Public Health Effectiveness of the FDA 510(k) Clearance Process: Measuring Postmarket Performance and Other Select Topics: Workshop Report

ISBN
978-0-309-16206-7

132 pages
6 x 9
PAPERBACK (2011)

Theresa Wizemann, Editor; Committee on the Public Health Effectiveness of the FDA 510(k) Clearance Process; Institute of Medicine

PUBLIC HEALTH EFFECTIVENESS OF THE FDA 510(k) CLEARANCE PROCESS

Measuring Postmarket Performance and Other Select Topics

Workshop Report

Theresa Wizemann, *Editor*

Committee on the Public Health Effectiveness of the FDA 510(k) Clearance Process

Board on Population Health and Public Health Practice

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This project was supported by a contract between the National Academy of Sciences and the US Department of Health and Human Services (HHSF223200810020I, TO 13). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the agency that provided support for this project.

International Standard Book Number-13: 978-0-309-16206-7
International Standard Book Number-10: 0-309-16206-8

Additional copies of this report are available from the National Academies Press, 500 Fifth Street NW, Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2011. *Public Health Effectiveness of the FDA 510(k) Clearance Process: Measuring Postmarket Performance and Other Select Topics: Workshop Report.* Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON THE PUBLIC HEALTH EFFECTIVENESS OF THE FDA 510(k) CLEARANCE PROCESS[1]

DAVID CHALLONER (*Chair*), Vice President for Health Affairs, Emeritus, University of Florida, Gainesville, FL
GARY S. DORFMAN, Professor and Vice Chair for Research, Weill Cornell Medical College, New York, NY
BARBARA EVANS, Associate Professor; Codirector, Health Law & Policy Institute; Director, Center on Biotechnology and Law, University of Houston Law Center, Houston, TX
LAZAR GREENFIELD, Professor of Surgery and Chair Emeritus, University of Michigan, Ann Arbor, MI
STEVEN GUTMAN, Associate Director, BlueCross BlueShield Assocation, Alexandria, VA
YUSUF M. KHAN, Assistant Professor, University of Connecticut Health Center, Farmington, CT
DAVID KORN, Vice Provost for Research, Harvard University, Boston, MA
ELIZABETH W. PAXTON, Director of Surgical Outcomes and Analysis, Kaiser Permanente, San Diego, CA
SHARI LAWRENCE PFLEEGER, Director of Research, Institute for Information Infrastructure Protection, Washington, DC
WILLIAM W. VODRA, Senior Counsel (*Retired*), Arnold & Porter, LLP, Washington, DC
BRIAN WOLFMAN, Visiting Professor of Law, Georgetown University Law Center, Washington, DC
KATHRYN C. ZOON, Scientific Director and Director of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

Study Staff

ABIGAIL E. MITCHELL, Study Director
HEATHER M. COLVIN, Program Officer
KATHLEEN M. McGRAW, Senior Program Assistant
NORMAN GROSSBLATT, Senior Editor
ROSE MARIE MARTINEZ, Director, Board on Population Health and Public Health Practice

[1]The role of the Committee on Public Health Effectiveness of the FDA 510(k) Clearance Process was limited to planning an information-gathering public workshop. This workshop summary has been prepared by the workshop editor as a factual summary of the presentations and discussions that took place at the workshop. Statements, recommendations, and opinions expressed are those of individual presenters and participants, are not necessarily endorsed or verified by the committee, and should not be construed as reflecting any group consensus.

Reviewers

This report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards of objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the process. We thank the following for their review of this report:

Richard DeRisio, Abbott Medical Optics
Larry Kessler, School of Public Health, University of Washington
John S. Rumsfeld, Veterans Health Administration
Terrence J. Sweeney, Philips Healthcare
Susan F. Wood, School of Public Health and Health Services, The George Washington University

Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the report before its release. The review of the report was overseen by **Kristine M. Gebbie,** Acting Joan Grabe Dean of the School of Nursing, Hunter College at the City University of New York. Appointed by the Institute of Medicine, she was responsible for making certain that an independent examination of the report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of the report rests with the authors and the institution.

Contents

TABLES, FIGURES, AND BOXES

Tables

Figures

Boxes

Abbreviations

AED	automatic external defibrillator
CART	Clinical Assessment, Reporting, and Tracking program
CDRH	Center for Devices and Radiological Health (FDA)
CE Mark	European conformity mark
CHMP	Committee for Human Medicinal Products
CMS	Centers for Medicare and Medicaid Services
EMA	European Medicines Agency
FDA	US Food and Drug Administration
FFDCA	Federal Food, Drug, and Cosmetic Act
GHTF	Global Harmonization Task Force
GRAE	generally recognized as effective
GRAS	generally recognized as safe
ICD	implantable cardioverter–defibrillator
IDE	investigational device exemption
ISO	International Organization for Standardization
MAUDE	Manufacturer and User Facility Device Experience Database
MDA	Medical Device Amendments of 1976
MDR	medical-device reporting

NCDR	National Cardiovascular Data Registry
NSE	not substantially equivalent
PMA	premarket approval
QSR	quality system review
SSED	Summary of Safety and Effectiveness Data
UDI	unique device identification
VA	US Department of Veterans Affairs
VHA	Veterans Health Administration

1

Introduction

At the request of the Food and Drug Administration (FDA), the Institute of Medicine (IOM) has convened a consensus committee to review the 510(k) clearance process for medical devices, also known as premarket notification (see Box 1-1). Section 510(k) of the Federal Food, Drug, and Cosmetic Act requires a manufacturer of medical devices to notify FDA of its intent to market a medical device at least 90 days in advance. That window of time allows FDA to evaluate whether the device is substantially equivalent to a product already legally on the market (called a predicate), in which case the device does not need to go through the premarket approval (PMA) process. A predicate can be a device that has been cleared through the 510(k) process, a device that was legally marketed before May 28, 1976 (a preamendment device), a device that was originally on the US market as a Class III device (PMA) and later downclassified to Class II or I, or a 510(k)-exempt device (see Box 1-2). A device is considered substantially equivalent to a predicate if it has the same intended use as the predicate device and has either the same technologic characteristics as the predicate device or has different technologic characteristics but does not raise new questions of safety and effectiveness and is as safe and effective as the predicate (FDA, 2000).

As part of its fact-finding process, the IOM Committee on the Public Health Effectiveness of the FDA's 510(k) Clearance Process planned two public workshops to gather information relevant to the statement of task. The committee's statement of task is focused on the 510(k) clearance process. However, it is not possible to review the 510(k) process thoroughly in isolation from other components of medical-device regulation and oversight. Therefore, although some of the topics included in the workshops were not

BOX 1-1
IOM Study on the Public Health Effectiveness of the FDA 510(k) Clearance Process

The IOM Committee on the Public Health Effectiveness of the FDA 510(k) Clearance Process will assess whether the 510(k) clearance process sufficiently protects patients and promotes public health. Specifically, the IOM committee will answer two principal questions:

- Does the current 510(k) process optimally protect patients and promote innovation in support of public health?
- If not, what legislative, regulatory, or administrative changes are recommended to optimally achieve the goals of the 510(k) process?

A final consensus report is expected to be released in the middle of 2011.

BOX 1-2
Definitions of Medical Device Classes

- **Class I** devices are subject to the least regulatory control. They present minimal potential for harm to the user and are often simpler in design than Class II or Class III devices.
- **Class II** devices are those for which general controls alone are insufficient to ensure safety and effectiveness and for which existing methods are available to provide such assurances.
- **Class III** devices are those for which insufficient information exists to ensure safety and effectiveness solely through general or special controls. Class III devices are usually those which support or sustain human life, are of substantial importance in preventing impairment of human health, or present a potential unreasonable risk of illness or injury.

SOURCE: FDA, 2000.

explicitly part of the 510(k) process, they are related to FDA's ability to identify safety concerns about medical devices throughout their life cycle.

The first workshop was held on June 14–15, 2010, in Washington, DC. At that workshop, information was presented to the committee regarding the legislative history of the Medical Device Amendments of 1976, which

instituted the 510(k) process; the regulation of medical devices by FDA, including the premarket notification process and FDA's compliance infrastructure; the structure of the medical-device industry innovation ecosystem and the effects of the regulatory framework on device innovation; and the global regulatory environment for medical devices, including efforts toward global harmonization (IOM, 2010). In addition, the committee heard brief statements from stakeholders on issues relevant to the committee's task during a public comment period.

The second workshop, summarized in this report, was held on July 28, 2010, in Washington, DC. Its primary focus was on monitoring the safety of marketed medical devices, including FDA's postmarket surveillance activities, analysis of safety concerns that resulted in medical device recalls, and non-FDA sources of adverse-event information. The committee also heard a presentation on the issues associated with the use of computer software in medical devices and additional perspectives on device approval and clearance processes.

This report summarizes the views expressed by workshop participants. Although the committee is responsible for the overall quality and accuracy of the report as a record of what took place at the workshop, the views contained in the report are not necessarily those of the committee.

David Challoner, chair of the IOM committee, reminded participants that the committee is in the process of assembling materials that it will examine in the course of developing its findings, conclusions, and recommendations. The committee has drawn no conclusions thus far, and comments made by participants during the course of the workshop should not be interpreted as positions of the committee or of IOM. In addition, probing questions asked by committee members during IOM information-gathering sessions are not indicative of their personal views.

ORGANIZATION OF THE REPORT

The following chapters summarize the presentations and discussions at the second workshop. Chapter 2 reviews FDA's postmarket surveillance activities, including the agency's current system and future plans for monitoring the safety of marketed devices, and product recall studies including a study commissioned for the committee. A variety of non-FDA efforts to monitor adverse events are associated with medical devices, and several such surveillance programs are discussed in Chapter 3. Chapter 4 summarizes an expert panel's discussion of postmarket-surveillance issues. Chapter 5 includes three presentations on other topics of interest to the committee. The first is a presentation of a commissioned paper on the trustworthiness of software in devices; the paper itself is included in Appendix D. The second presentation is a review of quality concerns about the clinical data

used in the PMA process. The last presentation is an example of industry concerns about transparency of, and delays in, FDA decision-making within the 510(k) process.

The workshop agenda, biographic sketches of the speakers and panelists, and the two commissioned papers presented at the workshop are available as appendixes.

REFERENCES

FDA (Food and Drug Administration). 2000. *Device Advice: Premarket Notification. Washington, DC, Food and Drug Administration*. http://www.fda.gov/medicaldevices/deviceregulationandguidance/howtomarketyourdevice/premarketsubmissions/premarketnotification510k/default.htm (accessed August 31, 2010).

IOM (Institute of Medicine). 2010. *Public Health Effectiveness of the FDA 510(k) Clearance Process: Balancing Patient Safety and Innovation*. Edited by Wizemann, T. Washington, DC: The National Academies Press.

2

Food and Drug Administration Postmarket Surveillance Activities and Recall Studies of Medical Devices

This chapter reviews the Food and Drug Administration (FDA) postmarket surveillance activities, including the agency's current system and future plans for monitoring the safety of marketed devices. FDA's surveillance activities are focused on identifying potential safety issues with devices currently on the market. If safety concerns are identified through surveillance activities, FDA can take several different types of actions, including recalling devices. The workshop included presentations on two studies that analyzed data on product recalls. The first of these studies was commissioned by the committee (see Appendix C for the commissioned paper describing the study). The second presentation is a summary of a separate study of recall data.

