Regulation of drugs and devices: an evolution

R A Merrill

Health Affairs 13, no.3 (1994):47-69
doi: 10.1377/hlthaff.13.3.47

The online version of this article, along with updated information and services, is available at:
http://content.healthaffairs.org/content/13/3/47

For Reprints, Links & Permissions : http://content.healthaffairs.org/1340_reprints.php

Email Alertings : http://content.healthaffairs.org/subscriptions/etoc.dtl

Not for commercial use or unauthorized distribution
To Subscribe: https://fulfillment.healthaffairs.org

Health Affairs is published monthly by Project HOPE at 7500 Old Georgetown Road, Suite 600, Bethesda, MD 20814-6133. Copyright © by Project HOPE - The People-to-People Health Foundation. As provided by United States copyright law (Title 17, U.S. Code), no part of may be reproduced, displayed, or transmitted in any form or by any means, electronic or mechanical, including photocopying or by information storage or retrieval systems, without prior written permission from the Publisher. All rights reserved.

Not for commercial use or unauthorized distribution
Prologue: The Food and Drug Administration (FDA) is a federal regulatory agency that has left a tangible mark on American life. Through a long, controversial, but often distinguished history, the agency has granted approval to products that it considered worthy of its imprimatur while rejecting others that failed to pass the rigorous regulatory tests applied. Of a wide variety of products that fall under its purview, pharmaceuticals and medical devices receive substantial scrutiny, although they are formally subject to different legal standards. In this essay Richard Merrill, the Daniel Caplin Professor of Law at the University of Virginia in Charlottesville, discusses the evolution of the agency’s regulation of drugs and devices. As a former chief FDA counsel (1975-1977) during parts of the Ford and Carter administrations, Merrill is uniquely qualified for the task. He not only knows the agency from the inside, but he also represented the Carter administration in negotiating the content of the Medical Device Amendments of 1976 with congressional staff and directly with Rep. Paul G. Rogers, who at the time chaired the House Energy and Commerce Subcommittee on Health and the Environment. Merrill has the reputation of being an eloquent writer who deals with complex legal material in an understandable fashion. He is a member of the Institute of Medicine (IOM) of the National Academy of Sciences and previously served on its Governing Council. Merrill was a member of the IOM’s Committee to Study FDA Use of Advisory Committees and also chaired the IOM’s Committee on the Nutrition Components of Food Labeling. Merrill, a Rhodes Scholar at Oxford University, graduated with honors from Columbia University and its law school.
Abstract: The U.S. Food and Drug Administration’s (FDA’s) drug approval process has evolved from a system in which a drug could lawfully be marketed unless the FDA were able to prove that the manufacturer knew that the drug would not work for the conditions for which it was promoted to one in which drugmakers require advance approval from the agency for almost every important step in testing, production, and marketing. The more modern system for regulating medical devices is the product of amendments to the federal Food, Drug, and Cosmetic Act in 1976. In those amendments, Congress sought to create a framework for control of device technology that would also facilitate innovation. This paper suggests that, notwithstanding this aspiration, both external pressures and internal practices are inexorably bringing device regulation closer to the “drug model.”

The U.S. Food and Drug Administration (FDA) exercises enormous authority, ranging from food labeling and composition to the development and marketing of virtually all medical products. Its jurisdiction embraces products whose annual sales account for a quarter of total consumer spending in this country. The agency has a reputation for being cautious in its regulation of drugs and medical devices, a reputation that is understandable, given its responsibility for ensuring the safety and efficacy of increasingly powerful chemical and biological agents and technologically sophisticated medical instruments.

This paper depicts the evolution of two systems of government regulation of medical products. The highly visible and politically charged development and marketing of new drugs for acquired immunodeficiency syndrome (AIDS) introduced many Americans to the complex process of drug regulation and approval. Federal regulation of drugs began in 1906, but the present system is the product of legislation passed by Congress in 1938 and significantly amended in 1962. At its core, the drug regulatory system requires FDA approval for every important step in the product development process. Medical devices are regulated under more recent legislation, the Medical Device Amendments of 1976, in which Congress attempted to combine rigorous review of high-risk devices with more modest requirements for most other products and, at the same time, encourage innovation. An examination of the FDA’s administration of the 1976 amendments, and particularly its actions of the past five years, reveals a system that seems to be moving inexorably toward the “drug model.” The expense and delay implied by this model justify concern for the future of innovation in medical device development.

Evolution Of The ‘Drug Model’

The 1906 Food and Drugs Act, the first major federal law governing therapeutic drugs, gave the FDA the authority to interdict the marketing of drugs that were adulterated or misbranded. A misbranded drug was one whose label bore “any statement . . . regarding such article, or the ingredi-
ents or substances contained therein which shall be false or misleading in any particular.”¹ The government argued in United States v. Johnson, the first major test of the 1906 act, that the quoted language was meant to prohibit the marketing of a drug for which deceptive therapeutic claims were made.² The Supreme Court, however, held that the statute was not aimed at “all possible false statements, but only at such as determine the identity of the article.” Though describing a drug as containing an ingredient that it lacked constituted misbranding, it was not misbranding to claim, falsely, that it cured cancer.

The Johnson decision prompted Congress to amend the 1906 act one year later by changing the definition of misbranded to read: “If [a drug’s] package or label shall bear or contain any statement . . . regarding the curative or therapeutic effect . . . which is false and fraudulent” (emphasis added).³ This guarded language apparently was thought necessary to meet concerns that the government could not constitutionally punish genuine but false opinions, and it proved a weak tool for controlling the marketing of drugs bearing deceptive therapeutic claims.

