The Rise and Fall of the Dolorimeter: Pain, Analgesics, and the Management of Subjectivity in Mid-twentieth-Century United States

NOÉMI TOUSIGNANT
Anthropologies of African Biosciences Research Group, London School of Hygiene and Tropical Medicine, 15–17 Tavistock Place, London WC1E 9SH, UK. Email: noemi.tousignant@lshtm.ac.uk

ABSTRACT. This article describes how two experimental technologies, the Hardy–Wolff–Goodell dolorimeter and the clinical trial, were involved in, and transformed by, American analgesic research. Introduced in 1940, the dolorimeter quickly became popular as an analgesic-testing technology. By the early 1950s, however, the main sources of funding for analgesic evaluation had shifted to Henry K. Beecher’s clinical trial methodology. To explain both the initial popularity of the dolorimeter and its displacement by the clinical trial, I examine the demands and resources generated by those who participated—as sponsors, investigators, collaborators, or subjects—in analgesic research and evaluation. These actors linked methodological designs to material resources, social interactions, and epistemological values, changing how pain-relieving efficacy both should and could be evaluated. They also mediated the interaction between specific expectations of, and investments in, analgesic evaluation and broader ideas about the reliability of drug evaluation and the subjectivity of pain. My analysis thus connects the changing social and material configuration of analgesic evaluation to the rise of clinical trials as well as increasingly psychological understandings of pain in order to frame the rise and fall of the dolorimeter. Keywords: pain, analgesic, clinical trials, subjectivity, objectivity, quantification.
In 1940, James Hardy, Helen Goodell, and Harold G. Wolff announced the creation of a method of unprecedented accuracy for quantifying pain thresholds. Their technique made use of an apparatus they called the dolorimeter, which was a type of algometer. These misleadingly named devices do not, strictly speaking, measure pain. Algometers are instead designed first to stimulate subjects in precise and potentially painful ways, and then to quantify the minimal intensity of stimulation that results in a response to pain. Originating in late-nineteenth-century psychophysical experimentation, this technology was initially used by trained observers to elucidate general laws of sensation. Soon, however, it became popular for measuring variations in sensory and pain thresholds among different types of untrained subjects. Thus, in early twentieth-century anthropological expeditions and criminal anthropology laboratories, algometers served to define the distinct sensitivity to pain of women, so-called deviants and non-Europeans. In contrast, Newsweek acclaimed the dolorimeter for revealing to the scientific community that we are “all brothers under the skin,” equal in our capacity to feel pain.

The Hardy–Wolff–Goodell dolorimeter was thus set apart from previous algometers by the astonishing stability of its results. Other algometrists had inflicted pressure or electric currents on their subjects. Hardy, Wolff, and Goodell instead focused the heat of a beam of light onto their foreheads. The pain it generated was easy to identify; it could be judged by anyone with equal certainty and detachment. This sensation of pain, they explained, was stripped of

4. Anon., “Brothers under the Skin: We All Feel Pain in Same Way, National Academy Is Told,” Newsweek (6 May 1940) in Harold G. Wolff Papers (1922–70), Box 1, Folder 6, Medical Center Archives of New York-Presbyterian/Weill Cornell, New York (hereafter, Wolff Papers).
its affective content and was impermeable to variations in mood, experience, or personality. Older algometers, they claimed, had measured the more labile and emotionally charged reaction to pain; only their dolorimeter had succeeded in isolating pure sensation, which was perceived by all normal subjects at the same level of precisely quantifiable heat intensity.5

The uniformity and precision of the dolorimetric threshold was especially welcomed by analgesic researchers. Calling it “elegant”6 and “ingenious,”7 they hailed the dolorimeter as a breakthrough in analgesic testing. Previous attempts to use threshold-measuring devices to compare pain-relieving effects had failed due to variability between subjects and trials.8 Psychosomatic researchers also found the dolorimeter useful for producing standardized pain stimuli and quantifying various types of response to pain.9 By 1950, over twenty research teams had published data generated by a dolorimeter; it had entered laboratories across the United States, and traveled to Britain and Canada.10 Students at Johns Hopkins and Cornell medical schools experimented with the method in their physiology and neurology classes.11 It could be purchased as a two-piece apparatus from the Experimental Engineering Corporation, and was also

commercialized by the Co-Design Corporation. This was a sign, wrote a Sterling-Winthrop researcher in 1948, of the stabilization of the technique; it also expressed the prediction of a growing market.

By 1953, however, the president of the Co-Design Corporation, Donald Williamson, complained of low sales of their recently redesigned dolorimeter: “Frankly, the volume on this instrument has been disappointing.” Why were scientists not buying the dolorimeter? Williamson wrote to a select group of researchers interested in pain measurement to find out: Was there a flaw in its design, or something wrong with the company’s marketing campaign? The problem with the dolorimeter apparently extended beyond its commercialization. From 1950, the number of articles reporting the evaluation of analgesic drugs with the dolorimeter—its most popular usage—leveled off. In 1953, the validity of dolorimetric analgesic tests was under attack in the pages of *Science*. The author of this critique was Henry K. Beecher, an anesthesiologist who had recently designed a clinical trial method adapted for analgesic evaluation. In the early 1950s, the main sponsors of analgesic testing in the United States had shifted their support from the dolorimeter to Beecher’s analgesic clinical trial.

Why did the dolorimeter’s popularity decline so suddenly and apparently so unexpectedly in the early 1950s? Perhaps Beecher’s new method of analgesic evaluation was evidently superior to the


14. Donald E. Williamson to Janet Travell, 16 June 1953, Box 19, Folder 30, Janet Travell Papers, George Washington University Archives, Washington, DC.

15. This trend is suggested in the number of publications reporting results using the dolorimeter in my own literature review and confirmed by the references enumerated in the bibliography provided by Kutscher and Kutscher, “Evaluation,” 228–30. It is of course possible that the dolorimeter was still in use after 1950, but that results were rarely published. For example, in 1953, one group reported that they still used the dolorimeter, but only as a test preliminary to clinical studies. See S. C. Cullen and E. G. Gross, “Analgesic Testing Methods,” *Bulletin of the Committee on Drug Addiction and Narcotics*, 1953, 626. Nevertheless, it seems unlikely that the method was widely used.

dolorimeter; maybe he was simply better at selling his method than Hardy, Wolff, and Goodell. The timing of this shift suggests that the persuasiveness of Beecher’s critique of the dolorimeter may have been amplified by two contemporary trends. The adoption of Beecher’s methodology followed a movement of therapeutic reform, in which American clinicians and pharmacologists promoted new criteria for defining the reliability of therapeutic evaluation. From the 1940s, this reliability was increasingly described as being grounded in clinical settings, large numbers of subjects, specific methodological devices (randomization, placebo controls and double-blinding), and statistical analysis—features that corresponded to Beecher’s method. 

Beecher also contrasted Hardy, Wolff, and Goodell’s divided model of pain, and their emphasis on the measurement of sensation rather than reaction, with his own understanding of pain as a whole, irreducibly subjective experience. The clinical trial, he argued, allowed for the evaluation of pain relief using a complex experience of suffering in its natural habitat—the sick patient. The victory of Beecher’s method might be seen as the result of a more general shift, taking place in the mid-twentieth century, away from the definition of pain as a psychophysical event generated by the peripheral nerves toward views of pain as a global experience that was modulated by the thinking, feeling self.

Yet, the broader emergence of the clinical trial or increasingly psychological models of pain do not explain how the clinical trial became an effective method for evaluating analgesics, one that was capable of displacing the dolorimeter, or why this happened precisely around 1950. Here I argue that the making of the analgesic clinical trial—as well as the initial popularity of the dolorimeter—must first be located within the history of analgesic research.