MONITORING DEVICE SAFETY: THE CENTER FOR DEVICES AND RADIOLOGICAL HEALTH'S CURRENT SYSTEM AND VISION FOR THE FUTURE

In addition to getting safe and effective products to market as quickly as possible, FDA must ensure that devices currently on the market remain safe and effective. Susan Gardner, director of the Office of Surveillance and Biometrics of the Center for Devices and Radiological health (CDRH), described the CDRH postmarket program as consisting of postmarket problem identification, postmarket problem assessment, and public-health response. Particularly in the 510(k) program, the diversity of products demands a diverse surveillance strategy, she said. FDA relies heavily on a pas-

sive surveillance system, and, Gardner noted, some initiatives for improving surveillance efforts are under way.

FDA surveillance systems include mandatory reporting though the Medical Device Reporting (MDR) system, voluntary reporting (primarily by health-care professionals or consumers) through the MedWatch system, hospital-based reporting through the Medical Product Safety Network (MedSun), and an international vigilance program in which reports are exchanged with global regulatory authorities.

Voluntary reporting was initiated in 1973 and now accounts for about 3% of the adverse-event reports that FDA receives, Gardner said. Mandatory reporting was initiated in 1984 for manufacturers and importers (accounting for 93% and 1% of reports, respectively) and in 1990, under the Safe Medical Device Act, for user facilities, including hospitals, nursing homes, surgical ambulatory centers, and so on (accounting for 3% of the reports). All together, FDA receives about 200,000 case reports a year and has a database of about 2.5 million reports.

Mandatory Reporting

For all device classes, FDA regulations require manufacturers to report deaths, serious injuries, and malfunctions to FDA within 30 working days of their becoming aware that a device may have caused or contributed to those events. User facilities are required to report deaths to FDA within 10 working days of recognition of an event and deaths and serious injuries to the manufacturer within 10 working days.

In addition to individual reports, FDA initiated in the late 1990s a program called summary reporting, which provides an abbreviated method for reporting device adverse events. The program relies on established codes (rather than text) for device events that are well known and allows the agency to assess the data for trends. Summary-reporting exemptions are granted only for a specific well-known product and a specific well-known adverse event. Whenever there is an incident related to a product that is outside those boundaries, the manufacturer must file a full individual report.

MedSun is a national network of 350 user facilities. Each facility has two liaisons to the program—an engineer and a risk manager—who are trained to recognize and report adverse events. The system uses electronic reporting to reduce the burden on staff. The emphasis of the program is on device use issues. In addition to the mandatory reporting requirements, FDA encourages user facilities to voluntarily report near-misses and close calls, which now account for bulk of the reports. The program has given the agency an additional connection to the clinical community beyond the reporting relationship, Gardner noted.

About 14 people are dedicated to reviewing postmarket surveillance

reports and another 14 or so work for the MedSun program, Gardner said, and 15–18 epidemiologists are closely linked to the MDR staff.

Identifying Signals Among the Array of Device Hazards

When evaluating adverse event reports, FDA looks for a broad array of device hazards (see Box 2-1). The challenge is to identify the issues, or "signals," that are important among the many reports. A signal is defined as information about a product that FDA regulates that suggests an unexpected risk to patients or users. About 2 years ago, Gardner said, the agency embarked on a "signal escalation" program designed to organize the signals that arise and make them more visible to others in the regulatory centers within FDA. Briefly, as reports come, they are reviewed by analysts who sort them according to an established triage system called Code Blue. Some reports are immediately pulled out and sent to the branch chiefs to review

BOX 2-1
Types of Device Hazards

- **Device failure** (for example, sutureless anastomosis device for coronary arterial bypass graft procedure comes apart and leads to cardiogenic shock)
- **Device malfunction** (for example, ventilator power is lost because of limited power supply capacity in power surge)
- **Use error** (for example, tissue is retained in arthroscopic shaver handpieces because of human factors or design issues)
- **Interactions** (for example, glucose is falsely indicated as increased because of use of glucose test strips and immunoglobulins)
- **Mismatch of parts** (for example, an electrosurgical pad–generator mismatch results in burns)
- **Environmental effects** (for example, heart rhythms analyzed improperly by automated external defibrillators because of too high humidity)
- **Allergic reactions** (for example, catheters are impregnated with chlorhexidine)
- **Toxic events** (for example, cornea is damaged by heavy metals after sterilization of ophthalmic instruments)
- **Software error** (for example, infusion pump interprets a single keystroke as multiple)
- **Packaging defects**
- **Poor maintenance**

(for example, reports of pediatric death, explosions, and burns). Additional information is obtained as needed from the manufacturer or user facility. Reports of interest are further assessed by a specialized group, and a signal is entered into the central tracking system, which is accessible to all staff. A reviewer can, for example, go into the system and see whether signals have been reported on a product being reviewed. Occasionally, reports may also be sent directly to the Office of Compliance if there is a potential compliance issue. Signals come from all offices in CDRH.

Enhancing Surveillance

In addition to the passive-reporting surveillance systems, FDA is implementing enhanced surveillance capabilities. Using MedSun, for example, the agency is conducting targeted surveillance through surveys; 10 surveys and special studies are going on now. Through the MedSun regional-representative pilot program, FDA representatives are visiting 15 hospitals and working directly with their staff to improve reporting. The agency has also formed networks within MedSun to collect real-time data in targeted areas: LabNet fosters reporting from laboratories, HomeNet from home-care agencies, and KidNet from pediatric and neonatal intensive-care units.

Another example that Gardner cited is the Clinical Assessment Reporting and Tracking program for catheterization laboratories (CART-CL), in collaboration with the Veterans Health Administration.[1] With 76 cardiac-catheter laboratories nationwide involved, the program fosters enhanced reporting of unexpected problems with devices by clinicians at the point of care.

FDA is also working on data mining of its vast database of event reports, looking for device–event associations. Gardner noted that FDA's current Manufacturer and User Facility Device Experience (MAUDE) database, built in the early 1990s, is being revamped.

The agency is building its electronic infrastructure and working toward electronic medical device reporting (eMDR). In 2009, nearly 70,000 reports were submitted electronically via an agency-wide portal. Gardner noted that regulation for eMDR is under review.

A particularly important initiative, Gardner said, is the development of a unique device identification (UDI) system. Most devices are bar-coded, but the bar code is "static," including only such information as the manufacturer, make, and model. There is no dynamic way to identify individual devices uniquely, for example, by serial number or by lot number and expiration date. Accurate identification is challenging, Gardner noted, given the variety of devices that FDA regulates, including software and implantable devices.

[1]The Department of Veterans Affairs CART system is discussed further in Chapter 3.

FDA is also conducting discretionary observational studies (for example, in collaboration with such registries as the American College of Cardiology's [ACC's] National Cardiovascular Data Registry [NCDR])[2] and claims-based studies.

Postapproval Studies

Postapproval studies may be ordered as a condition of approval for the highest-risk premarket approval (PMA) products. The studies are used to address important, but not essential, questions of device safety or effectiveness. All studies that are ordered now for PMAs are hypothesis-driven and have deadlines and deliverables, Gardner said. All postmarket studies are listed on a public database with their status. There are also postapproval studies conducted on devices cleared through the 510(k) clearance process.

Postmarket Surveillance Studies

Section 522 of the Federal Food, Drug, and Cosmetic Act (FFDCA) gives FDA the authority to require a manufacturer to conduct postmarket surveillance studies for class II and class III devices if failure of the device is reasonably likely to have serious adverse health consequences or is expected to have substantial use in pediatric populations, is implanted for longer than 1 year, or has life-supporting or life-sustaining use outside the device user facility.

In accordance with Section 522, FDA asks a company questions about a device, the company returns to FDA with a protocol for answering the questions, and then FDA goes through a process of approving or not approving the protocol. A manufacturer can take a number of surveillance approaches, for instance, if the program protocol does not necessarily require that it conduct a clinical trial. Approaches could include, for example, literature review, nonclinical testing, use of secondary data sources, followup with patients, clinical registries, and observational studies. Gardner noted that FDA can order a postmarket surveillance study for a predicate device if necessary.

There are drawbacks to these "Section 522 studies," Gardner said. Surveillance can be required for only 3 years. Because discussion, development, and review processes take time, there are no immediate answers to questions about a device. Thus, if there is a safety issue, the answer is not to order a Section 522 study; other FDA tools are more appropriate.

In 2007, as a result of an Institute of Medicine study of postmarket surveillance for pediatrics (IOM, 2005), the FFDCA was amended to allow FDA to order Section 522 studies as a condition of approval or clearance

[2]The NCDR is discussed further in Chapter 3.

and for longer than 36 months for devices that are expected to have substantial use in pediatric populations. That was a significant change.

On the Horizon

Gardner highlighted a number of forthcoming initiatives that she said will be important for postmarket surveillance. The Sentinel Initiative is an effort to develop a national, integrated infrastructure of electronic healthcare data systems for medical-product safety surveillance. It will augment, not replace, existing functionality, Gardner noted. The focus is on active surveillance. In the proposed model, the data sources would remain at their remote locations, maintained by their local owners, and FDA would be able to send queries to the data owners about specific safety questions (such as rates of implant revision or reintervention, rates of infection, and selected outcomes, such as myocardial infarction, stroke, and death). The system is not designed, however, to address much of what has been discussed regarding 510(k) products, such as out-of-box failures, software glitches, manufacturing defects, and packaging or labeling errors.

The agency is also forming new critical collaborations and leveraging established partnerships with agencies, academic institutions, and professional societies. Gardner cited one program, MDEpiNet, in which FDA is collaborating with academic centers to advance epidemiologic methods and training related to devices.

FDA is working with multiple stakeholders on the development of registries. As noted earlier, registries are one approach that can be used for Section 522 surveillance. They can be used for active surveillance, for short-term as well as longitudinal data. Other potential capabilities include linkage studies with Medicare claims data and mapping of registry data to electronic health records.

Another effort under way is an evidence synthesis project. The agency acknowledges that there are isolated groups or silos of postmarket data (such as MDRs, observational studies, and published clinical studies), and this project is addressing how to combine the multiple data sources and developing prognostic models of long-term device performance.

CDRH is also focused on better integration of premarket and postmarket data on device performance, Gardner said. She described several initiatives, including the Collaborative Review Program, in which center staff spend half their time in premarket review and half their time looking at postmarket adverse-event reports. The program is administratively challenging, Gardner noted, inasmuch as staff are reporting to two different areas, but there has been some success in taking postmarket information back to the premarket function. CDRH also has networks that integrate people across the center in specific fields (such as an orthopedic network and a

cardiovascular network), and it has a embarked on an improved knowledge-management program, using collaborative software and wiki products, to pull together all the information about a given product and make it easily and readily available to all center staff.

Gardner noted that resources, both funding and staffing, continue to be barriers to implementing some of those initiatives.

PREMARKET NOTIFICATION: ANALYSIS OF FOOD AND DRUG ADMINISTRATION RECALL DATA

William H. Maisel, former director of the Medical Device Safety Institute,[3] provided an overview of an independent analysis of FDA recall data that he was commissioned to perform for the committee.[4] The ideal measure of the success of the device approval or clearance processes, he said, would be the performance and reliability of each individual approved or cleared device. Analysis would need to be done for many thousands of devices, so it would be extremely difficult and impractical. Therefore, surrogates of device performance are used, including recalls and medical-device reports. FDA defines a recall as an action taken to address a problem with a medical device that violates FDA law. Recalls occur when a medical device is defective, when it poses a risk to health, or when it is defective and poses a risk to health.