From the perspective of those who believed that the government should have authority to prevent the marketing of unsafe or ineffective drugs, the 1906 act had two deficiencies. It allowed the FDA to act only after a drug was marketed. Also, to prevent the marketing of an ineffective drug, it was not enough to show that the product did not work; the agency had to prove that the seller knew that the claims it made were false. The 1938 Federal Food, Drug, and Cosmetic Act (FD&C Act) took the first major step to remove these obstacles to regulation.⁴

**Invention of “new drugs.”** The most important change-requiring premarket review for safety—was a response to the infamous Elixir Sulfanilamide disaster, in which more than 100 residents of Tennessee were poisoned by the solvent incorporated without testing in a new therapeutic potion. The 1938 FD&C Act defined a new category of products—“new drugs”—which could not be marketed without first notifying the FDA and allowing the agency time to assess their safety.⁵

This was the beginning of the modern system of premarket review for prescription drugs—a system in which marketing without FDA approval is now unlawful. The original premarket notification requirement, however, was more modest. It did not apply to drugs that were not new drugs, and it allowed manufacturers to decide for themselves whether a product fell within that category. Furthermore, the FDA’s authority was officially confined to assessing whether a new drug was safe.

The 1938 act did, however, relax the standard that the government had to meet to prove that a drug was misbranded. A seller’s genuine belief in the product’s value was no longer relevant. The question was whether the
The 1938 act made major changes in the FDA's regulation of drugs. Manufacturers more commonly consulted with the agency before marketing a new product, and the agency became increasingly involved in overseeing the design and conduct of clinical trials of experimental drugs. While the FDA's authority was nominally limited to reviewing the safety of new drugs, agency reviewers often felt obliged to consider new drugs' effectiveness as well, on the premise that whether a biologically active agent could be considered “safe” depended on whether it could provide benefits that outweighed the inevitable risk.

**Comprehensive licensure.** The modern U.S. drug regulatory system has its roots in amendments to the 1938 FD&C Act that Congress passed a generation later, partly in response to the grim effects of thalidomide. Although the 1962 drug amendments purported simply to elaborate the new-drug approval system, they in fact transformed it.

First, the amendments raised the standard a new drug must satisfy prior to marketing by explicitly providing for FDA review of effectiveness as well as safety. Moreover, the effectiveness requirement dramatically expanded the scope of the new-drug review process. While the rate of invention of active ingredients was increasing during the 1950s, it might never have generated a massive workload of drugs whose safety required evaluation. But for a single active ingredient many claims of therapeutic utility were possible—each of which was a “new-drug” claim that demanded review.

Second, the 1962 amendments converted what had in form been a premarket notification system into a premarket approval system, in which the maker had to wait for agency officials to affirm a drug’s safety and effectiveness. Drugmakers thus became prisoners of the agency’s indecision, its preoccupation with other issues, or its lack of resources.

The third major change was to enlarge the FDA’s authority over clinical tests of new drugs. Section 505(d) of the act specifies that a drug’s effectiveness must be shown by “substantial evidence,” defined as “evidence consisting of adequate and well-controlled investigations, including clinical investigations, by qualified experts.” And section 505 (i) gives the FDA authority to establish standards under which experimental drugs may be shipped to clinical investigators. In interpreting the “substantial evidence” requirement and setting standards for “investigational exemptions,” the FDA has become the most influential source of guidance on the design of drug tests in the country, perhaps in the world. And it is also centrally concerned with the manner in which clinical trials are carried out. This concern extends to both the protection of experimental subjects and the recording and reporting of study results.

What have been the systemic effects of these changes in the system for
drug regulation launched in 1962? The most obvious is the most fundamental. Under the 1906 law and the original 1938 act, the FDA exerted control over the therapeutic claims made for a drug, if at all, after it reached the market. Since 1962 the FDA’s primary function has been to judge, on the basis of evidence that the maker must supply, whether a new drug works. This simple shift in the burden of proof transformed the way in which drugs are developed, tested, and marketed.

Accompanying this shift was a more subtle change in the FDA’s view of its consumer protection role. The FDA has authority to prevent harm before it occurs. In the agency’s own publications, in press accounts of its performance, and in its dialogue with Congress, agency reviewers are reminded that they share with the manufacturer responsibility for any drug that causes harm or fails to work.

Emergence Of The New-Drug System

While the basic elements of the FDA system of drug regulation had been established by 1962, many important details are products of later legislation or initiatives of the agency itself.

Expansion of jurisdiction. The premarket approval requirement was from the outset limited to “new drugs.” A new drug was defined after 1962 as one that was not generally recognized by experts as safe and effective for its labeled uses. Although this definition embraced essentially all novel active prescription drug ingredients introduced after 1938, the act excluded several categories of drugs and left the status of others uncertain. The FDA’s early implementation of these provisions was marked by a series of efforts to narrow the exceptions and to confirm the new-drug status of products about which the statute left doubt. It is a complicated story, but readers should be familiar with the results of the FDA’s effort and appreciate its lessons.

Two concerns animated the FDA’s efforts to bring all prescription drugs within its “new-drug” jurisdiction. First, the agency could see no logic in allowing “me-too” copies to escape, even temporarily, the limitations that applied to the pioneer product. Second, it was efficient to determine what limitations ought to apply to a drug (and all copies) and implement these administratively by modifying the terms of its new drug application (NDA). Then, any deviation would automatically render the product illegal.

One cannot overstate the significance of the shift in regulatory leverage that has resulted from Congress’s adoption of premarket approval for prescription drugs and from the FDA’s successful efforts to extend its coverage. From an environment in which drugmakers could market any product that the government was unable to prove in court was or bore claims that were knowingly false, we have moved to a system in which no prescription drug
may be marketed unless and until the FDA is convinced that it is safe and
effective for the uses that the agency will allow on the label. Furthermore,
the FDA takes the position that virtually any change in an approved new
drug requires advance approval. Not only attempts to expand indications
but more modest changes in labeling, ingredients, the method or even the
location of manufacture, or packaging must first undergo FDA review.

**Oversight of clinical investigations.** The act prohibits the interstate
shipment of any new drug for which the FDA has not approved an NDA.
Since 1938, however, the agency has been empowered to grant exemptions
for investigational drugs. From this authority have grown two types of
regulatory requirements governing clinical studies.