Neither method was adopted readymade for analgesic testing; both became entangled with, and were transformed by, the pursuit of better analgesic drugs. Who participated in this quest, what did

19. Marcia L. Meldrum, “‘Departures from the Design’: The Randomized Clinical Trial in Historical Context, 1946–1970” (PhD diss., Department of History, State University of New York at Stony Brook, 1994), on how the clinical trial was adapted by Beecher for the evaluation of analgesic drugs.
they invest in it, and how did they define the qualities of analgesics? These are key questions for understanding the shifting conditions under which the dolorimeter and clinical trial were valued and validated as analgesic-testing technologies. New actors (sponsors, investigators, collaborators, and subjects) brought new ideas, interests, resources, and capacities to analgesic testing. By their choice, support, and use of specific technologies, these actors linked methodological designs to material resources, social interactions and epistemological values, changing how pain-relieving efficacy both should and could be evaluated. Anchored in a close examination of the practices of analgesic evaluation, this article explores the interaction—mediated by these actors—between interests in analgesics, conceptions of pain, criteria of reliability, experimental conditions, and the validity of specific technologies—the dolorimeter and the analgesic clinical trial.

To introduce both technologies, I first revisit the dispute between Beecher and the inventors of the dolorimeter. I describe how each method worked and, equally importantly, what it took to make it work. In interpreting this debate, I shift the focus from models of pain to models of control. Beecher, Hardy, Wolff, and Goodell argued about the nature of pain; yet their disagreement was, more fundamentally, about how to manage the vagueness and variability of experiences of pain for the purpose of analgesic evaluation. Each method was based on different ideas about what could and should be controlled in order to generate quantifiable pain relief; each achieved precision and consistency through contrasting targets and strategies of control. Implementing these models required different kinds of resources. Paying attention to pain-measurers’ own arguments about the advantages and drawbacks of each method, I describe how their successful implementation relied on a specific collection of expertise, authority, materials, people, practices, and other financial, spatial, and organizational resources.

In other words, each of the pain-testing methods worked under a different set of conditions and expectations. The key questions, then, are: When, how, and by whom were these conditions and expectations generated? The second section of this article follows the rise of the dolorimeter in the context of growing demands for analgesic testing in the late 1930s and 1940s. The dolorimeter was
popular, I suggest, because it was seen as providing a valuable—and affordable—kind of precision by sponsors and practitioners of analgesic testing in the 1940s. They were invested in making the dolorimeter work; but, significantly, they also did not invest the kinds of human and material resources required to produce precise clinical comparisons of pain-relieving potency. It was only in the late 1940s that new investments—monetary, but also professional and institutional—were turned toward clinical analgesic evaluation. By tracing various sources of support for Beecher’s work, I show how a broader movement to promote clinical trial methodologies converged with specific interests in, and arguments for, the clinical evaluation of analgesics. This support, and the types of resources it mobilized, created the conditions, I argue in the third section of this article, under which it became possible to effectively implement and promote the analgesic clinical trial.

In the conclusion, I come back to the relationship between ideas about pain—how its experience is generated and why it varies—and practices for managing its subjectivity. Beecher’s critique of the dolorimeter drew on changing ideas about how pain was modulated by environments, nerves, and individual ways of thinking and being. This broader shift enabled him to represent Hardy, Wolff, and Goodell’s divided model of pain—in which sensory perception could be separated from the broader experience of suffering, and still be called “pain”—as outdated. Yet it can also be argued that these two technologies did not simply follow, but also actively produced—by making operational for analgesic evaluation—new ways of thinking about the subjectivity of pain.

**INDIVIDUAL AND GROUP PAIN: TWO SOURCES OF PRECISION AND MODELS OF CONTROL**

In 1940, in the *Journal of Clinical Investigation*, Hardy, Wolff, and Goodell described their experimental set-up: a 1000 watt bulb projected a strong beam of light, which was focused through a fixed aperture onto a small area of a person’s forehead that had been blackened with China ink (to control for differences in skin pigmentation). The light, of variable intensities, was segmented into exposures of exactly three seconds by an automatic shutter. The subject’s task was simple: to identify the exposure that produced in him/her a sensation described as “heat finally ‘swelling’ to a
distinct, sharp stab of pain at the end.” Remarkably, this sensation was “easily recognizable, even by untrained subjects” and was consistently identified at the same level of light intensity. The lowest level of light intensity at which this sensation appeared was the subject’s pain perception threshold; it was quantified in hundreds of millicalories per square centimeter per second. The experimenter’s job was to keep the machine calibrated, to vary the intensity of each exposure by means of a rheostat and, when the subject said “pain,” to read the value indicated by the dial of a radiometer.20

In contrast, the anesthesiologist Henry K. Beecher wrote of his analgesic clinical trial in 1953: “We are concerned incidentally, of course, with simplicity. A method that can function with no apparatus other than a notebook and a pencil is manifestly more desirable and more broadly useful, other things being equal, than one that requires complex and delicate apparatus which needs calibration by a well-trained physicist.”21 Thus, while dolorimetric pain thresholds—in normal subjects compared with the same subjects who had consumed potentially analgesic substances—were measured as highly precise levels of intensity of radiation, Beecher’s subjects—postoperative patients—were asked to estimate the relief of their pain as “none, slight, moderate or complete”; later they would be asked only to choose between more than 50 percent relieved or not.22 And while Hardy, Wolff, and Goodell generated a purified form of pain sensation, Beecher worked with the messy experience of postoperative pain in real hospital patients.

Yet, we should not be misled by the appearance of simplicity that Beecher contrasted with the intricacies of the dolorimetric method. The analgesic clinical trial, no less than the dolorimeter, intervened to manage the vagueness and variability of human pain in order to create a measurable form of pain relief. The complexity, control, and precision of the analgesic clinical trial lay elsewhere and not in precision instruments, standardized stimulation, and consistent judgments; it was located in procedures of data collection, compilation,

and analysis, and required coordination between numerous collaborators; observers, consultants, and subjects.

The dolorimeter managed variability in perceptions and expressions of pain by intervening within the experience of individual subjects. "Tom, Dick and Harry, all ages, all races, if they are alert and attentive, exhibit the same pain threshold," explained Harold Wolff at a session of the Cornell Conferences on Therapy on the topic of "the Psychologic Aspects of the Treatment of Pain." This uniformity was revealed as a result of the dolorimeter's capacity to isolate the sensation of pain from the reaction to it; from subjects' fears and expectations, moods, memories, and personalities. The dolorimeter enabled subjects to maintain a neutral attitude toward the beam that was heating their foreheads. It produced a pain that was "free of suffering," well-defined, temporary, and unthreatening. By ensuring the accurate quantification and calibration of the stimulus, and by maximizing the distinctiveness of the sensation to be identified, the dolorimeter provided each subject with optimal conditions for making an exact and consistently reproducible judgment. In theory, the quality of the sensation itself ensured this standardization. In practice, users of the dolorimeter—including Hardy, Wolff, and Goodell—found that subjects had to acquire the capacity to discriminate the endpoint consistently and with emotional detachment through a period of "familiarization" or training. The dolorimeter itself standardized the stimulus, while its method of use was meant to standardize the experience of pain, and even the experiencing subject, at the level of the individual.

Beecher’s clinical trial method, on the other hand, exercised little control at the level of individual subjects or their feelings of pain. The source of pain was a surgical wound. Similar wounds, Beecher insisted, did not necessarily give similar pains. Protocols for the trials did not specify any means of standardizing the subjects themselves, beyond selecting postoperative patients for intelligence and cooperativeness. As with the dolorimeter, however, further efforts were made to make patient samples more uniform for the clinical trial. Beecher reportedly favored male subjects because “the menstrual cycle requires troublesome controls.”