The purposes of the study by Maisel were to analyze the available FDA recall data affecting 510(k) products, to provide an estimate of 510(k) product recall rates, to describe the causes of recalls affecting the products, and to identify risk factors for product recalls.

Data Analysis

Data for analysis were obtained from the FDA 510(k) database, specifically the 48,402 total 510(k) applications that were cleared by FDA from 1996 through 2009, and from the FDA recall database from 2003 to 2009.

Clearance Data

Maisel used FDA advisory-committee assignments during 1996–2009 to provide some idea of the types of products that were being cleared via the 510(k) process (Figure 2-1). Advisory committees that received the greatest number of applications were those for general and plastic surgery, orthopedic, general hospital, and cardiovascular products.

[3]Dr. Maisel is no longer with the Medical Device Safety Institute as of August 2010. He is currently employed at FDA.

[4]The complete commissioned paper is available as Appendix C.

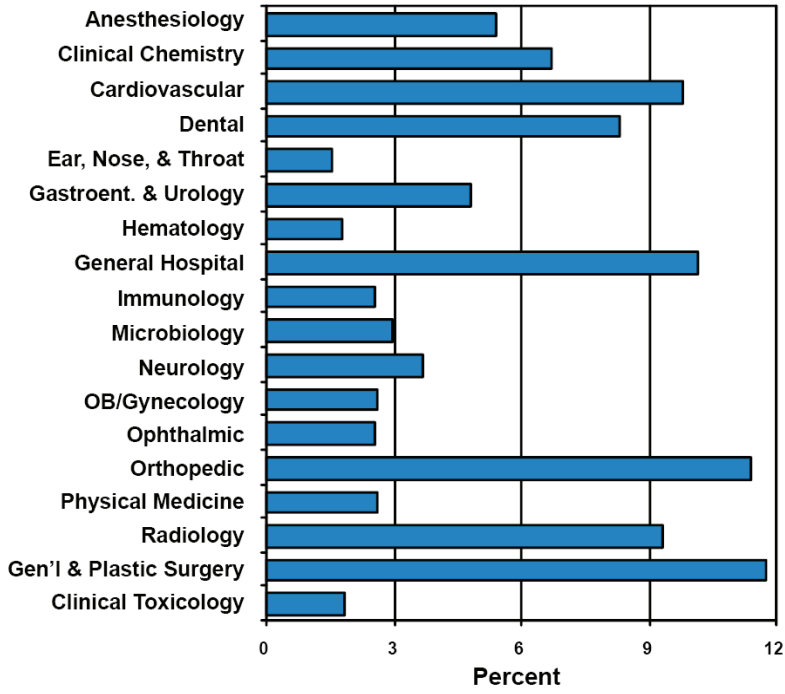

FIGURE 2-1 Breakdown of advisory-committee assignments for 510(k) submissions in 1996–2009.

More than 80% of the devices cleared via the 510(k) program were class II devices, about 10% class I, and just over 1% class III.

The vast majority, about 80%, of the applications in the database are traditional 510(k) applications. Special 510(k) applications, which are for modifications of products that conform to design control standards, make up about 16%, and abbreviated applications, which are for devices that can be cleared on the basis of standards or special controls, make up about 3%.

About 25% of the devices cleared during that period were implantable devices. Over 96% of these 510(k) implantable devices were not life-sustaining.

Recall Rates

Some recalls, Maisel said, can involve more than one 510(k)-cleared product. For the study, recall data were analyzed on the basis of individual 510(k) applications; that is, data were expressed as unique 510(k) applications that were subject to recall.

From 2003 to 2009, 400–500 unique 510(k) applications were affected by recall each year. Maisel reminded participants that several thousand devices are cleared through the 510(k) process each year. About 74% of individual recalled 510(k) devices were recalled a single time, and about 16% were twice. Overall, about 26% were recalled more than once; some were recalled more than 10 times. Analysis of the data is complicated, Maisel noted, in that companies sometimes expand a recall because new information becomes available. Therefore, a large number of recalls of some products should not be interpreted as multiple product failures in the absence of further in-depth analysis. About half the devices recalled in 2003–2009 had been cleared in 1996–2002.

Determining recall rates presents a variety of challenges, Maisel said. For example, how should one account for the fact that some devices have been on the market longer than others? Ultimately, Maisel said, it was decided to apply Kaplan–Meier methods to do a "survival estimate," essentially thinking of each device as though it were a patient entering a clinical trial.

The analysis of recall-free 510(k) "survival" of devices showed that 98.4% of the devices cleared in 2003–2009 remained on the market, free of recall, 1 year after the 510(k) decision. The proportion that remained recall-free for 2 years was 96.5%; and the proportion for 5 years, 91.5%. That approach, Maisel said, was thought to be the fairest way to conduct an assessment of the rate of the recalls for the 510(k) program.

Another way to look at the data (using Kaplan–Meier analysis) is to ask what percentage of devices are recalled during their first year on the market, their second year on the market, and so on. The data show that there is a higher rate of recall among 510(k) products during their first 2 or 3 years on the market—about 1.6–1.9%. The recall rate drops off to about 0.9–1.1% in years 5 and 6 on the market as device problems are resolved.

Causes of 510(k) Recalls

Analysis of the recall database for FDA's classification of causes of recalls showed that manufacturing process was the most common cause of a 510(k) recall, accounting for 28.8% of the recalls. Manufacturing-process errors included inadequate control of a process, inadequate environmental controls, and errors in storage, packaging, or labeling. Device design, that is failure of a device to perform as intended despite the product's meeting all its design specifications, accounted for 28.4% of recalls. Materials and components that were nonconforming, contaminated, degraded, counterfeit, or inadequately tested accounted for about 16%. Change control—changes in specifications, programs, or procedures that adversely affected components or devices—accounted for 11.9%. Employee errors and other miscellaneous concerns accounted for 7.1% and 7.5%, respectively.

Predicates

510(k) devices cleared in 2004–2009 were evaluated for the number of predicates cited by their manufacturers in their applications. More than 80% cited 1–5 predicates, about 10 cited 6–10 predicates, and fewer than 5% cited more than 10 predicate. From Maisel's analysis, it appears that applications that cited 1–5 predicates were associated with a lower rate of recall, and applications that cited 6–10 predicates were associated with a higher rate.

Maisel noted that some 510(k) submissions include bundled products, and more than one "device" in a 510(k) application might necessitate more than one predicate. Some of the devices with very high numbers of predicates cited are in vitro diagnostics, such as a laboratory analyzer that might perform multiple tests and require multiple predicates.

If there were multiple predicates, the age of the newest predicate cited was less than 5 years in more than 75% of the 510(k) submissions. Maisel said that devices for which the age of the newest predicate was 1–5 years had a slightly higher recall rate than nonrecall rate.

The age of the oldest predicates was less than 5 years in about 50% of cases, 6–10 years in 25%, and more then 10 years in 25%. Some of the predicates cited, Maisel said, were 15-20 years old, but there was no indication that older predicates were associated with an increased recall rate.

Type and Features of 510(k) Submissions

From 2003 through 2009, 75% of the devices that were not recalled were cleared through traditional 510(k) submissions compared with about 62% in the recall group; 22.3% of the devices that were not recalled were cleared through special 510(k) submissions compared with 34.2% of recalled devices. Maisel suggested that that is a signal that may warrant further investigation as to whether there is something about the special 510(k) process that increases risk. (Devices cleared through abbreviated 510(k) applications were infrequent in both devices recalled and not recalled.)

510(k) devices affected that were recalled were less often cleared by the Office of Device Evaluation (67.9%) than devices that were not recalled (77.0%) and were more often cleared by the Office of in Vitro Diagnostic Device Evaluation and Safety (OIVD) (32.1% vs 23.0% of nonrecalled devices). Again, Maisel said, that may be related to the types of products evaluated by each office and does not necessarily reflect the quality of review. There was a slight signal, he said, that applications undergoing third-party review had a higher rate of recall. There was no difference in the recall rate for implantable devices in the percentages that were recalled or not recalled. Life-sustaining devices were more likely to be recalled, but that could be

BOX 2-2
Maisel's Key Findings: Analysis of FDA Recall Data

- More than 3,000 devices are cleared for marketing each year under new 510(k) submissions. Overall, more than 48,000 510(k) devices were cleared 1996–2009.
- Recalls affect 510(k) devices 400–500 times annually.
- Three-fourths of 510(k) devices are recalled a single time, and one-fourth are recalled two or more times.
- The annual rate of recall of 510(k) products is highest during the first 3 years after clearance; lower recall rates are observed in postmarket years 5 and 6.
- Manufacturing-process and device-design issues are the most common cited causes of 510(k) device recalls.
- Applications citing a large number of predicates, clearance via a special 510(k) process, and third-party review are associated with higher rates of recall. Life-sustaining devices and class III devices are also recalled more frequently than others.

a threshold issue, Maisel suggested; that is, manufacturers are much more likely to issue recalls if they become aware of a potential for serious clinical consequences.

More than 85% of 510(k)-cleared devices were in class II, including devices both affected and unaffected by a recall. Class III devices, which tend to be more often life-sustaining devices, have a higher rate of recall.

Devices assigned to certain advisory committees were more likely to be recalled. Devices assigned to the anesthesia, chemistry, and cardiovascular advisory committees, for example, are generally higher-risk devices and had higher recall rates; but dental, immunology, and microbiology devices were less likely to be recalled. Maisel suggested that this information may be useful in deciding how to allocate resources to the fields in which recalls are more likely (see Box 2-2).

Medical-Device Reporting

As part of his task, Maisel said that he was asked to include MDR data in his analysis but that the available data were not ideally suited to analysis, because of incomplete reporting, insufficient information, and misclassification. He was able to link some of the MDR data with recall data, but Maisel stressed that the data should be interpreted with caution because of the limitations of the sample.

The database included 182,394 MDRs that were associated with 7,823 510(k) devices cleared in 1996–2009. Nearly two-thirds of the reports (66.4%) were associated with a device malfunction; almost one-third (29.5%) were associated with a patient injury, but fewer than 2% involved a patient death.

Of the MDRs in the analysis, 83,000 were associated with products that were ultimately recalled: 750 death adverse-event reports, 19,936 of reported patient injuries, and 60,291 of device-malfunction reports involved 510(k) applications that were associated with recalled devices. There is no information about whether these MDRs came in before or after the recalls, Maisel said. However, the data at least suggest that the recall process is associated with a substantial proportion of the MDRs.

FOOD AND DRUG ADMINISTRATION RECALL-DATA STUDY

The 510(k) clearance process has been subject to substantial criticism, said Ralph Hall, Distinguished Visiting Professor of Law at the University of Minnesota Law School, but there are no systematic means with which to assess whether the system is working. To address that, Hall embarked on a study to determine whether the 510(k) system permits products to enter the market without a "reasonable assurance of safety and effectiveness" and whether specific parts of the 510(k) process lead to greater or lower risk.[5]

Methods

There are three classes of FDA device recalls, which, Hall noted, with regard to risk are in reverse numerical order from device classification itself. Class I recalls involve the most serious safety issues—situations in which there is a reasonable probability that use of or exposure to a violative product will cause serious adverse health consequences or death. Class II recalls involve temporary or reversible medical issues or remote risks, and class III recalls involve nonsafety issues, such as regulatory violations (for example, marketing without proper clearance). The recall classification is determined by FDA.