One set of requirements is designed to assure the integrity of clinical
trials. The core of these is the requirement that investigators secure and
doctor the informed consent of trial subjects. The FDA has supple-
mented this requirement with a mandate for review by a local institutional
review board (IRB), and, to facilitate monitoring of compliance with both
requirements, the agency has established detailed specifications for IRB
operations and record keeping. The second set of requirements is intended
to increase the likelihood that clinical trials will produce acceptable evi-
dence of a drug’s safety and effectiveness. These requirements are set forth
in regulations; in test guidelines for specific therapeutic classes; and in FDA
reviewer critiques of clinical protocols.

These diverse instructions constitute a growing body of FDA “law”
governing drug trials. While agency officials claim that these instructions
embody what independent experts consider to be sound experimental de-
design, many manufacturers and clinicians contend that FDA reviewers tend
to demand more elegant studies than are needed to support sound judg-
ments about effectiveness. The FDA assumes, of course, that any trial will
follow the protocol and that the records submitted to the agency will
truthfully reflect the observations of investigators. But this assumption is
not simply a matter of faith; it is bulwarked by an elaborate system for
monitoring the veracity of drug sponsors and clinical investigators.

**Assurance of production capability.** A drug is deemed misbranded if it
is not produced in accordance with “good manufacturing practice”
(GMP). Not content to rely on discovery of substandard products or
routine inspections to enforce this requirement, the FDA has promulgated
regulations specifying what practices constitute “good manufacturing” for
drugmakers generally and detailed directives for categories of products.
Since 1991 the agency has made successful completion of a so-called GMP
inspection a condition for approval of any new or supplemental NDA.

**Midcourse corrections.** A decade after the 1962 amendments, Sam
Peltzman argued that the effectiveness standard, coupled with the require-
ment of affirmative FDA approval, cost patients more (by slowing the introduction of new drugs) than it saved them (by preventing other thalidomide incidents).\textsuperscript{16} Peltzman’s thesis has been elaborated by others, while defenders of the proof-of-effectiveness standard have sought to justify both Congress’s central choice and the FDA’s implementation.\textsuperscript{17} This debate provides the background for examining two efforts by Congress to compensate for effects of the new drug approval system.

**Orphan drug amendments.** That the 1962 amendments would increase the cost and time required to introduce new drugs was surely predictable. Drugmakers would, accordingly, focus their research dollars on drugs whose sales could yield a good return on their investment. Drugs for common, particularly chronic, diseases would be popular; drugs for uncommon diseases would attract less attention.

By the 1980s Congress was convinced that it needed to create or restore incentives for the development of treatments for so-called orphan diseases—those illnesses that afflict a limited number of people.\textsuperscript{18} The primary incentives were assurance of a period of market exclusivity—apart from any patent protection—for developers of approved drugs for uncommon diseases and tax credits for certain developmental costs.\textsuperscript{19}

**NDAs and patent term restoration.** Under the scheme Congress established in 1962, every new drug had to have its own approved NDA. After they failed to escape the FDA’s new drug jurisdiction, makers of generic copies of pioneer drugs sought access to the safety and effectiveness data on the basis of which the FDA had approved the pioneer NDAs. From the outset the agency ruled that this data constituted trade secrets exempt from disclosure under the Freedom of Information Act (FOIA) and protected by the FD&C Act itself.\textsuperscript{20} The result of this “second patent” was to prolong the period during which manufacturers of pioneer drugs escaped competition.\textsuperscript{21} Makers of pioneer drugs had a different complaint about the system, however. They cited the erosion of patent protection during the increasingly complex drug development and approval process.\textsuperscript{22}

Ultimately, these divergent views yielded a set of compromise amendments to the FD&C Act. In brief, the makers of generic drugs gained earlier access to the market because, once any patent on the pioneer product had expired, the FDA could approve NDAs for their generic products based on proof that they were functionally equivalent. Makers of pioneer drugs won patent extension under a formula that credits both time spent conducting clinical studies and time awaiting FDA review and approval.\textsuperscript{23}

---

**Lessons From The New-Drug Approval System**

The evolution of the FDA drug approval system contains important
lessons about regulatory structure and bureaucratic behavior.

**Enlargement of FDA jurisdiction.** In the early years of this century the developer of a new drug could set its own schedule for marketing, unimpeded by regulation. If the FDA had reservations about safety or effectiveness, it had the burden not only of initiating a challenge but of proving that the drug was dangerous or mislabeled. Now all new prescription drugs must have FDA approval before they can be marketed. The system also requires agency approval for any significant, and many not so significant, changes in the original product. Moreover, any change that requires approval may become the occasion for FDA review of a manufacturer’s ongoing compliance with other, only incidentally related, requirements. The FDA must also be notified about, and sometimes its approval obtained for, clinical investigations involving experimental drugs and studies of approved drugs for new uses. Accordingly, FDA concurrence is now a prerequisite for all critical steps in the development and marketing of drugs.

This means that manufacturers’ development plans are hostage to circumstances that impede the FDA’s ability to provide the approval they need. These “circumstances” include agency guidelines that require FDA reviewers to reconstruct and reanalyze the clinical data summarized in an NDA. They also include the agency’s budget for medical reviewers, the competence of reviewing staff, the agency’s ability to schedule meetings of advisory committees (whom it consults on most significant approvals), the field staff’s capacity to complete timely GMP inspections, and the competing demands on agency reviewers to fulfill other functions (such as the review of FOIA requests).

**A risk-averse environment?** A common claim is that FDA reviewers are afraid of making decisions that could allow the marketing of another thalidomide. They are said to be haunted by the spectre of “being hauled up” before a congressional oversight committee and pilloried for a mistake that cost lives. There is a germ of truth in this caricature, but the environment in which FDA drug reviewers function is more complex.