Two other techniques of standardization were described explicitly in research reports. For some years, Beecher maintained that what he called “placebo-reactors” should be screened out because they “diluted” the data. In addition, Beecher’s novel technique of using each subject as their own control—by administering different drugs sequentially to a single patient—allowed some variability to be eliminated. If patients were matched against themselves, the idiosyncrasy in their judgments about pain intensity could be cancelled out. However, both these techniques—eliminating placebo responders and using subjects as their own controls—were measures of economy rather than devices essential to the functioning of the method. They were useful but—if sufficiently large series of subjects were employed—not essential.

Abundance, rather than a capacity to make certain types of judgments, was the main virtue of Beecher’s experimental subjects. “The concept of group effect will perhaps always be necessary in dealing with general problems of pain and other subjective

28. Louis Lasagna, Frederic Mosteller, John M. von Felsinger, and Henry K. Beecher, “A Study of the Placebo Response,” Am. J. Med., 1954, 16, 770–79, determining a technique to screen out placebo responders from analgesic clinical trials was cited as one of the rationales for this study. However, later results showed that it was very difficult, if not impossible, to identify those who would respond to placebos consistently.
ailments,” wrote Beecher in 1953. While individual testimonies of pain and relief could be expressed in a numerical form, they were considered to be qualitative judgments. The calculation of the efficacy of a test drug was made on the basis of how many patients reported relief. Large numbers also flattened out variability: “When the series is large . . . one can control suggestion, inherent or implied, the presence of the investigator, practice effect, learning, motivation, interest, the subject’s anticipation of an unknown medication, his drug history,” wrote Beecher, adding that “sufficient numbers” would also “cancel out normal mood swings, above and below par.” Collective measurements of pain required large quantities of data, but also control over the process of data collection as well as statistical expertise to render data intelligible. Observers had to be consistent and neutral; they used standardized questionnaires to interrogate subjects and were blinded, that is, were kept ignorant of the specific drug their subjects had been given. Statisticians were hired to help design and validate the trials.

The dolorimeter and clinical trial thus obtained precision and consistency through different scales and targets of control. What did it take to make each method work?

Calibrating painful stimulation and its perceiver required a precise and mechanical source of stimulation; this equipment was relatively cheap; in 1940, it could be built for $300, and later bought for $850. However, the apparatus needed to be constructed and used identically by each team in order to obtain comparable results. We have seen that researchers found that the dolorimeter worked best when a small number of reliable subjects were used, and were trained over a period of time. These subjects needed to be attentive and available for repeated trials. The instrument could be used in a fairly small, quiet space—usually some sort of laboratory. It was usable by anyone who had access to the right type of subjects and necessary hardware, and the technical expertise to set up and calibrate the apparatus.

The need for trained subjects was underlined by Beecher as a serious flaw in the dolorimetric method. For Beecher, the difficulty of obtaining consistent threshold values in untrained subjects revealed the “leakage” of reaction to painful stimulation into judgments of its intensity. In other words, the dolorimeter failed to fully isolate the sensory experience of pain from the reaction to it. Beecher hypothesized, and later asserted, that pain was an indivisible experience; the theoretical separation of sensation and reaction could not be produced in human subjects. Trained subjects were also difficult to keep ignorant and neutral, Beecher pointed out; they learned how to recognize analgesics and became involved in the experiment. While dolorimetric users saw training as a means of controlling the quality of subjects’ judgments—its neutrality, precision, and consistency—Beecher instead cast trained subject as incompatible with properly controlled drug evaluation.

Beecher and the Cornell team debated this issue in the pages of *Science* in 1953. Yet, throughout the 1940s, analgesic researchers had accepted the use of trained subjects as an unexceptional practice for reducing variability in analgesic evaluation. New attitudes toward the training of subjects are expressive of a more general shift in views of legitimate and necessary ways of controlling therapeutic evaluation. Yet it is also important to note that the issue of subject training was not just epistemological; it was also economic. With training, it was possible to compare analgesics using fewer subjects and trials. The fact that this issue was not controversial in the 1940s can be explained by demands for highly precise, but also relatively cheap and quickly obtained, comparative data on analgesic efficacy.

The reliable evaluation of pain relief in clinical settings, using large numbers of naive subjects, required a very different set of practices, people, and resources. For all Beecher’s talk of simplicity, the analgesic clinical trial was more expensive, time-consuming, and difficult to coordinate than the dolorimetric method. Using simple interrogation—“fairly primitive questions” as one of Beecher’s colleagues would later say—to obtain comparisons of pain-relieving efficacy required a large team of observers, subjects, and consultants to collect and manipulate information under appropriate conditions.

and on a sufficiently large scale. Observers and statistical experts needed to be paid salaries and consultant fees; while recruiting subjects required the authority needed to access a clinical setting, coordinate clinical staff, and oversee patient treatment.

The validities of the dolorimeter and the clinical trial were explained using different models of pain and its subjectivity. Yet they can also be seen as proposing different models for controlling pain’s subjectivity that did not work under the same conditions. To understand why and how each model of control was attractive and workable, both technologies must be examined in interaction with the changing conditions under which they were implemented. In the following two sections, I examine how the dolorimeter and analgesic clinical trial were oriented toward, and transformed by, the pursuit of better analgesics—the definition of better shifting from less addictive and dangerous, to more effective and lucrative, as well as easier to supply and to control—from the late 1920s to the early 1950s.

THE RISE OF THE DOLORIMETER AND AMERICAN ANALGESIC RESEARCH

The dolorimeter was designed in the Cornell University laboratory of neurologist Harold G. Wolff. An obituary described Wolff’s conceptualization of pain as “an aspect of man’s relationship to his environment,” but also, in contrast, as “a discreet sensation.” His interest in pain stemmed from his work on headaches. In addition to conducting detailed research on intracranial blood vessels, Wolff took seriously the role of personality features and emotions in predisposing certain kinds of people to chronic headaches. He investigated this, for example, by measuring the impact of “stress interviews” on arterial pulsations in subjects prone to headache. Yet his research also motivated a search for a form of experimental pain that could be reliably measured. Laboratory notes from the late

34. Oral History Interview with Louis Lasagna, 8 September 1995 (Ms. Coll. no. 127.19), 6, John C. Liebeskind History of Pain Collection, History & Special Collections Division, Louise M. Darling Biomedical Library, University of California, Los Angeles.
35. “Obituary,” Box 1, Folder 1, Wolff Papers.
1930s describe experiments in which subjects—both patients and fellow researchers—were asked to describe the severity of headaches produced by histamine injections. Fellow researchers provided detailed accounts of their pain, but the information given by patients was much less precise; concerns were raised about the reliability of their judgment. The desirability of a measuring technology such as the dolorimeter appears to have been grounded in the perceived need, in the context of a broader research program exploring the psychosomatic determinants of painful conditions, for quantitative and objective—in the sense of bypassing subjects’ emotional and vague expressions of pain—information about sensory pain intensity.

Yet, Wolff quickly emphasized the value of the dolorimeter’s precise and consistent pain thresholds for the purpose of analgesic evaluation. In 1940, a first article describing the dolorimetric method in the *Journal of Clinical Investigation* was followed by a second describing a comparison of analgesic efficacy using the dolorimeter. Earlier, at a meeting in 1939, Wolff had already demonstrated the dolorimeter to William Charles White, chairman of the Committee on Drug Addiction (CDA), which had, since the late 1920s, launched a project to develop a nonaddictive analgesic. This suggested use connected the dolorimeter’s technical potential to the needs of this well-financed and politically backed program of research. As Caroline Acker has shown, the pursuit of a nonaddictive analgesic appealed to various American interest groups—including elite pharmacologists, the American Medical Association, the Public Health Service, the Drug Enforcement Agency, the Rockefeller Foundation, and the National Academies of Science—who supplied the Committee with funds, expertise, facilities, institutional networks, and scientific authority.