Hall said that although it is not perfect, the group of class I recalls offer the best safety-related performance measure of the 510(k) system. MDR data are not a good measure, he said, because the reports include known risks, there is inconsistent reporting, information is incomplete or inaccurate, and there is no quality control or confirmation. Medical-device reports are primarily anecdotal, he added. The number of products involved

[5]This study was not commissioned by the committee. Mr. Hall indicated that he would provide the committee with a written report of his findings at a later date.

in recalls also is not a useful measure, because there is no denominator, it is not possible to differentiate single-use products from multiple-use products or to determine failure rates or rates of harm, and recalls include nondefective products.

For his study, Hall focused on class I recalls in 2005–2009. The study included data derived from FDA databases, including databases of recalls, 510(k) cleared devices, PMA devices, product classification, and the Total Product Life Cycle; the 2009 Government Accountability Office (GAO) report; results of ancillary Internet searches; and direct communication with device companies and FDA.

A total of 474 class I recalls during the period were identified. There were often multiple records on a single recall event (involving, for example, different sizes, model numbers, or trade names); when these had been consolidated, it was determined that there were 118 unique class I recalls. Those recalls were then coded by Hall for a variety of factors, including approval or clearance pathway, whether a device was implantable, reason for the recall, device class, third-party review, and medical specialty. Hall indicated that in some instances the coding of the reasons for recall in the system was unclear. Therefore, he developed criteria to determine which category to place devices for which the reason for recall was unclear.

Recalls, Hall said, have three broad root causes: premarket issues, issues that the PMA and 510(k) processes are intended to check and prevent, such as design issues and clinical data gaps; postmarket issues, such as manufacturing issues, labeling mistakes, and sterilization issues; and miscellaneous actions, often taken by unrelated third parties, such as counterfeit products. Hall stressed that the 510(k) system can be expected to prevent only premarket issues, and any assessment of the correctness of clearance decisions or of the robustness of the 510(k) process should look only at premarket issues.

Hall acknowledged several challenges to the method. First, the study relied on public data and the accuracy of the databases. Second, there may be "missing" recalls that are not reported; this is a violation of FDA regulation, and Hall opined that there were probably few cases and that they were probably not major. Third, the study focused on class I recalls because they are the potentially high-end or high-impact safety issues, but there are questions about the consistency of the FDA recall class determinations.

Data Analysis

Of the 118 unique class I recalls in 2005–2009, six involved counterfeit devices, leaving 112 core recalls, for an average of 22.4 class I recalls per year. According to GAO, there are more than 50,000 listed devices (GAO, 2009), so there was a 0.2% recall rate over the 5 years, Hall said.

Hall organized the 118 recalls by primary reason for recall (see

Table 2-1). Of the 13 categories used, four are premarket issues: device design, software design, failure to identify a clinical risk, and failure to warn or inadequate instructions. The other categories are postmarket issues or other concerns, including counterfeit devices. Within categories, recalls were divided according to approval pathway (PMA, 510(k), or class I device).

When all class I recalls are considered, including those for counterfeit devices, premarket issues accounted for about 45% of the recalls involving 510(k) devices. If counterfeits are excluded and only the 112 valid devices are considered, about 48% of the class I recalls of 510(k) devices and about 43% of the PMA device recalls were due to premarket issues. Hall noted that his results indicated that the numbers for 510(k) and PMA devices are not statistically different. Correspondingly, about 55% of the recalls involve postmarket issues.

Most of the premarket issues leading to recall were design issues (including software design). On the basis of his results, Hall said, it is clear that there is a critical role for quality-system regulation (QSR), including bench testing and design controls to identify design issues without endangering patients.

TABLE 2-1 Primary Reason for 118 Class I Recalls

Primary Reason for Recall	PMA	510K	Class I	Other or Unknown	TOTAL
Manufacturing	6	31	2	1	40
Labeling error	0	4	0	0	4
Design issue	6	25	1	0	32
Software design	1	9	0	0	10
Software manufacturing failure	0	2	0	0	2
Supplier issue	2	5	0	0	7
Failure to identify clinical risk	0	0	0	0	0
Failure to warn or inadequate instructions	0	8	0	0	8
Missing parts	0	0	0	0	0
Sterilization	1	4	2	0	7
Regulatory violation	0	1	1	0	2
Packaging or handling	0	0	0	0	0
Other (such as counterfeit or sham)	0	6	0	0	6

Hall pointed out that recalls related to newly discovered clinical risks were identified. Although "inadequate labeling or instructions" could possibly be used to describe unidentified clinical risks, Hall was not able to assess that further with the available data. Overall, about 7% of the recalls were attributed to unidentified clinical issues. Those data suggest, Hall said, that conducting additional human clinical trials would have very little effect on the number of class I safety recalls.

To assess the robustness of FDA device review broadly, one must look at the rate of recalls compared with submissions, Hall said. The number of submissions, not approvals or clearances, is a better measure of robustness of the process because it includes situations in which products were not cleared (thus eliminating safety risks).

A denominator is needed to be able to determine how many class I recalls are related to each type of approval pathway. However, finding an exact denominator is impossible because there is no precise time relationship between submission, clearance, and initiation of a recall.

Thus, Hall looked at the total 510(k) submissions for the 10-year period 2000–2009, computed the per-year average, and then multiplied that average by 5 to estimate the average submissions over a 5-year period. Using the calculated denominator of 19,873 submissions over the 5-year study period 2005–2009 and the total of 89 class I recalls of 510(k) devices during that time, Hall concluded that 99.55% of 510(k) submissions did not result in a class I recall (recall rate, 0.45%). The total of 43 class I recalls due to premarket issues represented 0.22% of total submissions, so 99.78% of 510(k) submissions did not result in class I recalls due to premarket issues.

For comparison, Hall noted that 2.3% of Medicare hospitalizations result in a patient-safety event, there is a 2–4% risk of hospital-acquired infection, and more than 15% of patients over 65 years old receive a potentially unsafe prescription.

A similar analysis of PMA products resulted in similar findings. Class I recalls relative to submissions during the study period were comparable: 99.71% of devices were not subject to recall, and 0.12% were recalled for premarket issues.

Roughly 1% of all medical devices are PMA products. Not surprisingly, given the higher complexity and higher risk of PMA products than of other products, PMA products accounted for 14% of the class I recalls during the study period. About 67% of all devices are exempt or class I device products; these lower-technology, lower-risk products accounted for only 6% of recalls.

Hall further subdivided PMA devices by type of product according to the applicable section of the Code of Federal Regulations (CFR). The bulk of recalls involved cardiovascular devices (21 CFR 870) and hospital and personal use devices (21 CFR 880), and these had higher than average

BOX 2-3
Hall's Key Findings: Using Recall Data to Assess the 510(k) Process

- 99.8% of 510(k) device submissions did not experience a class I recall in a 5-year period.
  - Most 510(k) recalls (55%) involve postmarket issues.
  - Analysis of PMA and 510(k) systems yields similar results with regard to recalls.
- Issues related to some product types (AEDs and infusion pumps) exist.
  - Product-specific "rules" may be appropriate.
  - Continuing review of recalls can aid in early identification and intervention for problem product types.
- Data support the importance of QSR systems.
  - Additional human testing before clearance seems to have limited value.
  - Design controls, bench testing, and preclinical studies seem to be more effective and more ethical.
- The data do not clearly support the need for a fourth device class.

rates of premarket issues. There were very few recalls of orthopedic devices although these are long-term implantable products, no recalls of obstetrical and gynecologic devices, and only one recall of a radiology device. Hall posed the question of whether the data support the need for a fourth device classification. He opined that, on the basis of safety data, there is no clear, discrete delineation of products that would logically fit into a new "class IIB," that is, an enhanced 510(k) or limited PMA process.

The same data were analyzed by medical specialty, and the same general pattern was observed: recalls occurred predominantly in cardiovascular and general hospital specialties.

If the three-letter product codes are looked at, 54.2% of recalls were concentrated in five categories: automatic external defibrillators (AEDs), anesthesia products, cardiovascular devices (a broad catchall category), catheters, and infusion pumps. Two PMA product types, AEDs and infusion pumps, accounted for 28% of all class I recalls during the study period.

Hall raised the question of whether product-specific guidance and special controls would be an appropriate response, and he noted that a detailed root-cause investigation of the products may be warranted.

In summary, Hall concluded that FDA has an excellent safety record on

the basis of his analysis of class I recalls. About 99.8% of device submissions did not experience a class I recall during the 5-year period studied (see Box 2-3).

He stressed that QSR is extremely important given the prevalence of design issues and is probably more important than additional human clinical studies.

Hall highlighted several questions for consideration, including

- What aspects of postmarket surveillance have the greatest effect on identifying recall needs?
- What are the true root causes of safety recalls (such as common factors, human factors, and complexity)?
- What are the potential effects of changes in the 510(k) system (such as effects on FDA resources and time, on whether the added burden of changes would be proportional to safety benefits, and on effects on patient access)?
- Which parts of clearance and approval submissions make a difference relative to improving safety decisions?
- What is the role of multiple predicates in recall situations?

REFERENCES

FDA (Food and Drug Administration). 2000. *Device Advice: Premarket Notification. Washington, DC, Food and Drug Administration.* http://www.fda.gov/medicaldevices/deviceregulationandguidance/howtomarketyourdevice/premarketsubmissions/premarketnotification510k/default.htm (accessed August 31, 2010).

GAO (General Accountability Office). 2009. *FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved Through the Most Stringent Premarket Review Process.* http://www.gao.gov/new.items/d09190.pdf (accessed August 16, 2020).

IOM (Institute of Medicine). 2005. *Safe Medical Devices for Children.* Washington, DC: The National Academies Press.

3

Non–Food and Drug Administration Sources of Adverse Event Data

The Food and Drug Administration (FDA) partners with external groups to collect and analyze postmarket surveillance data. The committee requested information from some of FDA's partners as well as other groups that collect adverse event information related to medical devices. The committee was presented with overviews of four non-FDA sources of adverse-event data with a focus on the collection of information about specific clinical activities or specific kinds of devices.

THE NATIONAL CARDIOVASCULAR DATA REGISTRY: OPPORTUNITIES AND CHALLENGES IN POSTMARKET SURVEILLANCE

"We are awash in data," began Frederick A. Masoudi, associate professor of medicine at the Denver Health Medical Center and the University of Colorado and senior medical officer of the American College of Cardiology National Cardiovascular Data Registry (NCDR). It is not necessarily a lack of data that is the problem, he said, but the lack of an ideal data source.

An ideal data source for postmarket device surveillance, he suggested, is one that includes

- Clinical data (avoiding administrative data whenever possible).
- Standardized definitions (collection of the same data elements for the same event).
- Detailed phenotyping of patients (collection of a wide array of clinical characteristics that could be used for risk adjustment).

- Real-world populations (that is, not in the context of a clinical trial).
- A population denominator.
- Followup for adverse events.

The mission of the NCDR is "to improve the quality of cardiovascular patient care by providing information, knowledge, and tools; implementing quality initiatives; and supporting research that improves patient care and outcomes." Although the NCDR was not set up primarily for postmarket surveillance, Masoudi said, it is one of its benefits. The NCDR is actually a suite of cardiovascular-disease registries (see Box 3-1). Several of the NCDR registry programs are focused on procedures and devices, Masoudi noted.

The data-collection platform is similar in all the NCDR registries, using electronic data collection to capture a wide array of detailed demographic and clinical characteristics for each patient, details of procedures that are performed, and inhospital complications. The data include detailed information on the devices used for each procedure.