It is complex both because drug therapies are complex—their assessment requires application of several different disciplines—and because review occurs years after the key evidence has been assembled and involves, in critical positions, persons who may have had no role in the design or planning of the clinical trials whose results they are assessing. A former director of the FDA’s Center for Drug Evaluation and Research (CDER) once likened the review process to the process for oversight, review, and approval of a doctoral thesis. It is not, he stressed, a process to which the reviewer comes with an inclination to disapprove the work submitted; the reviewer wants the candidate to succeed. But he or she (or they) invariably discover problems with the research already undertaken: Certain questions
have not been explored; key variables have not been controlled for; research notes may be confused.

The resulting exchanges inevitably prove interactive. As two problems are resolved, at least one new issue surfaces. Sometimes experiments have to be repeated. “why didn’t you do it this way?” is a recurrent question, whose answer requires explanation, argument, occasionally even appeal to supervisory judgment. And all of this occurs in a context in which meetings are scheduled with reluctance—because they exhaust the time of all participants and can raise suspicion about the integrity of the process—and also with difficulty. As a consequence, the process typically moves slowly and often with mounting frustration for those who cannot turn to other activities until this one is completed.

This process is not necessarily unsound. Even with procedural reform and added staff, the drug approval process will remain one in which the incentives to achieve prompt resolution are very different. Under the current regulatory system, the side with the weaker interest in resolution is the one that must be satisfied before a product may be distributed or used.

**History Of Medical Device Regulation**

A comparison of the FDA’s regulation of drugs with the evolution of the system for regulating medical devices reveals differences but also clear parallels. And the pressures on and within the FDA to subject new devices to requirements similar to those that it has applied for many years to drugs have never been more intense.

**Prior to the 1938 act.** The 1906 act did not apply to medical devices. Indeed, prior to World War II there was no device industry to speak of. Several companies supplied instruments to physicians and equipment to hospitals, but few thought that these items presented risks other than those associated with untrained use. Outside dentistry, few medical products were intended for implantation in, or even prolonged application to, the human body. However, numerous mechanical or electrical contrivances were promoted for their capacity to arrest disease or improve health. Virtually all were considered frauds by both federal regulators and medical practitioners. Accordingly, by the time Congress turned its attention to amending the 1906 act, there was little objection to extending the act’s prohibitions against adulteration and misbranding to so-called devices.

Initially it was proposed to expand the new act’s definition of drug to include devices, a merger that made no difference at the time. It was not until reports of the Elixir Sulfanilamide disaster in 1937 that Congress decided to mandate premarket review of new drugs, but by that stage in the act’s evolution drugs and devices were defined as distinct categories-a
change made to conform with common sense. No one at the time appears to have asked whether premarket review was appropriate for some devices.

**Regulation under the 1938 act.** The 1938 FD&C Act gave the FDA jurisdiction over medical devices for the first time, but was essentially confined to challenging the sale of products that it believed were adulterated (unsafe) or misbranded (bearing misleading claims of effectiveness). Enforcement was mixed. The agency successfully challenged the label claims made for what officials termed “quack” devices, but it was often frustrated in its efforts to remove from the market altogether products for which agency scientists believed there could be no legitimate medical use.27

The agency paid little attention at first to devices marketed to and used by trained medical practitioners. While a few simple implants, such as surgical nails, had gained acceptance in medical practice, the era of artificial organs, cosmetic implants, and heart pacemakers lay in the future. Furthermore, concerns about medical equipment ranked low in priority within the FDA’s Bureau of Drugs, which until the 1970s had responsibility for administering the device provisions of the law.28 The FDA’s meager regulatory authority over devices was brought to Congress’s attention during its consideration of the 1962 drug amendments; however, to secure prompt passage of drug reform legislation, a proposal to require premarket approval for medical devices was put aside—leaving the agency with only the tools that the original 1938 act provided.29

The 1960s saw rapid innovation in medical technology, and FDA officials became convinced of the need to review some new devices for safety and effectiveness before they were introduced. On several occasions the agency seized upon the act’s expansive “drug” definition to claim that certain diagnostic tools or other items of medical equipment were in fact “new drugs” that required approval before they could be introduced.30 Perhaps surprisingly, this gambit succeeded. In *United States v. Bacto-Unidisk*, the Supreme Court sustained the FDA’s claim that an antibiotic sensitivity disk—which at no time came into contact with a patient’s body—was a “drug,” which meant that it first had to be approved by the FDA.31 A similar agency charge against the maker of a piece of equipment used to tie off blood vessels during surgery was also upheld. The FDA also successfully held that certain contact lenses were “new drugs.” And it convinced the manufacturer of the Copper-7 intrauterine device (IUD) that the product was a new drug because its contraceptive action was the result of the chemical action of the copper filament. In every case the agency’s central argument was that the risks presented by the product were of the sort that led Congress to require premarket approval for new drugs.

In truth, these cases were not systematic initiatives. The FDA recognized that rebaptizing products as “drugs” tortured the language of the act and...
would eventually strain the credulity of judges. Moreover, as the agency's chief counsel explained to Congress, pursuit of this “new-drug” theory had a price.\textsuperscript{32} If all IUDs were declared new drugs, every IUD that had previously been sold was illegal. A nationwide recall of the several hundred thousand IUDs women were using without obvious adverse effects would be disruptive, costly, and risky.

**Enactment of the 1976 amendments.** By 1970 there was broad recognition that the existing law provided an inadequate framework for regulating the rapidly expanding universe of medical devices. Resorting to the new-drug provisions of the law offered no comprehensive solution. Moreover, a consensus was emerging that the new-drug model was not the right instrument for regulating most devices.