40. Caroline J. Acker, *Creating the American Junkie: Addiction Research in the Classic Era of Narcotic Control* (Baltimore: Johns Hopkins University Press, 2002), 65. Until 1940, funding for the Committee was provided in large part by the Rockefeller Foundation.
By the 1930s, this project had created a demand for improved human tests of analgesic efficacy. The later popularity of the dolorimeter can be understood in part through the aims of the CDA’s research program and its demands for analgesic-testing technologies. A nonaddictive analgesic meant splitting the desirable therapeutic properties of opiates from their unwanted side effects. Designing a drug that targeted pain (and cough) selectively depended on the chemical dissociation of these properties; it also required pharmacological means for measuring their respective magnitude in living organisms. Given the high volume and untested toxicity of substances produced during the early phase of the program, analgesia tests were initially run on animals. The Committee’s pharmacologist, Nathan B. Eddy, had designed a successful method in which cats’ tails were pressed until they squeaked.

In the midst of high hopes for the CDA’s first breakthrough drug, desomorphine, arguments broke out over the validity of animal, and then of various human, tests for determining the drug’s dosage and analgesic effectiveness. The reliability of evaluations of analgesics in ex-drug users, who were made available to the CDA through its links with the U.S. Public Health Service (PHS), was soon contested when they indicated desomorphine to be more addictive than anticipated. The Committee then secured access, with the help of the Surgeon General, to the chronically ill patients of Pondville Hospital; yet desomorphine was again shown to be addictive. This seems to have motivated a shift in the Committee’s objectives from a non-addictive to a less addictive drug. Eddy encouraged this by suggesting that a high ratio of analgesic potency to addictiveness might still make desomorphine a

---

41. Nathan B. Eddy, The National Research Council Involvement in the Opiate Problem, 1928–1971 (Washington, DC: National Academies of Science, 1973), 19. In the first two-and-a-half years of the program, sixty-six substances were sent by the chemistry laboratory, based at the University of Virginia for pharmacological testing at the University of Michigan.
44. H. S. Cumming (Surgeon General) to Henry D. Chadwick (Commissioner of Public Health), 23 June 1934, Projects: Development of Nonaddictive Analgesics: Clinical Studies: Pondville Hospital, CDA-NASA.
valuable drug.\textsuperscript{45} This new goal demanded a finer calculus of analgesic efficacy in humans.

From 1935, finding a better human test of analgesia became a priority for the Committee. Two methodological avenues were investigated: autonomic measures of pain correlates, such as pulse or blood pressure, and pain–threshold measurements.\textsuperscript{46} Neither worked out, but their exploration provides a sense of what CDA members were looking for in a test of analgesia: a measure of pain that was stable, impersonal, preferably bypassing the language, will and consciousness of subjects, and provided precise information about pain-relieving potency to weigh against measures of addictiveness.\textsuperscript{47} Clearly, the Committee’s main priority at this point was not to obtain a general sense of how well analgesic drugs worked therapeutically. The reason for the failure of these methods (results fluctuated too much between subjects) shows that the main obstacle to this ideal objectivity was identified as the influence of personal factors on the experience and expression of pain.\textsuperscript{48}

CDA members had little hope that such objectivity could be found in clinical testing. When they did launch trials of Metopon, the next promising drug developed by the CDA, in a network of clinical sites from the mid-1930s, they frequently pointed to gaps in the objectivity of these studies.\textsuperscript{49} Strict study protocols were

\begin{itemize}
\item[45.] N. B. Eddy to W. C. White, 24 April 1935, Projects: Development of Nonaddictive Analgesics: Clinical Studies: Pondville Hospital, CDA-NASA. In this letter, Eddy points out that the drug might still be valuable because of its intensity, but that, in order to determine this, information about the minimal effective clinical dose was needed, on the basis of which it would be possible to conduct a trial “under the most rigid and close control of our organization.”
\item[48.] The Director of Laboratories at Lexington Hospital reported to the Committee that “judgemental and interpretive factors were of more significance in producing pain than any particular degree of stimulation” and to abandon intensive work along these lines: E. Williams, “A Quantitative Measure of Analgesia (Summary of Work Done at Lexington in Past Year, November 12, 1938” in L. Kolb to W. C. White, 12 November 1938, Projects: Development of Nonaddictive Analgesics: Clinical Studies: PHS Hospital: Lexington General, CDA-NASA.
\item[49.] The studies were carried out at the University of Michigan Hospital, Pondville Cancer Hospital, Walter Reed Army Hospital, Marine Hospital, and eventually the
designed to minimize some of the variability in nurses’ and patients’ expectations, intelligence, pathologies, and judgment through the devices of double-blinding (nurses and patients were to be kept ignorant of which drugs were administered), standardized data forms, and criteria for selecting reliable and comparable subject groups. Yet, investigators complained that data came from impressions of pain relief, “not, unfortunately, upon any quantitative measure of analgesic effect”; this was seen as requiring the use of a mechanical device. Indeed, plans for the first study at Pondville Hospital stated the desirability of including a threshold-measuring technology in the clinical protocol.

Clinical protocols, unsupported by mechanical pain assessment, failed to meet CDA members’ expectations of objectivity in analgesic testing. But the protocols also failed to provide the level of precision they sought as a result of the conditions under which they were implemented. CDA members and investigators frequently complained about their lack of control over the clinical studies. The shortage of eligible patients made it difficult to form control groups on the basis of matching severity of pain; it was difficult to obtain the cooperation of attending staff; nurses failed to fill out the standardized forms; and certain patients were deemed to be unreliable. Investigators had little clinical authority—necessary for patient recruitment, control over their treatment, and coordination of staff—and inadequate means of standardizing the production of data between different sites.

Massachusetts General Hospital. Access to these sites was enabled by the Committee’s alliances.


51. “Report on Work on Analgesia at the University of Michigan, October 6, 1937,” Projects: Development of Nonaddictive Analgesics: Clinical Studies: University of Michigan Hospital, CDA-NASA; “Memorandum Number 2, June 18, 1936.”

52. L. E. Lee to W. C. White, 8 December 1938, Projects: Development of Nonaddictive Analgesics: Clinical Studies: Pondville Hospital, CDA-NASA; “Admissions to Pondville Hospital are at a low ebb at present, consequently suitable study cases appear very infrequently,” said O. E. Denney to N. B. Eddy, 13 November 1939, Projects: Development of Nonaddictive Analgesics: Clinical Studies: Marine Hospital, 1938-1939, CDA-NASA. Descriptions of individual patients and the problems in using them for analgesic studies can be found in the Committee’s correspondence. For example, in this letter, Denney describes one patient as being: “a highly emotional, unstable person.”
Thus, it was not only the CDA members’ ideas about objective analgesic testing that lessened the value of clinical studies. The actual lack of precision in the data they obtained, which resulted from their inability to control experimental conditions, also influenced their judgment of the objectivity of clinical testing. The value of Metopon continued to be uncertain; committee members resigned themselves to distributing limited amounts of the drug to clinicians in exchange for their impressions of its efficacy.\textsuperscript{53}

This is when the dolorimeter arrived. With its promise of precise data and easy implementation in humans, it contrasted sharply with the CDA’s clinical studies and their complexity. Unsurprisingly, CDA members were immediately enthusiastic; they invited Wolff to present his method in 1940, and researchers were soon dispatched to his Cornell laboratory to learn how to use it.\textsuperscript{54} Howard Andrews, a physiologist at the PHS Narcotic Hospital in Lexington, Kentucky, reported after his visit to Cornell: “I feel that this technique fills a long-felt need in our substitute drug program,” and was eager to begin work immediately “because of the fact that the complete evaluation of the addiction liability of a drug requires a knowledge of its analgetic power.”\textsuperscript{55} Committee members marveled at the consistency of the dolorimeter’s results, but also at its cheapness—the set-up could reportedly be duplicated for about $300, and even less, about $50, given that the Lexington laboratory already owned a potentiometer.\textsuperscript{56} They also hoped that their most accessible source of subjects, ex-drug users incarcerated at