Each institution that submits data to the NCDR receives quality benchmark reports that can be used to support self-assessment and quality improvement by the institution.

BOX 3-1
NCDR Registries

CathPCI Registry—for diagnostic catheterization and percutaneous coronary intervention (PCI) procedures. Includes data since 1998 from over 1,100 hospitals on more than 9 million patients.

ICD Registry—tracking implantable cardioverter defibrillators (ICDs). Includes data from 1,445 hospitals on more than 250,000 patients.

CARE Registry—for carotid artery revascularization and endartectomy procedures.

ACTION-GWTG Registry (Get With The Guidelines)—acute–myocardial-infarction patients.

PINNACLE Registry (Practice Innovation and Clinical Excellence)—an ambulatory care, practice-based registry.

IMPACT Registry (Improving Pediatric and Adult Congenital Treatment)—tracking cath procedures for congenital heart conditions.

Postmarket Surveillance with the National Cardiovascular Data Registry

As an example of postmarket device surveillance with the NCDR, Masoudi cited a study led by Paul Varosy looking at complications in implantable cardioverter defibrillators (ICD) patients (Dewland et al, 2008). It has been suggested that dual-lead ICDs may be superior to single-lead devices, but, Masoudi said, there are concerns that the use of dual-lead devices when there is not a clear need may be associated with higher rates of complications.

On the basis of the registry, 206,000 patients who underwent ICD implantation were identified. Patients who received biventricular ICDs or had clear indications for dual-chamber devices were excluded from the analysis. Of the remaining patients who had no clear reason to receive a dual-lead device, about half had received single-lead devices and half dual-lead devices ("discretionary" dual-lead ICD placement).

The analysis showed a substantially higher rate of major inhospital complications and a higher risk of death associated with dual-chamber devices than with single-chamber devices. Both findings were statistically significant, Masoudi noted, even after risk adjustment based on a wide array of patient characteristics that are collected in the registry.

Another example Masoudi cited used the CathPCI Registry to assess the rates of bleeding complications associated with different closure devices used at the groin site after angiography and percutaneous coronary intervention (PCI) (Travis et al., 2005). After adjustment for various characteristics among the different patient categories, the analysis showed that the VasoSeal device was associated with a significantly higher risk of adverse outcomes after angiography than other hemostasis devices. As a result of the analysis, Masoudi said, use of the VasoSeal device was removed from the market.

Several postmarket surveillance collaborations between the NCDR and FDA are going on, Masoudi said. FDA task orders include assessment of device use in carotid revascularization, ICD lead safety, risk factors for ICD malfunction, data elements and metrics for an atrial fibrillation (AF) ablation registry, and a dataset for the National Congenital Heart Disease Registry. The NCDR and FDA are also drafting a white paper discussing the value of registries, such as the NCDR, for postmarket surveillance.

Several challenges to device surveillance using the NCDR, Masoudi said, include followup, data quality, and integration of data collection into care. Cost is an overriding concern; registries are extremely expensive to develop and to maintain.

Strategies for Longitudinal Followup

A number of strategies have been used to address the challenge of longer-term surveillance using registries which are primarily hospital based,

Masoudi said. For example, NCDR data have been linked to administrative data, such as those from the Centers for Medicare and Medicaid Services (CMS), to provide followup of clinical events (for example, later hospitalizations, death, and complications that might be coded in administrative datasets). In some cases, that has been done probabilistically: instead of exact patient identifiers, several aspects of patients' episodes of care are used to find a match with the Medicare data. Although it is probably not adequate for the purposes of device-based postmarket surveillance, Masoudi said, the probabilistic matching approach (as opposed to direct patient matching) has been shown to be a valid method of data analysis. This is particularly relevant in light of the limitations on patient identifiers as a result of the Health Insurance Portability and Accountability Act privacy rule.

Masoudi referred to a recent report from the Institute of Medicine that concluded that the privacy rule does not do enough to protect patient privacy in all situations and substantially impedes research or efforts to improve public health. The privacy rule is vague in many respects, he said, and it is often assumed that the most conservative approach is the most appropriate one, particularly because the penalties for nonadherence to the privacy rule are very high. As a result, efforts to comply with the rule interfere with important research, in this case in identifying linkages between NCDR data and direct Medicare data.

Another strategy for longitudinal followup is collaboration with health systems that own all their administrative data, such as Kaiser Permanente and the Veterans Health Administration. That allows direct connection between patients in the registries and followup in the system instead of a probabilistic matching strategy.

A final strategy described by Masoudi involves forming a relationship between inpatient registries, such as the CathPCI Registry and the ICD Registry, and outpatient registries, such as the new PINNACLE Registry. The fragmentation of the health-care system creates some challenges to this approach, Masoudi said, for example, how to ensure that the same patients who are followed in a hospital with the ICD Registry are then followed in an outpatient setting with the PINNACLE Registry or another outpatient registry.

The NCDR has a number of data-quality procedures in place as part of a larger quality program, for example, data-quality checks where sites are given green-light, yellow-light, or red-light status on the basis of the quality of the data that they submit to the registries (completeness of the data and range checks of the different data fields).

A limited audit is also performed on a clinical level by the NCDR, but because of insufficient resources, this is done for relatively small numbers of charts and clinical records, Masoudi said. A data-quality program that

ensures the fidelity of data, particularly if it involves clinical-data abstraction, requires substantial resources.

When a physician or other clinical professional enters into a patient chart that, for example, the patient has class III angina, that information does not automatically get into the NCDR. It needs to be re-entered in many cases, and this adds to the challenge of fostering participation in the registries.

One very successful approach to address this, Masoudi said, has been instituted in the Department of Veterans Affairs (VA) health system. In the Clinical Assessment, Reporting, and Tracking System for Cardiac Catheterization Laboratories (CART-CL) program, for example, data collected in the process of routine clinical care are tailored to the NCDR data fields and then entered into the program for device surveillance.

THE DEPARTMENT OF VETERANS AFFAIRS CARDIOVASCULAR ASSESSMENT, REPORTING, AND TRACKING PROGRAM: INTEGRATION OF REAL-TIME DATA COLLECTION INTO THE PROCESS OF CLINICAL CARE

Paul D. Varosy, director of cardiac electrophysiology in the VA Eastern Colorado Health Care System and assistant professor of medicine at the University of Colorado Denver, provided an overview of the VA CART program, describing it as a new paradigm for cardiovascular-disease surveillance and a potential model for medical-device surveillance.

Data Resources

The VA-wide electronic health record, known as the Computerized Patient Record System (CPRS), is an outgrowth of the national VA medical-record system in place since the 1970s. CPRS was developed in the middle 1990s as a rich graphical user interface for health-care data and incorporates text notes and reports, laboratory data, electronic order entry, pharmacy records, and images (such as electrocardiograms and radiology records). It is organized and managed at the regional level by the 23 Veterans Integrative Service Networks, which are all linked to a single nationwide network.

Clinical and administrative data from the VA facilities are warehoused at the Austin Information Technology Center. Data are aggregated and processed for multiple potential uses, including workflow tracking, quality of care and quality assessment, and health-services research. Varosy added that access controls prevent breeches in data security.

Data are also obtained from the VA Office of Patient Care Services, which has clinical oversight over cardiovascular services. The office has specific programs for quality monitoring and improvement and national

programs, such as the pacemaker and ICD surveillance programs and the CART program.

Limitations of Administrative Data

There are some important concerns about administrative data, such as data that are aggregated into the VA datasets, Varosy said. First, most parts of the clinical record are not entirely field-specific. There are text notes and reports that need to be abstracted, either by manually going through the individual records (which is labor-intensive) or by some complex natural-language processing extraction that is cumbersome and difficult to interpret. There can also be a "loss in translation" of information and a lack of clinical granularity of the data.

A second concern is lack of standardization. With regard to heart function, for example, left ventricular ejection fractions can be measured in various ways, including echocardiography, radionuclide ejection fraction measurement, radionuclide myocardial perfusion study, cardiac catheterization, and magnetic resonance imaging. Results can be reported in different places in accordance with different standards, such as a specific number of an ejection fraction, for example, 36.7% as measured with one of the various methods; an estimate of 35–40%; or a qualitative estimate, such as "moderately depressed left ventricular function." The need to collate information from different sources based on different studies and different ways of codifying the results makes it difficult to interpret the data.

Finally, Varosy said, dependence on administrative coding is problematic in a system where coding is not tied to reimbursement. For example, in the VA system there is relatively little incentive to ensure correct coding according to Current Procedural Terminology (CPT) and *International Classification of Diseases 9th Revision* (ICD-9) when there is no reimbursement on the basis of these codes. However, analyses of the VA database often use the codes for disease surveillance and quality assessment.

As a result, abstraction of data after care is necessary because data collection is generally not integrated directly into the process of clinical care.

Offering an analogy, Varosy suggested that using administrative data, such as *ICD-9* codes or CPT codes, for disease surveillance is like trying to monitor air traffic by reviewing jet-fuel receipts (if a plane refueled in Omaha and then again in Newark, it must have flown from Omaha to Newark). The air-traffic control system is instead designed to see where planes are flying in real time. That is not the case for health-care data.

Department of Veterans Affairs Patient-Care Services Clinical Programs

For nearly 30 years, VA has been a leader in remote pacemaker monitoring. The VA Pacemaker Surveillance Program includes remote followup of pacemaker function, administrative tracking of clinical and administrative cohorts, and support of clinicians and voluntarily enrolled patients. Patients without easy access to a VA hospital are monitored by telephone every 3 months to assess all pertinent characteristics (such as battery life) remotely. The program is administered from two sites: Washington, DC, and San Francisco.

The VA National ICD Surveillance Center (VANISC) was established in 2003 on the basis of the successful pacemaker surveillance program. VANISC monitors voluntarily enrolled patients who have ICDs remotely for arrhythmia episodes and reports results directly to the patients' providers. It also facilitates disease surveillance and research studies.

To provide a sense of the scope, Varosy said that the western pacemaker surveillance program and the ICD program combined (both using secure data servers based in San Francisco) remotely monitor more than 18,000 veterans, including more than 12,000 veterans who have implantable ICDs. In FY 2009, the staff of 13 reviewed and provided support for over 32,000 pacemaker transmissions and over 40,000 ICD transmissions.

The remote monitoring programs have limitations, Varosy noted. Enrollment in the programs is voluntary. Linkage of remote monitoring programs to electronic health records is problematic, and there is a lack of infrastructure to connect remote and in-clinic device followup. Most important, he said, the ascertainment of long-term clinical outcomes is challenging (outcomes data are necessary for quality improvement, device performance and surveillance, and health-services research).

The Cardiovascular Assessment, Recording, and Tracking Program: A New Paradigm for Care

CART is a clinical tool that improves the efficiency of care by integrating data collection with electronic health records and facilitating report generation. The user interface, designed with clinicians in mind, incorporates VA-wide standardization and allows completion of reports in real time, often, Varosy said, before a patient is even off the examination table in the cardiac-catheterization laboratory.

The integration of data collection into the transaction of health care, Varosy said, allows transactional quality management, real-time patient-safety monitoring, real-time device surveillance, and nearly real-time health-services research.

Critical to the success of CART, Varosy said, are strategic collaborations with clinical champions in the 77 catheterization laboratories in the VA

system nationwide, the VA Office of Patient Care Services, the VA Quality Enhancement Research Initiative (QUERI), the VA Office of Quality and Performance, the VA Office of Information and Technology, and outside VA (for example, in the NCDR and in FDA). Varosy noted that VA has well-established connections with FDA, including monthly CART conference calls with FDA in which information about signals or unexpected problems with devices reported by clinicians at the point of care are shared.