The first clear conception of a system for regulating devices was offered in 1970 by an internal Department of Health, Education, and Welfare (HEW) committee chaired by Heart Institute Director Theodore Cooper.\textsuperscript{33} The Cooper committee's conceptual blueprint built on a central principle: No single form of regulation, such as drug-like premarket approval, would fit all medical devices. The committee called instead for a system whose requirements were calibrated to the issues of safety and effectiveness presented by specific devices. This was to be accomplished by assigning all existing devices to one of three classes, ranging from premarket approval to simple policing for faulty manufacture or mislabeling.\textsuperscript{34}

The Cooper committee report inspired FDA officials and congressional staff to begin drafting the legislation that became the 1976 Medical Device Amendments. Expert advisory committees were to play a central role in FDA decision making and were to begin by classifying all devices then on the market. Premarket review of safety and effectiveness would be required for only a minority of devices.\textsuperscript{35}

Anyone who now reads the claims made by agency officials and Congress is likely to be amused by their optimism. They promised regulation that would assure the safety and effectiveness of medical devices and at the same time "foster innovation" of new products.\textsuperscript{36} Few acknowledged the inevitable tension between these goals. The 1976 amendments were promoted as a new type of regulatory statute, one that augmented the FDA's weapons against dangerous and worthless products and subjected high-risk technologies to premarket review, but also authorized less intrusive regulation of the majority of devices. The amendments, as implemented, should be judged against this promise.

**Design of the amendments.** The device amendments increased the length of the FD&C Act by more than one-third and purported to resolve all important questions about the FDA's authority.\textsuperscript{37} The agency was given power to prescribe GMP requirements for devices, to ban worthless or
dangerous products administratively, and to require notification, replacement, and/or refund by makers of defective products.” Device manufacturers were assured a host of procedural safeguards, including rights of judicial review as guarantees against arbitrary action. These changes, however, rank behind four key decisions reflected in the amended act.

(1) Classification: Congress provided that all medical devices should be classified into one of three classes. The law permitted the FDA to “downclassify” a device if it later concluded that the requirements otherwise applicable were too stringent. The FDA did not complete the classification process until 1988, later than predicted, but not a bad performance given the complexity of the task.

(2) Three-tier controls: Devices in Class I would be subject to so-called general controls, essentially requirements to prevent adulteration or misbranding. Devices in Class II would be subject to or at least eligible for categorywide performance standards. Only devices in Class III would be subject to premarket approval. The FDA’s final classification of all devices on the market in 1976 assigned roughly 30 percent to Class I, 60 percent to Class II, and 10 percent to Class III.

(3) Comparable regulation of old and new devices: The FDA’s difficulties in applying the effectiveness requirement to pre-1962 drugs led Congress to adopt the general principle that devices should not be disadvantaged by the inevitably gradual implementation of the new law’s substantive requirements. Post-1976 products similar to devices already on the market could not justifiably be refused market entry until the agency was prepared to set standards or demand testing for the type of device. Accordingly, the amendments provided that a post-1976 device that was “substantially equivalent” to a device already being sold could enter the market immediately and comply on the same time table as the pre-1976 incumbents with any requirements that the FDA later adopted.

(4) Premarket notification: The drafters of the 1976 amendments realized some mechanism was needed to distinguish between devices made after 1976 that were indeed similar to marketed products and genuinely novel devices. The mechanism was a requirement that the FDA be notified, ninety days in advance, of a seller’s intention to market any device that it had not previously marketed. The idea was to give the agency an opportunity to verify the seller’s claim that the new product was substantially equivalent to a device already on the market. This section 510(k) process has become the main route by which new products reach the market?
amendments is at points critical of the agency’s performance, my primary objective is to illustrate how the structure of the law, combined with the agency’s instincts, has produced a system that imposes many of the same burdens on medical innovation as the new-drug model imposes.  

The first managers of the FDA’s new Bureau of Medical Devices believed that regulators, inventors, and manufacturers should work cooperatively. Most did not share the suspicion of manufacturers’ motives held by many FDA field inspectors and some reviewers of drugs. And they took seriously the claims by architects of the 1976 amendments that regulation should not discourage the development of new devices.

**Preamendment Class III devices.** The 1976 amendments contemplated that FDA approval would be required for only a minority of devices made before the amendments were enacted, those assigned to Class III, as well as substantially equivalent postenactment devices. Neither group would be subject to review, however, until the FDA announced a deadline for the submission of premarket approval (PMA) applications. The statute prescribed no timetable for such announcements.

The FDA proceeded slowly to require submission of PMA applications for pre-1976 Class III devices and their equivalents. The agency first had to complete the classification process. It also had to promulgate regulations to implement several new provisions that required regulations to make them operative. Furthermore, the agency was soon faced with PMA applications for new devices that had no pre-1976 equivalents and a much larger number of section 510(k) premarket notifications. Both of the latter were subject to statutory deadlines for review, and the bureau’s management initially proclaimed its determination to comply.

It was therefore easy to neglect the premarket approval requirement for old Class III devices and their equivalents. Many of these products had been around for years; some had undergone clinical testing; and patient usage in most cases did not raise concerns. More than a decade passed without any determined effort by the FDA to “clean up” the backlog of Class III devices for which PMA applications were going to be required—unless some were eligible for down-classification, a process subject to major procedural obstacles. One result of these priorities was to leave the FDA vulnerable to the later discovery of evidence that some “old” Class III devices might in fact not work or, worse, might be dangerous. The recent struggle over silicone breast implants epitomizes this problem. The episode dramatized the FDA’s neglect and raised doubts about its assessments of clinical safety and effectiveness.

**Standards for Class II devices.** Class II was created for devices for which performance “standards” would be sufficient to assure safety and effectiveness. What the drafters visualized as a standard is reasonably
clear: a set of performance specifications embodied in a formal regulation governing a type of device, such as heart monitors. What is not clear is how often Congress expected that the FDA would find it necessary to prescribe standards for devices in Class II—more than 50 percent of all pre-1976 devices.

One view is that this authority would be reserved for those devices about which enough was known both to reveal the risks of individual product variation and to allow establishment of categorical specifications. In this view, it was no shock that the FDA failed for more than a decade to propose the adoption of a standard for any Class II device. Other observers, however, believed that this amounted to failure to implement the new law. FDA officials responded that the statute did not make standards mandatory for Class II devices, claimed that more pressing problems had commanded attention, and called attention to procedural and resource impediments to use of this authority.