\textsuperscript{56} L. E. Lee to W. C. White, 24 July 1939, Projects: Development of Nonaddictive Analgesics: Clinical Studies: PHS Hospital: Lexington General, CDA-NASA; noted “the uniformity of Dr. Wolff’s results with a small group of subjects is remarkable”; “Report to the Surgeon General on his visit to Wolff’s laboratory by H. L. Andrews, July 28, 1939,” likewise commented “The fact that the threshold has practically the same value for all normal individuals is another feature in favor of the method.” N. B. Eddy to Howard L. Andrews, 30 June 1939, Projects: Development of Nonaddictive Analgesics: Clinical Studies: PHS Hospital: Lexington General, CDA-NASA; H. L. Andrews “Report to the
Lexington, could be used for dolorimetric testing. The technique fit both the Committee's ideas about objectivity in pain measurement and the kinds of resources it commanded in terms of funds, subjects, and scientific expertise.

Wartime events soon interrupted the Committee’s activities, but also stimulated new interest in synthetic analgesics among American pharmaceutical manufacturers. Following the diffusion from Germany of information about the first fully synthetic opiates, meperidine (Demerol) and methadone, the pharmaceutical industry replaced the CDA as the main sponsor of analgesic testing. The arrival in the United States of meperidine in the early 1940s and of methadone just after the war was marked by clusters of reports in the scientific press, the majority of which originated in industry-sponsored research. These reports signaled the advantages of fully synthetic narcotics in the context of increased wartime needs for pain relief and possible threats to American opium supplies. The lucrative potential of strong synthetic analgesics was also seized by the pharmaceutical industry, which set its research departments to work on developing and comparing derivatives and variants of the German molecules, as well as various potentiating agents.

The dolorimeter proved to be popular for evaluating these new drugs.

---

57. L. E. Lee to W. C. White, 24 July 1939.
58. Demerol was also called meperidine in the United States, and known in various countries as dolantin (the name it was first given by the German researchers), isonpecaine, dolantol, dolasol, pethidine (often used by British researchers), mefedine, and lidol; this drug was first synthesized by Eisleb and Schaumann in Germany and approved in the United States by the Food and Drug Administration in 1942. Robert C. Batterman and Clifton K. Himmelsbach, “Demerol—A New Synthetic Analgesic,” J. Am. Med. Assoc., 1943, 122, 222–26. Methadone was initially called amidone by German chemists; information about this drug was released in the United States by the U.S. Department of Commerce just after the war. It was soon approved by the FDA under the trade name of Dolophine, manufactured by the pharmaceutical firm Eli Lilly and Company, which had been quick to synthesize and test the compound. C. C. Scott and K. K. Chen, “The Action of 1,1-diaryl-1-[(dimethylaminoisopropyl) butanone-2, a Potent Analgesic Agent,” J. Pharmacol. Exp. Ther., 1946, 87, 63–71.
Although the dolorimeter had never been extensively used under CDA sponsorship, the appeal it held for Committee members fore- shadows the interest of industry representatives. Both valued similar qualities in analgesic data: quick, cheap, precise, comparable to information about addictiveness, and, ideally, obtainable under labor- oratory conditions by nonclinical researchers. Pharmaceutical companies showed little interest in paying for expensive and time- consuming clinical studies, which had not yet proven capable of providing highly precise information about analgesics. The kind of researchers they paid to determine analgesic efficacy— pharmacologists and physiologists—faced similar, indeed even greater obstacles in running effective clinical studies than had Committee-sponsored investigators.61

In the late 1940s, during this burst of commercial interest in synthetic analgesics, the dolorimeter reached the peak of its popularity. Pharmaceutical companies’ interest in analgesics converged with their increased funding of academic research, especially in pharma- cology.62 The dolorimeter was the first method used to compare the effectiveness of meperidine (Demerol) and morphine in humans, and its use was paid for by Hoffmann–La Roche to investi- gate the potential of its meperidine-type (piperidine) derivatives.63 The dolorimeter was taken up by employees of Lilly Research Laboratories in 1946 to evaluate methadone and other promising compounds.64 Pharmaceutical firms, including Hoffmann–La Roche, Whitehall Pharmacal, and Smith, Kline & French, also funded the development and evaluation of the method itself.65 Industry researchers sang its praises: “Of the methods proposed and

61. While CDA coordinating investigators were also, for the most part, nonclinicians who lacked authority and control over clinical testing, at least their access to clinical set- tings was facilitated by Committee connections.


used in the last decade, it is safe to say that the Hardy–Wolff–Goodell procedure has gained the widest acceptance,” commented Miller of the Sterling-Winthrop Research Institute, a statement echoed by Chen of Lilly who declared it to be “the best.”

Even at the height of its popularity, few denied that using the Hardy–Wolff–Goodell method required adjustments and compromises. It was found to lack sensitivity to the effects of weaker analgesics such as aspirin. A more significant issue was the difficulty in replicating the consistency in threshold values obtained by Hardy, Wolff, and Goodell. Yet researchers readily adapted the method to include a period of training for both human and animal subjects in order to obtain stable and reliable threshold values. Only later, as we have seen, was this practice criticized as a serious and potentially invalidating flaw in the dolorimetric method. During the early postwar expansion in analgesic development, subject training was seen as a legitimate, unexceptional means of managing variability in the measurement of analgesia. It was possible to argue for the epistemological value of training; it echoed with a conception of scientific observation that had been common in earlier psychophysical experimentation, in which judgment was enhanced by individual subjects’ expertise and education in the judgment of inner sensations. Yet training also had practical implications; it allowed for a reduction in the number of subjects needed for drug evaluation, and was easy to implement with the kinds of subjects—research colleagues, medical students—that were submitted to dolorimetric testing. Less variation in results also meant that fewer trials were needed to obtain interpretable data.

67. Miller, “A Critique,” 34–50. The lucrative market for weaker over-the-counter analgesics may have made this a big disadvantage from the perspective of the pharmaceutical firms.
69. Hardy, Pain Sensations, 82–83, explicitly calculated how training lowered the threshold of significance of data: 5–10% changes in pain intensity for trained subjects, in
These were the kind of data for which pharmaceutical companies were willing to pay in the 1940s. The affordability of the dolorimeter did not depend only on the price of the apparatus ($850 from the Co-Design Corporation, or $300 to build in 1940). More importantly, the instrument could be operated by pharmacologists or physiologists with presumably modest grants, given the amount of time (a matter of days) and subjects (as few as three) required to complete a trial. It could be used in academic settings or pharmaceutical company laboratories.

The dolorimeter fit into a gap between animal studies (convenient, relatively cheap and easy to standardize, but with limited translatability to humans) and clinical ones (perhaps more authentic, but much more difficult, and expensive, to standardize). In the early 1940s, this gap—which was epistemological, but also financial, logistical, and professional—had been particularly wide. When, rarely, clinical studies were conducted, investigators found it difficult to recruit subjects and control for patients’ pathologies and “psychological make-up”; they did not use placebo comparisons despite calls for doing so. Thus, while clinical studies were seen as necessary for confirmation of therapeutic efficacy, they were not considered to be adequate substitutes for obtaining initial quantitative information about the potential value of new analgesics. Very soon, by the end of the decade, the gap between laboratory and clinical methods was narrowed by the investment of new kinds (in both quantity and quality) of resources for analgesic evaluation.