Using the CART-CL program as an example, Varosy demonstrated the user-friendly CART interface consisting of checklists, drop-down boxes, and the opportunity to add text comments. He noted that many of the fields can be prepopulated directly from electronic medical records for data on such items as medications, allergies, vital signs, laboratory studies, and past medical history. For CART-CL, the coronary-angiography documentation process includes images with intuitive interfaces designed specifically for clinicians, which allow rapid and granular notation of specific lesions observed. Once all the data are entered, a uniform text report is generated that can be copied and pasted directly into the text-reports field in an electronic medical record. At the same time, all the data are captured into a database that resides in a secure server within the VA firewall.

Since 2005, nearly 140,000 total cardiac procedures have been documented with the CART-CL system by nearly 3,000 providers (including cardiology fellows in training and attending cardiologists). For decades, Varosy said, VA has been basing its workflow tracking and administrative information on the data resources from the Austin Information Technology Center. He pointed out, however, that for FY 2008, the CART program recorded 7,972 total cardiology procedures, but on the basis of purely administrative codes within VA the Austin Information Technology Center recorded 4,079 cardiology procedures (slightly more than half the number recorded in the CART system).

Transactional quality management in CART includes immediate e-mail reporting of major complications (for example, inhospital or intraprocedure stroke, death in a laboratory, or the need for emergency cardiac surgery during a cardiac-catheterization procedure). Immediate, secure, encrypted e-mail messages are sent to the chief cardiovascular consultant, CART leadership, and the CART Quality Management Committee chair. Committee review and preliminary recommendations occur within 24–72 hours after an event. Formal root-cause analysis or other interventions necessary to produce systemwide improvements in care may occur later. There are also monthly site quality-assurance reports, monthly and biannual national procedure and adverse-event count reports submitted to the VA central office and the CART Quality Management Committee, and quarterly regional reports submitted to the network administrators and chief medical officers.

Beyond the original CART-CL application, VA is creating modules to

address peripheral arterial intervention (CART-Peripheral), arrhythmia procedures (CART-EP), inhospital cardiac arrest (CART-CPR), and ambulatory care (CART-Ambulatory).

Data elements used are based on well-established data standards, including those of NCDR, and work is under way to construct a data link so that data captured in real time in CART-CL will be sent directly to NCDR. Similar integration efforts are planned for other CART modules and NCDR databases. In implementing CART-EP, VA hopes to provide a clinically useful reporting tool that will integrate with the ICD and pacemaker surveillance programs and allow transactional data collection.

The CART program, Varosy concluded, moves beyond the era of after-the-fact data collection to one of transactional data collection, leveraging real-time data to allow not only clinical care but real-time quality management, real-time workflow tracking, and real-time health-services research.

USE OF REGISTRIES FOR POSTMARKET DEVICE SURVEILLANCE

Eric D. Peterson, professor of medicine and associate director of the Duke University Medical Center and director of cardiovascular research at the Duke Clinical Research Institute, reiterated some of the issues that can arise after a device reaches the market, including rare events not observed in premarket evaluation, downstream safety events and long-term followup of devices that remain in patients for long periods, use in a wider array of high-risk patients, off-label use, device–drug interactions, and device–health-care provider interactions (the "learning curve").

Among the challenges for FDA, Peterson said, are the rapid evolution of technology, which can make device studies quickly obsolete, lack of incentive or ability of device developers to conduct high-quality premarket and postmarket evaluations, and the fact that the clinical community (including patients) can be very eager for access to new devices and often does not demand clinical evidence or does not support the conduct of clinical studies (enrolling patients in clinical studies is challenging).

A Role for Registries

Can clinical registries address some of those challenges? Peterson asked. A creative sentinel system to track medical devices once on the market could, he said, provide an idea of actual device use (a denominator for analysis), gather data on off-label use, and identify potential device-safety signals.

Peterson provided examples of existing clinical-device registries. A device manufacturer may establish its own device-specific registry, as Medtronic has done with its pacemaker registry. There are also multisponsor device registries, such as the Interagency Registry for Mechanically Assisted Cir-

culatory Support (INTERMACS) database, a collaboration of the National Heart, Lung, and Blood Institute, CMS, and FDA. Professional organizations may house registries, such as the clinical cardiovascular registries established by the Society of Thoracic Surgeons (with over 900 centers registering all bypass, valve, cardiac, and thoracic surgery), and the NCDR (discussed by Masoudi above). Peterson added that relevant data have come from randomized controlled trials, and in the future data may be available directly from electronic health records.

Efforts are now under way to build more in-depth device information capacity into each of those registries progressively and to link data from clinical registries with claims data (particularly Medicare claims) to provide information on longitudinal outcomes, such as rehospitalization and device explantation.

Examples of Food and Drug Administration Partnership with Registries for Purposes of Postmarket Surveillance

Peterson highlighted several examples of the use of clinical registries for postmarket assessment of devices. Drug-eluting stents, Peterson said, are remarkable tools in the hands of interventional cardiologists but presented a substantial challenge to FDA because of the very rapid adoption of the devices for both label and off-label uses. Analyzing data from the NCDR, Peterson and colleagues found that the adverse-event rate appeared to be higher for off-label indications, but it was not clear whether this was attributable to the device or to the patient population in which it was used (Rao et al., 2006). Another study using the data bank from the Duke Heart Center showed that patients who had drug-eluting stents in place and who did not remain on dual antiplatelet therapy for long periods had higher adverse-event rates and mortality (Eisenstein et al., 2007). Those findings and other signals from clinical trials and other larger databases prompted FDA to initiate a public debate on safety issues of drug-eluting stents and to establish an advisory panel. Peterson noted that after those publications and presentations, the use of drug-eluting stents in practice declined.

Ultimately, the Agency for Healthcare Research and Quality (AHRQ) and FDA commissioned a database to examine the comparative effectiveness and safety of drug-eluting stents vs bare-metal stents in a national PCI cohort. The DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Cardiovascular Consortium linked NCDR data on 262,700 PCI patients from 2004 through 2006 to CMS claims data (based on indirect identifiers) to assess long-term outcomes (up to 3 years after stent placement). Peterson said that the analysis could not confirm a unfavorable safety signal for drug-eluting stents (the results with drug-eluting stents were equal to or better than those with bare-metal stents).

As another example, Peterson described an early initiative in which the Duke Clinical Research Institute partnered with FDA and the Society of Thoracic Surgeons to look at the safety and use patterns of transmyocardial revascularization (TMR). Diffusion of the technology into clinical practice was rapid and included off-label use of TMR combined with coronary arterial bypass graft procedures (Peterson et al., 2003). A followup study is under way with a retrospective cohort to compare long-term clinical outcomes of the use of two existing laser devices.

Other examples that Peterson cited included a large FDA-sponsored partnership with the Society of Thoracic Surgeons to compare clinical outcomes of the use of two marketed endoscopic vein-harvesting devices and a series of studies being conducted by the DEcIDE Consortium, supported by AHRQ in partnership with FDA and the Society of Thoracic Surgeons, to evaluate the clinical effectiveness and safety of marketed biologic vs mechanical aortic valve prostheses in older patients.

Those organizations are now working together to develop standardized nomenclature to allow linking of data across datasets, Peterson said. There is also interest in using a registry as a backbone to carry out large clinical trials.

In conclusion, Peterson said that postmarket device surveillance is and will remain an important issue for health care. Clinicians need to be actively involved and to demand better device information and identify device issues. Ideally, clinical registries can be used to provide novel solutions for effective and efficient postmarket surveillance.

AUTOMATED POSTMARKET SAFETY SURVEILLANCE: THE DELTA SURVEILLANCE PROJECT

Medical-device safety surveillance today is primarily passive, said Frederic S. Resnic, director of the Cardiac Catheterization Laboratory at Brigham and Women's Hospital and assistant professor of medicine at Harvard Medical School. Automating prospective postmarket surveillance involves integration of high-quality data sources, appropriate safety expectations, monitoring systems, and secure data exchange.

Challenges

Resnic concurred with previous speakers regarding the array of challenges associated with device safety monitoring. Completeness and level of granularity of current datasets is a primary challenge. The lack of unique identifiers for devices affects the utility of clinical registries for understanding who has been exposed to which devices. Surveillance is affected by how quickly data become available, and appropriate and comprehensive

outcome ascertainment remains challenging. There are also the issues of data ownership, data security, and patient privacy.

Another major concern is signal detection and methods. Appropriate expectations need to be set, and appropriate comparators chosen. For an active, automated surveillance system to function, information needs to be converted into an alert that could notify regulators when a device appears to be heading out of bounds relative to performance expectations.

Given the availability of high-quality data and the ability to detect a signal, the next challenge is signal interpretation. Any alerts generated in observational surveillance must be verified through detailed clinical and statistical exploration. Device–operator, device–patient, device–drug, and device–device interactions all come into play.

Idealized Safety-Monitoring System

In one approach to surveillance, multiple data sources (such as hospitals) submit information to a centralized database or data owner, and the combined dataset is monitored. For reasons of data security and privacy, however, some data owners may be unable or unwilling to deposit their data into a central database. Rather, they share their data in a virtual fashion. Thus, a monitoring system must also be able to handle distributed datasets.

Resnic described an ideal monitoring system as one that is continuously updated, with data provided in as close to real time as possible, and that has an array of statistical analytic options. Systems should be able to handle multiple prospective analyses simultaneously, he said, and such analyses would be running in the background with various alerting thresholds.

The DELTA Surveillance System

In an effort to address some of those challenges, Resnic and colleagues are developing the Data Extraction and Longitudinal Time Analysis (DELTA) System, a Web-based platform designed to perform automated, real-time monitoring of device postmarket safety. DELTA has a matrix of analytic options (see Figure 3-1) that support exploration of potential safety events and draw inferences from both a frequentist perspective and a Bayesian perspective for uniform, stratified, and risk-adjusted expectations. The system generates and e-mails alerts as appropriate. DELTA is designed to run continuously in the background, much like industrial process-control systems in manufacturing plants, Resnic noted.