My purpose is not to challenge the FDA’s explanation. What is significant about its failure to use its standard-setting authority is that a large number of pre-1976 devices, and their postenactment equivalents, in the view of agency critics, remained unregulated throughout the 1980s.

**Section 510(k) premarket notification.** Congress decided that post-1976 devices should not be subject to heavier burdens than equivalent pre-1976 devices. Section 510(k)’s premarket notification requirement was the mechanism by which it sought to implement this policy of regulatory parity. This requirement applied to every product introduced after the 1976 amendments became effective. Its purpose was to allow the FDA to determine whether a new product was sufficiently similar to a preenactment product that it should be regulated in the same way, on the same schedule. However, Congress provided no definition of “substantial equivalence.” An interpretation limited to products displaying nearly identical features, materials, and performance would have quickly created a backlog of products that required either premarket approval or reclassification.

The Bureau of Medical Devices instead embraced a standard of substantial equivalence that focused less on identity of design and materials and more on comparability of function and clinical performance. To determine whether a device was substantially equivalent, the bureau often asked for reports of clinical experience. Although some manufacturers balked at these requests, most came to appreciate that the agency was more likely to agree that a product was equivalent if it could be shown that it was no more risky and no less effective than similar marketed devices. The number of premarket notifications each year soon dwarfed the number of PMA applications the FDA received. And, of the roughly 5,000 devices entering the market each year, 98 percent elicited a judgment that the product described
was indeed substantially equivalent and eligible for immediate marketing.\(^{59}\)

The section 510(k) route to the market for post-1976 devices thus became more difficult, to the extent that the FDA insisted on evidence from clinical experience and displayed an unwillingness to base equivalence on design and materials alone. But it was also much wider than a formalistic interpretation stressing physical identity would have allowed.

The FDA further broadened the section 510(k) route to market by adopting a policy of what became known as “piggybacking.” It accepted, as “equivalent,” devices that were similar to a product marketed after 1976 that was in turn similar to a product marketed before 1976. The makers of a new device could gain entry to the market, therefore, by pointing to a “predicate product” that itself had been marketed for the first time after 1976 if its lineage could be traced to a preenactment device.\(^{60}\)

The result of these early policies was to make it easier for many post-1976 devices to enter the market, where they remained minimally regulated until the FDA found the time, resources, and resolve to establish a performance standard or call for the submission of PMA applications. This did not mean that no new medical devices were subject to premarket approval. New products for which no equivalent predicate device could be identified represented only a small percentage of all those reported to the FDA, but the number increased as the 1980s passed.\(^{61}\)

**PMA applications.** Although new devices automatically subject to premarket approval represent only a small fraction of new medical products, it is an important fraction for several reasons. One is that until very recently the number of PMA applications has risen faster than the FDA’s capacity to review them. Government budgetary constraints always burden innovation in products that require affirmative agency approval. By definition, moreover, devices that are not eligible for marketing under section 510(k) are more likely to represent innovations. In addition, improvements in products that originally required FDA approval become difficult because any significant change requires approval of a supplemental PMA.

There are strong indications that the FDA’s approval standards for postenactment devices subject to premarket approval increasingly will resemble those for new drugs. The statutory provisions governing premarket approval for devices do resemble those for drugs.\(^{62}\) A PMA application is required to contain full reports of studies to show safety and effectiveness; investigational devices may be distributed only in compliance with an “investigational device exemption,” and advisory committee consultation ordinarily is required before approval.\(^{63}\)

Other provisions, however, reflect Congress’s intention that medical device regulation should not mimic the drug model.\(^{64}\) The most important, clearly, are provisions that confine premarket approval to devices for which
less regulation cannot assure safety and effectiveness. But there is also language in section 515 that describes the kind of evidence required to support approval of a Class III device:

(A) Except as authorized by subparagraph (B), the effectiveness of a device is . . . to be determined, in accordance with regulations promulgated by the Secretary, on the basis of well-controlled investigations, including clinical investigations where appropriate, by experts qualified by training and experience to evaluate the effectiveness of the device, from which investigations it can fairly and responsibly be concluded by qualified experts that the device will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling of the device, (B) If the Secretary determines that there exists valid scientific evidence (other than evidence derived from investigations described in subparagraph [A]) (i) which IS sufficient to determine the effectiveness of a device, and (ii) from which it can fairly and responsibly be concluded by qualified experts that the device will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling of the device, then . . . the Secretary may authorize the effectiveness of the device to be determined on the basis of such evidence.

This language was included to confirm that the agency may appropriately rely on less formal evidence than the FDA has prescribed for new drugs.

Notwithstanding this message, the FDA appears poised to insist on studies for medical devices that are as elaborate as those it requires for new drugs. In April 1992, when the agency was wrestling with the problem of silicone breast implants-most of which had been marketed prior to 1976-FDA Commissioner David Kessler formed the Committee for Clinical Review. The committee was chaired by Robert Temple, then director of the Office of Drug Evaluation I in the FDA’s CDER. Most of the other members of the Temple committee were actively involved in the new-drug evaluation process. The committee examined the reviews performed by device center staff for approximately two dozen “pending and approved applications.” Notably, its review embraced both PMA applications and section 510(k) notification submissions whose burden, under the law, is simply to demonstrate substantial equivalence to some predicate product.

Based on what it conceded was a small sample, the Temple committee identified “certain patterns of deficiencies in the design, conduct, and analysis of clinical studies . . . in enough of the applications to suggest that these deficiencies represent a common problem, one regularly encountered by CDRH [the Center for Devices and Radiological Health] in review [of] PMA’s and 510(k)’s.” The committee’s damning critique found numerous examples of the following deficiencies: (1) failure to utilize the most appropriate kind of control or to identify any control group at all; (2) use of poorly defined historical controls, or control groups for which data on critical clinical details were not available; (3) use of sample sizes inadequate to answer questions; (4) poor specification and characterization of patients
entering studies; (5) poor or no assessment of the comparability of patients in treatment and historical control groups; (6) failure to define study endpoints (success, failure, complications) clearly and consistently; and (7) failure to consider and utilize blinded evaluation of endpoints where endpoints are subjective.