---
70. Sales pamphlet, Wolff Papers.
WORKING ON THE ANALGESIC CLINICAL TRIAL, DISPLACING THE DOLORIMETER

Various factors converged in the late 1940s to facilitate the emergence of the clinical trial as an effective technology capable of displacing the dolorimeter in analgesic evaluation. New actors—the U.S. Army, clinicians, as well as the members and allies of the newly renamed Committee on Drug Addiction and Narcotics (CDAN, successor to the CDA)—became involved in analgesic testing. They mobilized new resources and reoriented existing sources of funding, namely the pharmaceutical industry, toward the development and implementation of clinical testing methods. These resources—money, but also clinical authority and skills—were converted into concrete experimental conditions that increased the possibilities of control over clinical analgesic testing, and thus the quality of results generated in clinical settings. As the analgesic clinical trial emerged as a reliable and precise measuring technology, its advocates—especially Henry K. Beecher—were able to persuasively pit it against the dolorimeter in terms of validity, models of pain, and types of objectivity. A few years earlier, this would have been impossible.

An early newcomer to analgesic testing was the U.S. Army. World War II had raised concerns about the blockage of opium supplies, creating an interest among military authorities in evaluating methadone as a potential substitute for opiates. The Army expressed the need for a different type of information about analgesic efficacy from that which had previously been sought by the CDA and the pharmaceutical industry. In 1947, the Surgeon General of the Army noted the lack of “extensive clinical and practice experience” with synthetic painkillers. He seemed more interested in the clinical acceptability of the drug than in the precise ratios of analgesic to addictive potency.

This concern may have motivated the choice of a clinician, the Harvard anesthesiologist Henry K. Beecher, by the Medical Research and Development Board (MRDB) to evaluate the efficacy of methadone in the late 1940s, although this engagement may

simply have followed from Beecher’s previous connections with the MRDB.\footnote{73} In any case, there are good reasons why Beecher might have been interested in perfecting and promoting a clinical method for testing analgesics. Beecher, like other American anesthesiologists at the time, was concerned with improving the status, autonomy (especially from surgery), and expertise of his specialty. Upon taking up the post of Chief of the Anesthesia Service at Massachusetts General Hospital, he created a laboratory that he built up by securing new sources of research funding and projects.\footnote{74} By demonstrating that anesthesiologists could be researchers, particularly clinical researchers, as well as experts on pain, Beecher sought to expand professional opportunities for himself and his colleagues. During a time of expanding opportunities in analgesic research, Beecher was keen to claim that clinical pain was more authentic than experimental pain, that clinical evaluation could accurately determine analgesic effectiveness, and that postoperative patients were the most suitable subjects for such studies. Inviting fellow anesthesiologists to take up the clinical evaluation of drugs and especially analgesics, Beecher participated in an effort to emphasize the profession’s clinical skills, ability to contribute to the advancement of science, and expertise in pain-management, particularly during the postoperative period.\footnote{75}

\footnote{73} William S. Stone to Henry K. Beecher, 5 January 1950, Box 22, Folder 15, Henry K. Beecher Papers (HMS c64), Francis A. Countway Library of Medicine, Harvard Medical School, Boston, Massachusetts (hereafter, Beecher Papers). Beecher’s correspondence with the MRDB does not indicate, however, whether his sponsors specifically requested him to carry out this evaluation in a clinical setting. This funding may simply have been an extension of a larger grant for the Study of Sedatives provided to Beecher, on the basis of his expertise as an anesthesiologist and his military connections, to study other matters of military interest such as the use of barbiturates.


Before promoting the analgesic clinical trial, however, Beecher had worked on it. As a clinician and chief of a hospital anesthesia service, Beecher brought new resources to analgesic evaluation. Access to many suffering clinical subjects (postoperative patients), and authority over the treatment of their pain, as well as over both clinical and research staff, provided Beecher with more control over experimental procedure than previous investigators, often nonclinicians, had had. To reinforce this control, Beecher recruited crucial allies: technicians who were hired specifically to gather data according to standardized procedures, and statisticians who collaborated in designing experiments and processing results. This assistance could only be paid for with large and relatively long-term grants.

The MRDB grant allowed Beecher to begin standardizing his method (in collaboration with Jane Denton). He then attracted the attention and eventually the resources of the new CDAN. When the Committee met again in 1947, the American pharmaceutical industry had taken over the job of analgesic drug innovation. The new role of CDAN was advisory; it was to provide federal agencies such as the FDA and the Army and Navy with impartial information about the benefits and dangers of new analgesic drugs. Thus, the old CDA’s focus on innovation in analgesic drugs shifted with CDAN to innovation in testing methodologies. Making drug testing more rigorous was meant to remedy the perceived inability, discussed during CDAN’s first meetings, of physicians to judge therapeutic efficacy on the basis of clinical experience, and of pharmaceutical manufacturers to sponsor reliable drug evaluations without external guidance.

Such suspicions, especially when voiced by the secretary of the AMA’s Therapeutic Trials Committee (TTC) who attended CDAN’s first meeting, link the Committee’s quest for reliable drug testing methods with a movement of therapeutic reform in American medicine. Analyzed by Harry Marks, this movement was spearheaded by an elite group of “therapeutic reformers”—

76. The CDA clinical studies, for example, were coordinated by the pharmacologist Nathan B. Eddy. The largest studies at Pondville Hospital were implemented by CDA researchers, initially by a nonclinician, and then by a clinician without a professional appointment at the site of study. Several of the clinical studies carried out in the 1940s were carried out by pharmacologists. See also Meldrum, “Each Patient.”
including members of the TTC—who campaigned for the adoption of specific criteria of drug evaluation to protect therapeutic decision-making from the influence of the pharmaceutical industry.\textsuperscript{77} Within CDAN’s framework, this movement also converged with regulatory concerns. For the Committee’s members and its federal allies—the PHS, FDA, and the Bureau of Narcotics—unchecked commercial aspirations brought additional threats to public health. Reliable drug testing methods were needed to protect American consumers from exposure to therapies with unfavorable ratios of addictiveness to efficacy. In the postwar years, after the successful wartime implementation of large, cooperative trials of antibiotics, therapeutic reform became associated with the methodological devices of the clinical trial—randomized selection of comparison groups, double-blinding, placebos, and statistical analysis.\textsuperscript{78} Beecher was able to tap into this movement to promote his own innovations in analgesic testing. In 1949, his army-funded trials of methadone, conducted with Jane Denton, were published in the \textit{Journal of the American Medical Association}. The articles were invited and prefaced by the TTC secretary, with the Council’s approval of the method as a “distinct advance in the methods available for quantitative evaluation of the therapeutic efficacy of [analgesic and narcotic drugs]).”\textsuperscript{79}

As they began searching for suitable grantees in 1948, CDAN’s members were impressed by Beecher’s work. They noted in particular his use of paid technicians as observers, an innovation enabled by his MRDB grant. The decisive feature of Beecher’s method, however, may have been its reliance on the use of large numbers of subjects. This now appeared to be essential to reliable drug testing. Indeed, what had recently been seen as the dolorimeter’s main advantage—its suitability for use with a small numbers of subjects—was denounced by Committee members as its main flaw. As a clinician-researcher himself, Isaac Starr, the Committee’s new chairman, may also have encouraged the selection of Beecher’s clinic-based method for CDAN sponsorship. Starr was obviously familiar
with the methodological logic of the clinical trial. It was upon his suggestion, made during a CDAN meeting, that Beecher adopted the use of cross-over analysis (comparing different drugs sequentially within the same subject, rather than in separate comparison groups), in which randomization would be applied to the order of drugs given to a single subject rather than to subjects’ allocation in comparison groups. This modification would become a cornerstone of Beecher’s method.\(^{80}\)