For the development and validation of DELTA, Resnic used historical data from the Massachusetts state registry. In 2002, he explained, the Massachusetts Department of Public Health implemented mandatory clinical-outcomes registries for invasive cardiac services, which required all

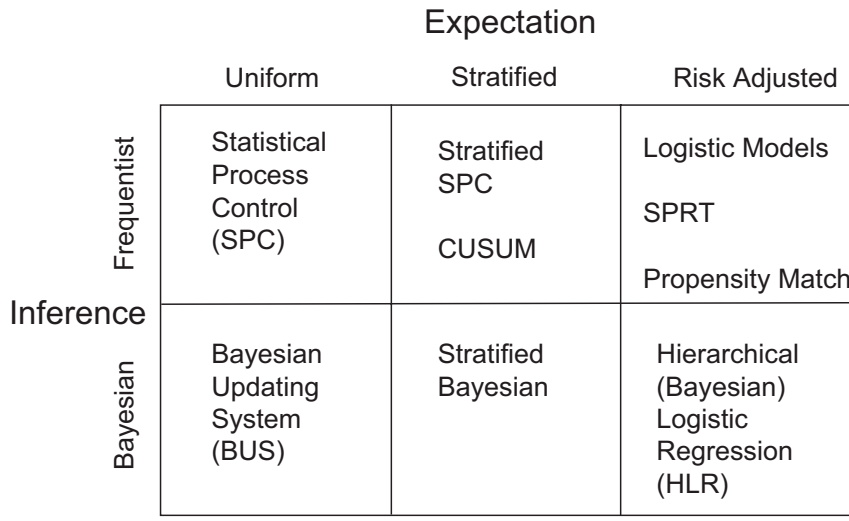

FIGURE 3-1 Statistical methods used in the DELTA System.
NOTE: CUSUM, cumulative sum control chart; SPRT, sequential probability ratio test.

hospitals in the state to submit data to the NCDR CathPCI Registry and the Society of Thoracic Surgery database and to harvest and report those data quarterly to the state registry. At the state level, reports are rigorously adjudicated and audited, and outcomes are linked to vital statistics and inpatient claims data (Resnic noted that this is a direct link, not a probabilistic link). For the retrospective-surveillance demonstration study, 74,000 cases of coronary intervention performed from the launch of the state registry in 2003 through 2007 were used. Devices that were first marketed within 6 months of the launch of the state registry were evaluated, and patients who received such devices were compared with propensity-matched patients who received competing devices of the same class (for example, a drug-eluting–stent patient under study was matched to another drug-eluting–stent patient according to about 35 variables).

A multicenter study that will test the prospective surveillance functionality of DELTA is under way (see Figure 3-2). For the purposes of the study, each participating hospital has a local DELTA system. Through secure data transmissions, the local DELTA agent communicates de-identified, encrypted data to the central DELTA collecting server. Only the data necessary to facilitate the analysis are sent.

To address concerns about facility disclosure of data to a central repository, in this case DELTA, the study will test three levels of data access: case-level data aggregation to the central database, with fully de-identified

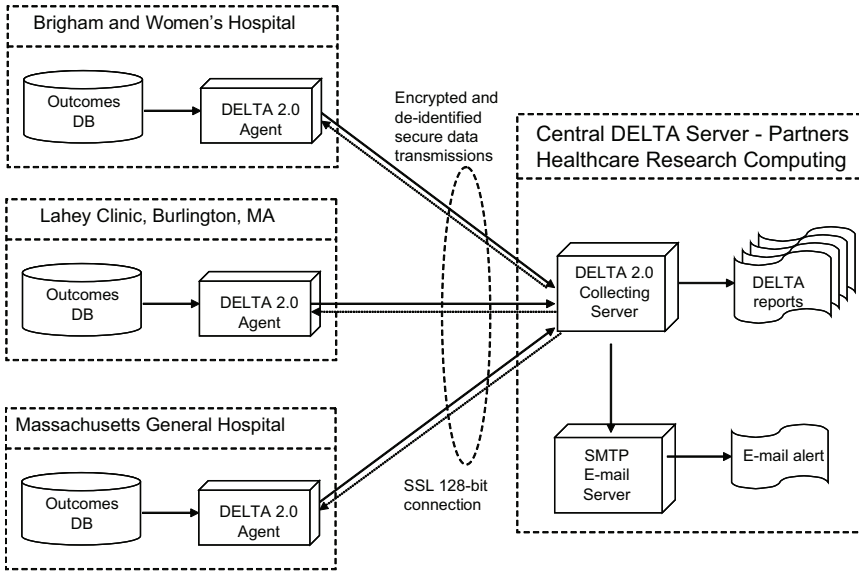

FIGURE 3-2 Prospective DELTA Network Study.
NOTE: DB, database; SMTP, simple mail transfer protocol.

data submitted; case-level outcome aggregation, in which only encrypted case identification, outcomes, and predicted outcomes are sent to the central server; and aggregated results of analyses performed at the local level, with no case-level information.

Costs, Resnic added, are primarily for collecting the data. In Massachusetts, a single implementation of DELTA costs around $25,000–30,000 for the hardware and the software license. The cost of collecting data is borne by the hospitals. However, because of other regulatory requirements for quality management within the hospitals and because many institutions in Massachusetts have adopted the CART-CL model, in which data are collected in real time as part of the clinical process, it is hard to tease out the actual cost of data collection, which has become part of the routine process.

Summary

Detection of low-frequency postmarket safety signals for medical devices challenges traditional methods of statistical surveillance, Resnic concluded. The ideal surveillance system is a time-efficient, high-sensitivity alert system that is designed to trigger detailed investigations of potential safety concerns. Such systems clearly require accurate, granular outcomes data and

device-specific identifiers. Resnic noted that the Massachusetts mandated cardiac registry provides such information.

The DELTA system is a prospective approach to surveillance. It provides flexible statistical and risk-adjustment methods for multiple simultaneous analyses and, Resnic said, meets the design requirements for many of the features of an automated safety surveillance system. Resnic noted that alerts based on analyses must be considered hypothesis-generating and require epidemiologic confirmation. In a future in which registries exist for many products, Resnic opined, systems like DELTA could be used to target regulatory resources and focus efforts on device pairs in which there is an outcomes signal.

Continuing testing of DELTA in a multicenter network study will provide an opportunity to evaluate the applicability and potential role of automated surveillance as a complement to existing methods and as a component of overall active surveillance strategies for new medical devices, Resnic said.

There is plenty of opportunity for further study in existing high-quality registries, Resnic noted. Future DELTA studies include a more in-depth exploration of the Massachusetts cardiac quality data focused on longitudinal outcomes, pilot studies with the VA CART-CL and the NCDR CathPCI Registry, an Orthopedic Implant Registry study, and a cardiac surgical valve safety-surveillance pilot study.

REFERENCES

Dewland, T. A., C. N. Pellegrini, Y. Wang, G. M. Marcus, and P. D. Varosy. 2008. Abstract 4127: Dual chamber ICD selection is associated with racial and socioeconomic disparities and increased complication rates among patients enrolled in the NCDR ICD Registry. *Circulation* 118:S834-S835.

Eisenstein, E. L., K. L. Anstrom, D. F. Kong, L. K. Shaw, R. H. Tuttle, D. B. Mark, J. M. Kramer, R. A. Harrington, D. B. Matchar, D. E. Kandzari, E. D. Peterson, K. A. Schulman, and R. M. Califf. 2007. Clopidogrel use and long-term clinical outcomes after drug-eluting stent implantation. *Journal of the American Medical Association* 297(2):159-168.

Peterson, E. D., P. Kaul, R. G. Kaczmarek, B. G. Hammill, P. W. Armstrong, C. R. Bridges, and T. B. Ferguson Jr. 2003. From controlled trials to clinical practice: monitoring transmyocardial revascularization use and outcomes. *Journal of the American College of Cardiology* 42(9):1611-1616.

Rao, S. V., R. E. Shaw, R. G. Brindis, L. W. Klein, W. S. Weintraub, and E. D. Peterson. 2006. On- versus off-label use of drug-eluting coronary stents in clinical practice (report from the American College of Cardiology National Cardiovascular Data Registry [NCDR]). *American Journal of Cardiology* 97(10):1478-1481.

Travis, D. R., S. Dey, B. Albrecht-Gallauresi, R. G. Brindis, R. Shaw, W. Weintraub, and K. Mitchel. 2005. Risk of local adverse events following cardiac catheterization by hemostasis device use—phase II. *Journal of Invasive Cardiology* 17(12):644-650.

4

Postmarket Surveillance of Medical Devices: Panel Discussion

After the presentations, workshop speakers assembled for a panel discussion of postmarket surveillance of medical devices. Gardner, Hall, Maisel, Masoudi, Peterson, Resnic, and Varosy were joined by Susan Alpert, vice president of regulatory affairs and compliance at Medtronic, and Larry Kessler, professor and chair of the Department of Health Services of the University of Washington, former director of the Office of Science and Technology of the Food and Drug Administration (FDA), and former director of the Office of Surveillance and Biometrics of the FDA Center for Devices and Radiological Health.

Moderator and committee member Lazar Greenfield provided Alpert and Kessler with the opportunity to make opening remarks. In the discussion, panelists expanded on the topics of predicates; device development, including the conduct of clinical studies; data collection and data-sharing; unique device identifiers; encouraging broader participation in device surveillance systems; and risk communication.

SUSAN ALPERT: INDUSTRY DEVICE SURVEILLANCE

As device manufacturers develop technologies and evaluate them, Alpert said, they have expectations for the performance of the devices in the marketplace. Manufacturers track and trend performance input from the field against those expectations and against previous experience. Each field report is evaluated to determine whether it meets the requirements for reporting to FDA and to global regulators and whether there is a potential need for a modification in design or manufacturing.

Many companies have voluntarily established their own product registries. Alpert highlighted two of Medtronic's voluntary databases that are being used for studies. The Systems Longevity Study is looking at survival of implanted leads to determine durability and long-term functionality. For over 27 years, she said, Medtronic has been capturing information about its cardiovascular devices. That is important because generally when a lead fails it is not extracted, so the manufacturer cannot gather forensic information on the device. She noted that information on more than 75,000 leads in 14 countries has been evaluated thus far.

The Implantable Systems Performance Registry was created in 2003 to monitor Medtronic's infusion and neuromodulation devices (such as implanted drug pumps and spinal-cord stimulators). The registry includes data on more than 5,000 patients in 50 centers. Alpert added that Medtronic evaluates and publishes its data in semiannual product-performance reports.

Another example of device surveillance that Alpert described is Medtronic's CareLink monitor. Many active implantable devices have electronic monitoring systems, often bedside monitors, to which the devices automatically send information about the patients and the performance of the devices every night. Through CareLink, this information can be transmitted to the provider for remote monitoring.

Alpert stressed that both industry and FDA depend heavily on physicians and end users as reporters. She asked the committee to consider how industry could interface differently with the clinical community to achieve better access, not only to information about devices in practice but to the products that need to be returned, so that forensic work can be done to assess defects.

LARRY KESSLER: ADVANCING SURVEILLANCE

Kessler offered several suggestions, directed to three constituencies, for advancing device safety and surveillance.

Congress, he said, should provide additional resources for FDA. Resources are needed to focus on the risk issues related to 510(k) products, not just the traditionally high-risk products. Development of registries may be helpful, he said, as would additional Section 522 studies. The rate-limiting factor, he said, is identification of the problems.

FDA regulation on unique device identification is critical, Kessler said. Although it was required legislatively in 2007, FDA has not yet issued a rule, and it needs to do so.

FDA should also enforce known engineering standards and foster attention throughout the centers to issues of risk, benefit, and quality systems, Kessler said. The risk-management strategies devised by manufacturers when they are developing 510(k) products should be widely available to

both premarket reviewers and postmarket and compliance staff. That is generally not the case at the 510(k) level, he said.

Finally, clinical professional societies should encourage clinicians to ask questions about devices and foster enrollment of patients in device registries and other studies.

PREDICATES

A committee member raised the issue of the high percentage of 510(k) product recalls that are attributed to design flaws and raised the question of how the source of a flaw may be associated with predicate devices.

Alpert agreed that a device on the market that is used as a predicate may be somewhat different from what was cleared as a result of multiple small changes. But a 510(k) submission for a new device is more than just a statement that something is similar to something else, she said. A tremendous amount of testing and evaluation is involved in a 510(k) submission, in some cases including side-by-side testing against another device. There are also 100 or so 510(k) device-specific FDA guidance documents that require a manufacturer to conduct specific kinds of testing according to specific standardized test methods.