Three features of the Temple committee’s review are particularly noteworthy. One is the fact that it was undertaken at all. Kessler’s decision to authorize the inquiry conveyed a message that the device center’s processes were not reliable and that its personnel were not qualified. Second, the committee itself accepted without question two propositions that, if implemented, will surely slow the approval of new devices. One is that the same qualitative criteria should govern studies to show substantial equivalence under section 510(k) and studies to prove safety and effectiveness under section 515. The other is in the committee’s broader declaration that “the fundamental principles underlying evaluation of any therapeutic intervention, whether it is a drug, device, diet, or surgical procedure, are the same.” Perhaps of greatest significance, Kessler has endorsed the report and promised that its recommendations will be implemented.

**Legislative Adjustment Of Device Regulation**

As in the case of drugs, Congress has on two occasions revisited the system it created for regulating medical devices to modify its original design.

**Corrections of regulatory impact.** The 1984 Drug Price Competition and Patent Term Restoration Act also offered patent term restoration for medical devices subject to regulatory review that significantly curtails their patent life. A handful of premarket approval devices have gained patent extensions under the 1984 act. While the Orphan Drug Act did not apply to devices, Congress has since added medical devices to the statutory provision that allows federal financial assistance for the development of devices for rare conditions.

**Strengthening device regulation.** Prompted by congressional dissatisfaction with implementation of the 1976 device amendments, the 1990 Safe Medical Devices Act strengthened the FDA’s hand in controlling entry of new products and in monitoring use of marketed products.

The 1990 amendments relieved the FDA of any obligation to promulgate performance standards for all Class II devices. New authorities to require postmarketing surveillance of device performance and directives to establish so-called tracking requirements for certain high-risk devices were added. The latter were justified, in part, as allowing the FDA to relax its requirements for initial approval because it could exercise tighter control over the use of devices once they were introduced. There is little evidence
that agency reviewers are being more flexible in approving initial marketing. Indeed, serious effort to implement the recommendations of the Temple committee will offset any gains in the pace of review.

A third set of changes surround the section 510(k) premarket notification process. Congress was persuaded that the time originally specified for agency response—ninety days—was too short for many devices, given their complexity, so it doubled the time limit. It also required the FDA to affirm that a device is substantially equivalent before it may be introduced—whether or not the statutory time limit has expired. At the same time, the 1990 act confirmed the FDA’s authority under section 510(k) to consider as “substantially equivalent” devices that differed in design or technology from marketed products if they are shown to be as safe and effective.

Transformation Of Regulation Of New Medical Devices

What lessons can be drawn from this comparative account of the evolution of federal regulation of medical devices? And what predictions can one make about the future effect of regulation on medical device development?

The recent evidence indicates that the rate of introduction of new medical devices has slowed over the past three years. New approvals of PMA applications declined from forty-six in 1990 to twelve in 1992. The number of 510(k) premarket notifications submitted to the agency rose from 5,063 in 1986 to 6,341 in 1992; during the same period the number approved declined, from 5,359 to 4,844, and the time required for FDA review increased substantially.

Last year the FDA received approval for a significant enlargement of the device center’s product review staff, to help cope with its increasing workload. Other things being equal, this strengthening of the agency’s internal core of scientists should reduce the review time for new medical devices, but other things may not remain equal. There is every indication that the center will become more demanding in its review of clinical trials for new devices. At a September 1993 workshop an agency spokesman said that of more than 600 PMA applications approved since 1976, only a few contained clinical data of the quality the agency now expects.

How the device center will review 510(k) notifications in the future is also difficult to predict. As amended in 1990, section 510(k) allows the FDA to require the submission of “clinical data” for devices that employ a different technology from the claimed predicate product. This change was billed as confirming the FDA’s past practice, which had resulted in requests for clinical data for just over 5 percent of all 510(k) notifications. Since 1990, however, that rate has tripled to 15 percent.

Another key question before the FDA—and developers of new
devices—whether the same kinds of studies may be required in 510(k) notifications as are required for premarket approval of new Class III devices. At the September 1993 workshop, agency officials suggested that “trials” may not be required, but that the “clinical data” demanded should display comparable sophistication and rigor in design—a statement whose ambiguity is hardly reassuring.

The unmistakable signs are that the medical device review process will increasingly resemble the process for approving new drugs. One reason is that the FDA has embraced advances in clinical trial design and statistical analysis. The same phenomenon is evident in the drug approval process, notwithstanding pressures to relax approval requirements for drugs to treat AIDS and cancer. The political climate in which the FDA operates also has driven agency reviewers toward greater caution. As has been true for a generation, members of Congress who take an interest in the FDA are more likely to criticize decisions to approve products that turn out to display unexpected hazards than to challenge the agency for being cautious or slow. However, it is not only advances in clinical trial design and persistent congressional suspicion that are transforming the FDA’s approach to new medical devices. These forces are influential precisely because the law under which the FDA functions is structured to reward caution and facilitate delay. The feature that exemplifies this bias is the requirement that all new drugs—and now most devices—must have FDA approval before they can be marketed. Such a requirement may be justified in an era when many drugs and devices are complex, powerful, and potentially very profitable, but we must understand what consequences flow from conferring this sort of authority to a government agency.

It means, first, that the makers (and other potential beneficiaries) of new products bear most of the cost of budgetary constraints on the agency. If a manufacturer requires FDA approval before it can commercialize a product, the manufacturer will be disadvantaged by any circumstance that makes gaining such approval more difficult—for example, if authority to grant approval is centralized in a few persons, or if the number of applications for approval increases, or if agency personnel lack the necessary expertise.