Committee members decided to approach the pharmaceutical industry—despite their possible misgivings about its commitment to objectivity in drug evaluation—for contributions to their research grant program. In fact, industry representatives expressed a strong interest in improving analgesic testing, particularly in humans. Why, in 1949, were they no longer satisfied with the dolorimeter, which had so recently been praised by industry researchers? There may have been a growing sense that the kind of information provided by the dolorimeter, obtained quickly from only few subjects, would soon no longer be sufficient for purposes of drug innovation, marketing, and eventually, regulation.\(^{81}\) Clinical trials may also have been seen as able to simultaneously evaluate efficacy and safety, although Beecher conducted separate studies of analgesia and side effects. Still, drug manufacturers did express a preference for cheaper, laboratory-based methods—something along the lines of an improved dolorimeter—over Beecher’s expensive and time-consuming design. However, CDAN’s members explicitly excluded industry sponsors from the selection and guidance of funded research.\(^{82}\) And these sponsors continued to channel

---

80. Meldrum, “Each Patient.”

81. At this time, drugs did not have to be proven effective in order to be approved by the FDA. Soon afterwards, however, the Durham-Humphrey Amendment Act of 1951 to the Food, Drug and Cosmetics Act of 1938 codified the categorization of prescription-only and over the counter drugs. For strong analgesics, the determination of addictiveness, which was a major component of the drug’s safety, was closely tied to the minimal effective dose. Getting a strong analgesic approved as an over-the-counter drug had significant commercial implications for pharmaceutical firms, as shown by the case of dextropropoxyphene (Darvon). See Meldrum, “Departures from the Design.”

82. Beecher’s research continued to be funded even after CDAN’s industry sponsors began to complain about its lack of immediate practical application and expressed a desire to initiate a re-exploration of the more economical laboratory methods of analgesic testing. N. B. Eddy to Starr, 11 May 1953, Box 2: General, 1947–June 1959: CDAN: Division of Medicine: NASA (hereafter CDAN-NASA).
their funds through CDAN, possibly for its scientific authority but also because of the benefits of pooling their contributions with their competitors’. This produced larger grants than individual companies were willing to support. More time and money did indeed have an impact on the feasibility, as well as the quality, of clinical analgesic evaluation.

Belief in the value of clinical trials, the orientation of funds—by the MRDB and CDAN—toward the implementation of clinical analgesic evaluation, and the actual effectiveness of Beecher’s method were mutually reinforcing. CDAN grants enabled Beecher to refine and diffuse the method he developed under Army sponsorship. He continued to provide salaries to technicians, who were more amenable to standardization and likely to enforce controls such as double-blinding than busy ward staff. Beecher also drew on his grant money to hire the services of a statistician, Frederick Mosteller, to help design and analyze the studies. Statistical analysis, standardized observation procedures, and access to large amounts of subjects were essential given Beecher’s conceptualization of the experimental subject as a collective one, whose main virtue was abundance rather than stability, certainty, or detachment. The precision and accuracy of the analgesic clinical trial did not rely on individual judgments of pain relief but on their aggregation.

By tinkering with explicit as well as tacit experimental practices, Beecher was able to improve the consistency of his data. In his publications, Beecher implied that his experimental design was so

83. “Minutes of Conference with Representatives of Drug Manufacturers, July 1, 1949,” Box 1: Minutes, July 1949 Conference with Representatives of Drug Manufacturers, CDAN-NASA.
84. A significant portion of these grants went to technicians’ (observers) salaries. For example, in a 1952 grant application, Beecher requested $5,980 for this purpose out of a total of $16,679 (part of this budget was for a separate study of side effects, so the salary’s proportion of the cost of clinical trial work was even higher). See H. K. Beecher, “Annual Report and Application for Renewal of Support to the Committee on Narcotics and Drug Addiction of the National Research Council, 21 January 1952,” Bulletin of the Committee on Drug Addiction and Narcotics, 1952, 225.
85. Beecher, “Annual Report,” 222–23, said “the aid of a top level statistician is essential. Professor Mosteller has become interested in this problem and if an application is approved will be available for this aspect of the work.” Harry Marks emphasizes the role of alliances between clinicians and statisticians in the postwar emergence of clinical trials, and draws attention to the mutually beneficial nature of this alliance by describing it as a “gift relationship,” The Progress of Experiment, 148–55.
powerful that any alert, suffering patient (as long as there were enough patients in the study) was suitable for the method. We have seen that in practice, however, certain kinds of subjects were excluded from trials, most notably women. The time and freedom provided by CDAN sponsorship also allowed Beecher to run experiments to determine the influence of factors such as the placebo response on trial results; he had initially intended to screen out “placebo-reactors.”

By the early 1950s, Beecher had gathered enough evidence in order to persuasively argue that it was possible to extract accurate information about analgesics even from sick patients’ vague and variable judgments of their idiosyncratic experience of pain. These data did not come cheaply or easily; Beecher agreed that this was “painstaking and tedious work.” Yet, he worked tirelessly to convince both sponsors and fellow investigators that it was worth investing in “a costly field, but one that promises to yield on cultivation an astonishingly rich harvest.”

Indeed, Beecher’s methodological innovations shaped CDAN-sponsored studies of analgesic evaluation for two decades. The analgesic clinical trial was taken up by Beecher’s assistants, anesthesiologist Arthur Keats, and the clinical pharmacologist Louis Lasagna, and was adapted by Raymond Houde for use in cancer patients at Memorial Sloan-Kettering Cancer Hospital. The work of the Memorial and MGH teams served as a model for a cooperative study of analgesics in VA hospitals that ran in the 1960s. Beecher and his colleagues’ work on analgesics also inspired the evaluation of psychiatric drugs. They showed that simple rating scales, when applied and aggregated under the right conditions, could provide reliable information about subjective drug effects. Harry Marks has shown that therapeutic reformers relied on statistical concepts and allies to argue that simply adding up specific objective measurements of pathological change was not sufficient to guarantee the objectivity of drug evaluation. Statistical validation at the aggregate

level was necessary to objectify—for judging general therapeutic efficacy—measures of blood counts, temperature, or even death.90 When no such physiological measures of drug effect were available, as in the case of analgesics or antidepressants, the possibilities offered by aggregate-level objectivity were even more promising. Beecher was able to show that something as simple and imprecise as number-coded impressions of pain relief could serve to consistently and precisely measure changes in collective pain. This played a significant role in the rise of the large and lucrative business of assessing psychiatric drugs, particularly antidepressants.91

CONCLUSION

In the early to late 1940s, and then from the late 1940s, two different configurations of money, research settings, experimenters, subjects, expertise, and demands favored different types of analgesic-testing technologies. The Hardy–Wolff–Goodell dolorimeter was well suited to the production of relatively cheap but low-volume data using few subjects. The method offered results that were highly precise at the level of the individual; this precision was obtained by a focused type of psycho-sensory control over the judgment of pain that combined instrumental means with a learning process. By contrast, Beecher’s analgesic clinical trial produced high volumes of relatively expensive data. It exercised collective control through procedural and statistical means that depended on money and rules, as well as the authority to enforce them, targeted toward the coordination of multiple actors and the information they produced.

By controlling the evaluation of pain relief in different ways, each method was designed and operated on the basis of a different notion of what pain was, and especially of how the experience of pain was modulated by experimental variables and interventions. Hardy, Wolff, and Goodell maintained that the experience of pain could be divided, in theory but also in practice, into sensation and reaction. Pain sensation was conceptualized as a straightforward trajectory of impulses traveling from the stimulus through the sensory apparatus to the perceiving subject. In this model, the intensity of

90. Marks, *The Progress of Experiment*.
stimulation bore a stable relationship to perceived intensity. That is, as long as reactions to the stimulation were eliminated through the use of an appropriate stimulus and of experienced, normally constituted and self-disciplined subjects.

