So, were the investigators ethically bound to close the institution, or was it ethically viable instead to study how to make the institution safer? At the time of the hepatitis experiments, parents and physicians were eager to get children admitted to Willowbrook because institutionalization was thought to be the best thing for the children and for their families. Placement in Willowbrook—that is, placement in a specialized school where retarded children could have access to the services of experts—was at the time seen by many as a symbol of an enlightened approach to the plight of retarded children. Objecting to the institutionalization of children at Willowbrook in the 1950s and early 1960s, based on our contemporary approach to mental retardation in children, is open to a charge of anachronism, as well as of a certain arrogance that we are more ethically evolved than those who preceded us. Given the view that institutionalization was a beneficial policy for children and their families, Krugman and colleagues did what they could to improve the chances that institutions were safer for their child residents. Accusing Krugman of ignoring the suffering of the children at Willowbrook only to further his own agenda makes no sense in this context.

Correcting the Distorted Legacy

Because of the mistaken views of Beecher and others about the scientific objectives of the hepatitis research, Krugman's studies at Willowbrook are persistently cited as an example of unethical pediatric research. Yet many in the medical community who correctly understood the scientific and social context of the research have honored Krugman's work at Willowbrook, as have many of the families of the children in the research. The mistakes of Beecher's analysis should be held to account for much of the continued misunderstanding. The errors are not simply of historical interest, because Willowbrook continues to be invoked in order to cast doubt on the ethics of researching the medical and social problems of retarded or otherwise socially vulnerable children. The use of Willowbrook in such a manner dangerously discourages research as a means to ameliorate health conditions for vulnerable populations of children.

Participation in medical research can be a powerful vehicle by which we devote social resources toward understanding the medical problems of specific populations, as the parallel example of women in clinical research makes clear. Excluded from participating in research, in part by misplaced ethical concerns over the effect of research on a possible pregnancy, women were assumed to benefit from the products of research if men were shown to have benefited from this research. The result was twofold: The unique medical issues of women were ignored, and different physiological responses of women to standard care were rendered invisible. It is a similar mistake to continue to allow the experiments at Willowbrook to cast a restrictive ethical pall over the participation of vulnerable children in medical research.

References