It also means that the agency responsible for deciding whether the standard for approval has been met also determines what that standard is. While the law ostensibly specifies what the manufacturer of a new product must show, its inevitably general language does significantly constrain the FDA’s appetite for elegant and costly information.

Third, a system that forbids marketing until approval is granted provides a potent means of securing compliance with collateral legal requirements. Some may be central to the FDA’s approval decision, such as proof that a
formulation is capable of delivering a drug that has been proved effective in another form. But the number of merely useful requirements that can be linked to product approval—and enforced by making compliance a prerequisite—is very large.

Concluding Comments

One cannot appreciate contemporary regulation of medical devices without understanding how prescription drugs are regulated. While Congress in 1976 sought to draw a sharp distinction between the framework that it (and the FDA) had created for controlling prescription drugs and the new system it was creating for devices, it did not entirely break the old mold. And in the hands of an agency that has had more than thirty years’ experience administering a “cradle-to-grave” system of control over prescription medicines, it may have been inevitable that administration of the device law should increasingly resemble the “drug model.”

Some may argue that this evolution was not only predictable but fully justified. It comes with a price, however, in both cost and delay in the availability of new devices. Those who consider these excessive will be tempted to search for improvements in the FDA’s system of review and approval. There is now much interest in the prospect that merely increasing the FDA’s resources—chiefly, perhaps exclusively, through fees charged to makers of products subject to regulation—will reduce the time required to gain agency approval. This promise of user fees is not yet proved. But even if additional resources are shown to enable the FDA to expedite its performance of functions that now take so much time, they do not address—a more fundamental question: Should government approval be a precondition for as many steps in product development, marketing, and improvement as it now is? Without some effort to address the appropriate scope of the FDA’s responsibilities, real reform of medical device (and drug) regulation will not be possible.

The author thanks M. Elizabeth Magill, University of Virginia Law School Class of 1995, for her assistance in preparing this paper, and Peter Barton Hutt for his valuable comments.

NOTES

1. 34 Stat. 770 (1906).
2. 221 U.S. 488 (1911).

5. FD&C Act, Sections 505(a) and 505(c), 52 Stat. 1052 (1938).


10. FD&C Act, Section 505(i).

11. 21 C.F.R., Part 312.

12. 21 C.F.R., Section 314.126; and Federal Register 44 (6 April 1979): 20796.


15. See Federal Register 53 (25 May 1988): 18905 (FDA refusal to approve NDAs for failure to comply with good manufacturing practice, or GMP, regulations); Federal Register 52 (10 March 1987): 7318; and Federal Register 52 (6 August 1987): 29724, affirmed Copanos and Sons, Inc. v. FDA, 854 F.2d 510 (D.C.Cir. 1988) (FDA withdrawal of NDAs for failure to comply with GMP).


32. Regulation of Medical Devices (Intrauterine Contraceptive Devices), Hearings before the House Government Operations Subcommittee on Intergovernmental Relations, 93d Cong., 1st Sess. (1973), as quoted in Hutt and Merrill, Food and Drug Law, 734-735.
36. Cooper, “Device Legislation,” 172. There Cooper stated that “the purpose of the legislation should be not only to avoid hazards, but also to promote needed device development.”
37. The Medical Device Amendments of 1976 ran forty-five pages long. 90 Stat. 539 (1976). By comparison, the 1938 act was nineteen pages long.
38. FD&C Act, Sections 520(f)(1)(A), 516, and 518.
39. Ibid., Section 517.
40. Ibid., Sections 513(e), 513(f)(2)(A), and 520(1)(2).
42. Hutt and Merrill, Food and Drug Law, 750.
43. FD&C Act, Sections 513(f)(1)(A) and 515(b)(1).
44. Ibid., Section 510(k).
45. Hutt and Merrill, Food and Drug Law, 755. A 1991 Advisory Committee on the FDA reported the following breakdown of the 97,000 items requiring “major product evaluation” by the Center for Devices and Radiological Health from the enactment of the 1976 device amendments through fiscal year 1990: 61,970 premarket notifications, 27,105 investigational device exemption supplements, 5,304 premarket approval supplements, 2,345 investigational device exemptions, and 832 premarket approval applications. Final Report of the Advisory Committee on the FDA (May 1991), C2-C3. In 1990 a House committee estimated that 98 percent of the devices entering the market each year do so on the basis of a claim that they are substantially equivalent to an earlier device. Safe Medical Devices Act of 1990, H.R. 101-808, 101st Cong., 2d Sess. (1990), 14.
46. For a more comprehensive, but hardly unbiased, assessment, see Less than the Sum of Its Parts: Reforms Needed in the Organization, Management, and Resources of the FDA’s Center for Devices and Radiological Health, House Energy and Commerce Subcommittee on Oversight and Investigations, 103d Cong., 1st Sess. (May 1993).
47. By contrast, the 1962 Drug Amendments exempted preenactment new drugs from the effectiveness standard for only two years and implied that the FDA by that deadline had evaluated evidence of effectiveness for all such products. Drug Amendments of 1962, P.L. 87-781, Section 107(c).
49. FD&C Act, Section 513(f)(2)(A). See also Hun and Merrill, Food and Drug Law, 770-772.
51. FD&C Act, Section 513(a)(1)(B).
52. Ibid., Section 514.
53. Hutt and Merrill, Food and Drug Law, 750.
57. 21 C.F.R., Section 807.81 and following.
58. Hutt and Merrill, Food and Drug law, 755.
61. As of 1991 the FDA had received 832 premarket approval applications and more than 5,000 premarket approval supplements. Final Report of the Advisory Committee on the FDA.
62. FD&C Act, Section 515.
63. Ibid., Section 520(g); and 21 C.F.R., Part 812.
64. Ibid., Section 513.
65. Ibid., Section 515.
67. The committee’s report does not acknowledge the different legal standards applicable to the two sorts of decisions. Final Report of the FDA Committee for Clinical Review (March 1993).
70. 104 Stat. 4511.
73. Ibid.