Beecher, however, argued that pain was indivisible at the level of experience. The reaction to painful stimulation came into play as soon as the organism began processing sensory information; it preceded conscious awareness of an experience of pain. Thus, the essential sameness of sensitivity to pain in different subjects was theoretical; in contrast, the experience of pain, which integrated subjects’ reactions, was for Beecher, “never alike for any two individuals and, indeed, with the passing of time and accumulation of life experience, is never exactly the same for the same individual from one time to another.”

This idiosyncratic reaction was thus an unavoidable and even crucial component of any experience worth calling pain. The dolorimetric pain threshold, as Beecher saw it, was in reality a mixture of sensation and reaction that paraded itself as pure sensation; in reality, there was no direct or predictable correlation between intervention and experience. Too many factors other than the intensity of stimulation influenced the experience of pain. Indeed, Beecher firmly rejected the principle of proportionality between the magnitude of injury or stimulation and the amount of pain experienced.

Marcia Meldrum has suggested that Beecher adopted this position as a response to his observations during World War II. In the course of his military service, Beecher noticed that many badly wounded soldiers refused morphine and denied experiencing pain. He concluded that strong emotion, rather than any modification of sensitivity, accounted for this disjuncture between wound severity and the intensity of pain. In the postwar years, there were other

95. Though the specific explanation given by Beecher was later replaced by reference to the action of endogenous opiates (such as endorphins), this study would often be cited by researchers who sought to reform conceptions of pain by pointing out that sensory-physiological events were only a small, sometimes insignificant part of the total subjective pain experience.
sources for new, psychologically saturated models of pain experience. An expansion in psychosomatic research—which became better funded and defined in the United States from the late 1930s—stimulated experiments on how personality and psychopathologies modulated responses to pain.\textsuperscript{96} Greater experience of and attention to chronic conditions such as low back pain, along with the development of psychiatry, seem to have brought clinicians to reflect on psychogenic and psychologically influenced experiences of pain.\textsuperscript{97} More specifically, two American doctors confronted with many cases of nerve-injured and chronically suffering patients, especially during World War II, searched for psychological components in the etiology of persistent pain. Influenced by William K. Livingston, the anesthesiologist John J. Bonica developed the notion of chronic pain as a clinical entity in itself—rather than simply a symptom of an underlying somatic condition; its etiology and treatment integrated significant psychological dimensions. The publication of Ronald Melzack and Patrick Wall’s gate-control theory of pain in 1965 drew on earlier research, giving rise to a more dynamic and integrative model of the nervous system; pain signals transmitted from the periphery appeared to be modulated at various levels before they reached consciousness. Thus, Melzack and Wall emphasized the central processing of pain stimuli—which integrated memory, interpretation, emotion, and personality—over the peripheral transmission of nervous information.\textsuperscript{98}

Still, the ways in which pain was defined for and by analgesic testing practices did not simply result from these shifts. First of all, it

\textsuperscript{96} For example, Malmo et al., “Standardized Pain”; Chapman et al., “Measurements of Pain.”


\textsuperscript{98} Baszanger, \textit{Inventing Pain Medicine}, describes how Bonica defined a new clinical entity, chronic pain, which was characterized not only by its duration but by its psychological components, which developed over time even if the pain had a physical origin. This type of pain affected the whole person and thus required a multidisciplinary clinical approach in a team including psychologists and psychiatrists. See also William K. Livingston, \textit{Pain Mechanisms: A Physiologic Interpretation of Causalgia and Its Related States} (New York: Macmillan, 1944); John J. Bonica, \textit{The Management of Pain with Special Emphasis on the Use of Analgesic Block in Diagnosis, Prognosis, and Therapy} (Philadelphia: Lea & Febiger, 1953); Ronald Melzack and Patrick Wall, “Pain Mechanisms: A New Theory,” \textit{Science}, 1965, 150, 971–79.
is misleading to follow Beecher in representing Hardy, Wolff, and Goodell’s conception of pain as anachronistic, mechanistic, and merely psychophysical. Indeed, Wolff and his colleagues were active in the development of psychosomatic research on stress, personality, emotion, pain, and chronic disease; they participated in redefining pain as a psychologically complex and highly personal experience. Yet, they claimed that the objectivity of analgesic evaluation depended on generating a less personal and therefore less variable experience of pain. They succeeded in doing so (as did several other teams of dolorimetric users who sought this kind of objectivity) through specific types of stimulation, subjects, instructions, practices, and expectations.

The idea that pain’s subjectivity was too complex to be controlled directly was not new in the late 1940s. Researchers who had failed to obtain consistent pain thresholds had already concluded this: reporting to the CDA in the mid-1930s, the Director of Laboratories in Lexington Hospital explained fluctuations in response to electric shocks by stating that: “judgemental and interpretive factors were of more significance in producing pain than any particular degree of stimulation.”99 The novel dimension of Beecher’s notion of pain was not that it was complex and personal, but that its relief could be reliably, consistently, and precisely quantified without being reduced to something simpler, less personal, and more predictable. In other words, Beecher made this definition of pain manageable for analgesic experimentation through the use of specific technologies of data collection and analysis, of practices that depended on certain kinds of authority, as well as material and human resources.

Beecher’s rich conception of pain not only justified the analgesic clinical trial; it was also justified by it. While the variability of experiences of pain and its relief could not be directly modified, it nevertheless presented itself in regular patterns in the bird’s eye view. The distribution of clinical data on analgesic efficacy was normal.100 It clearly distinguished between placebos and active

99. E. Williams, “A Quantitative Measure of Analgesia (Summary of Work Done at Lexington in Past Year, November 12, 1938)” in L. Kolb to W. C. White, 12 November 1938, Projects: Development of Nonaddictive Analgesics: Clinical Studies: PHS Hospital: Lexington General, CDA–NASA.
drugs. And it turned out, with the investment of time, effort, resources, and contacts, to be reproducible.\footnote{101}

However, obtaining this bird’s eye view was resource-intensive. Researchers, including some who had strongly advocated analgesic clinical trials, did not abandon hope of developing effective laboratory-based pain-measuring techniques. Even Beecher, who had declared experimental pain to be incommensurable with clinical pain in the early 1950s, sought, in the 1960s, to recover a form of experimental pain that could be more easily controlled and manipulated for analgesic testing. The method he developed was a tourniquet pain test, in which the subject was required to perform a certain amount of muscular work with limbs to which the blood supply was cut off. This induced an experience of pain which, because it was sustained and included an element of anxiety, more successfully mimicked clinical pain. But it was not clinical pain: it could be produced and manipulated predictably under laboratory conditions, at an individual level, in normal subjects. Such a test was important, Beecher explained in 1966, because: “practically, there is great need for a method of appraising in man, conveniently and accurately, the effectiveness of new pain-relieving agents. There is great need for a method which will not require the tedious use of pathological pain. Thus, the present findings appear to have wide usefulness to the pharmaceutical industry.”\footnote{102} Conscious of the social and material realities of analgesic testing, and of the demand for quicker, cheaper methods, Beecher had been willing to attempt to redefine what counted as an acceptable and quantifiable form of pain.

\textbf{Acknowledgements}

Many thanks to Thomas Schlich, Cornelius Borck, George Weisz, and anonymous reviewers for their comments. I also thank colleagues from the Social Studies of Medicine department at McGill University, the Medicines as Social Objects (MEOS) research group at the University of Montreal, and the Anthropologies of African

\footnote{101. The broader diffusion of the analgesic clinical trial relied on CDAN sponsorship and networks of communication; the first researchers who took up the method worked directly with Beecher and/or were funded by CDAN and participated in its meetings.}

Biosciences group at the London School of Hygiene and Tropical Medicine.

FUNDING

I gratefully acknowledge the financial support of the Fonds Quebecois de Recherche sur la Société et la Culture, Associated Medical Services, Inc., and the Leverhulme Trust.