Alpert said that industry is required to track all changes of products but noted that not all changes rise to the threshold of submission to FDA. Industry uses the FDA guidance document that specifies the kinds of changes that need to be reported to FDA.[1]

Kessler added that when evidence of a new problem emerges, the agency can require Section 522 studies for previously cleared 510(k) products that are on the market.

There are alternatives to the 510(k) predicate approach, Kessler said, and he urged the committee to look at other models, such as the Global Harmonization Task Force guidance Essential Principles of Safety and Performance of Medical Devices, which is used in the European Union and elsewhere (GHTF, 2005).

DEVICE DEVELOPMENT: DESIGN AND STUDIES

Kessler said that there are not enough opportunities for comprehensive dialogue early in device design and development. Design occurs at one point, interaction with clinicians at another, discussions with FDA at yet another, and then discussion with Centers for Medicare and Medicaid Services (CMS) and other contract carriers. Earlier, more comprehensive

[1] *Deciding When to Submit a 510(k) for a Change to an Existing Device*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm080235.htm#page27.

discussions are important, he said, for finding out what questions clinicians and their patients have.

Alpert noted that industry does engage the clinical community in most of the original design of products, and most of the changes that are made later are driven by users.

Kessler said that there is an eagerness to get new devices to the market and into use, but academicians need to get the word out that real-time pragmatic studies are necessary.

Maisel said that there should be discussion about the type and strength of evidence needed to ensure safety and efficacy instead of the assumption that the randomized clinical trial is the gold standard for every 510(k) device.

The primary incentive for any action by a manufacturer is getting a product to market. Peterson said that the reason that drugs companies conduct large randomized trials and device manufacturers do not is that regulations require trials for approval to market a drug. On the positive side, the current 510(k) structure fosters a high degree of innovation in devices; on the negative side, it is challenging to entice industry to do studies because they can get a product cleared for marketing fairly easily without them.

Clinical demand (or the lack thereof) for information also plays a role in performance-data collection. One reason that Medtronic does a lot of studies that are not required by either payers or regulators, Peterson said, is that physicians and patients have demanded data. In many fields of medicine, however, there has been no demand for information.

A difference between devices and drugs, Maisel said, is that conducting a 4-year randomized trial of a device would essentially be holding back the advancement of technology. Our goal and our measurement, he said, should be benefit to public health. That means coming up with a total-product approach that balances the risks posed by bringing a product to market and the need to obtain information.

For many devices, a clinical study is necessary to understand the final benefit–risk ratios and identify problems that might arise in a particular patient population. But for the technology itself, most information can be obtained better at the bench than in large clinical trials. There is a distinct developmental difference between technology and pharmaceuticals.

One of the major roles for registries, Varosy said, is in confirming the validity of translation of evidence from randomized trials into real-world populations. There is still a major role for real-world evidence even after the primary questions have been addressed by randomized trials.

A committee member noted that some 510(k) devices are tools, such as diagnostic radiology equipment and ultrasonography devices that have a plethora of potential clinical applications. A manufacturer could not be expected to test each application with a clinical end point before entering the

market. Are the necessary data the same for tool claims and clinical claims? he asked. What are the problems when a device comes to market as a tool and seeks a clinical indication later?

Hall responded that a lot of European device indications are more "engineering indications." A particular device is capable of performing a particular task (for example, cut and ablate). That raises the "specific vs general indication" problem. An ablation device ablates tissue. How would one obtain an indication for ablation of heart tissue? Many of the indications are procedural, not clinical, he said.

Once a device with general indications is on the market, Kessler said, there needs to be an integrated approach that is patient-centered and clinician-centered to figure out what is known and not known about the indications and to begin real-time studies (which need not be trials). In the case of trials, Kessler noted that a lot of device studies are done with 30, 50, or 100 patients to be allowed to use the more specific indications (compared with thousands or tens of thousands for a drug).

That comes back, Alpert said, to intended use vs indication. The issue has been challenging for the agency for many years, she said, because these are tools, and we try to define more specifically where they can be used and what benefits they can provide for specific populations.

Alpert reiterated Kessler's point that in Europe there is no effectiveness requirement; the focus is on safety and performance. Effectiveness questions are related more to reimbursement than to market entry.

A question was asked about whether, from a consumer-protection perspective, it should be made clearer which indications are not supported by data. Alpert noted that device labeling already clearly states what is known about a device and its indications. Hall concurred and raised the issue of interfering with the practice of medicine. There is a well-established and legal practice regarding off-label use of products, he said, and the American Medical Association has an explicit policy statement on this.

No study and no system can provide the whole the answer, Kessler said, or fully achieve the goals of the 510(k) process, which are the continuous evolution of products and continuous assurance of safety and effectiveness. Kessler drew attention to Gardner's presentation: the hazards and failures most seen with devices are frank failures, use errors, interactions, mismatches, environmental effects, and so on.

Alpert concurred that one size does not fit all. Industry is not resisting clinical trials that are appropriate, she said. Many 510(k) devices are low-risk tools, others pose moderate risks, and for some robust clinical information is useful for understanding uses and expectations. It is a matter of what the appropriate information for specific kinds of products is and how to obtain it.

POSTMARKET DATA COLLECTION AND DATA-SHARING

A committee member noted that databases include the fields that are relevant to the researchers at the time they are developed. It is always problematic to scale up the data-collection efforts or to modify the data that are collected in order to address new areas of interest. Resnic pointed out that in the case of the National Cardiovascular Data Registry (NCDR), the problem of tracking patient-identifier information (Health Insurance Portability and Accountability Act regulations preclude direct linkage) has been successfully worked around with probabilistic matching.

Resources are already being spent by industry on postmarket studies, Resnic said, but the data are often kept behind the company firewall. There are some requirements for sharing of postmarket data with FDA, but those data may not be made available to the rest of the community to learn from (including the next manufacturer).

Alpert pointed out that all adverse events data are shared with FDA and are public. However, when a manufacturer is paying to conduct a study and collect data for the purposes of modifying one of its devices and improving it relative to a competitor product, those data are often proprietary. In initiating additional postmarket studies, she said, Medtronic generally seeks to publish the data, but there is also competitive information that may not be released.

A committee member wondered whether it would be possible to have a core set of data in a precompetitive, publicly accessible space with appendixes of private data for specific studies.

UNIQUE DEVICE IDENTIFIERS

Developing standardized nomenclature for devices and a unique device identification (UDI) system will aid the ability to track devices among institutions, Peterson said. The data are already stored electronically somewhere, but they are not easily shared.

Varosy agreed that having a UDI system would make the process of documentation easier, more feasible, and translatable across systems.

Although UDIs are important, Alpert said, several other things also have to happen, including building electronic health records that have the capacity to collect the information. The big impediment is to put a particular identifier on a product rather than whether it is the hospitals, physicians' offices, or clinics that capture and use the data; right now, they do not do that, she said. Hospitals use the bar code on the device for supply-chain and billing purposes, but it is not used for anything else.

PARTICIPATION IN SURVEILLANCE SYSTEMS

Perhaps the biggest factor in the Department of Veterans Affairs (VA) system that has facilitated reporting, Varosy said, is having a uniform electronic health record as a foundation. Moreover, although a lot of health information technology is "inflicted" on doctors by developers, a key aspect of the VA CART-CL was that it was developed directly with clinicians to be a clinically useful tool.

The question is one of resource allocation, Resnic added. From an information-systems perspective, the reporting process is similar, whether the issue is the number of failures of tire tread in a manufacturing plant or unexpected falls out of a new hospital bed that is cleared through the 510(k) process. The question is whether it is cost-effective for health-care institutions to develop registries that can capture high-granularity data and determine a denominator.

Alpert noted that for one of the two registries she spoke of, Medtronic pays physicians. The payment is minimal, but it helps to ensure that information is entered into the database, she said.

The acceptance of a data-collection tool depends somewhat on the burden of collecting the data, Resnic said. There has to be some consensus early on that the data are worth collecting for the purposes of those participating institutions. There are multiple reasons why institutions participate in the NCDR, he said, one of which is to receive data from the NCDR for performance and quality benchmarking. Another incentive, Resnic said, is that many payer organizations, including CMS, view participation in the NCDR as evidence of a high-quality organization that is monitoring outcomes.

Peterson added that once a product is on the market, payers can require participation in registries as a condition of payment.

Resnic noted that every new data element that one tries to add to an existing system is going to come at some cost to the population of participants, and expansion has to be a collaborative development effort.

Hall pointed out that many 510(k) products are home-use products. The surveillance systems that have been discussed work much better in a hospital setting. One of the challenges is to facilitate data collection outside the hospital procedure-based structure.

RISK COMMUNICATION

A committee member asked, With continuous accrual of information in the postmarket period, at what point do actions need to be taken to communicate the information?

Resnic stressed again that in monitoring of datasets, the signals identi-

fied are really only hypothesis-generating. In the conduct of the DELTA studies in collaboration with FDA, Resnic said, results were shared with the FDA postmarket staff, who shared them with the premarket staff who looked at the internal FDA datasets, including engineering data and preapproval and postapproval studies, to see whether there were concordant signals. Such hypothesis-generating signals must be vetted through a relatively rigorous mechanism to avoid undue alarm in patient communities. Risk communication is a science unto itself, he said, and an initiative on risk communication is under way in FDA.

Maisel said that communication should be guided by ethical principles of what a patient would want to know and would need to know to make an educated decision about care.

Hall added that as a result of modern information technologies, there are no longer separate communication pathways for physicians and patients. The communication challenge is incredibly difficult.

REFERENCE

GHTF (Global Harmonization Task Force). 2005. Essential principles of safety and performance of medical devices (SG1-N41R9:2005). Study Group 1 of the Global Harmonization Task Force. http://www.ghtf.org/documents/sg1/sg1n41r92005.pdf (accessed July 2, 2010)

5

Other Select Topics

The committee used this workshop, as well as the June 14–15, 2010 workshop, as opportunities to hear about a variety of issues related to the medical device regulatory lifecycle. This chapter summarizes three separate presentations. The first presentation is about software in medical devices. The second is on the Food and Drug Administration's (FDA) use of evidence in premarket approval (PMA) process. The last presentation is an example of industry concerns about transparency of, and delays in, FDA decision-making within the 510(k) process.

TRUSTWORTHY MEDICAL-DEVICE SOFTWARE

Without software, many medical treatments could not exist, said Kevin Fu, assistant professor in the Department of Computer Science of the University of Massachusetts Amherst. The question is not whether devices should use software but rather how the complexities of software and its risks can be better understood. Fu presented an overview of a report that he was commissioned to prepare for the committee to summarize the role of trustworthy software in the safety and effectiveness of medical devices.[1]

"Software trustworthiness" is a system property that measures how well a software system meets operating requirements allowing stakeholders (such as patients, health-care professionals, and service providers) to trust the operation of the system. Software trustworthiness is closely tied to safety and effectiveness, and diminished trustworthiness can lead to lack

[1]The complete commissioned paper is available as Appendix D.

of safety, effectiveness, usability, reliability, dependability, security, privacy, availability, and maintainability, Fu said.

Safety and Effectiveness

There can be overconfidence in the function of software, Fu said. Complacency can be based on the belief that if the software appears to function, nothing can go wrong; this is not always the case. Fu cited one example from the late 1980s involving the Therac-25, one of the first linear accelerators to use software aggressively in the control of radiation treatments. After reports from health-care professionals of injuries and deaths from machine malfunctions (which resulted in radiation overdoses), the manufacturer investigated and reported that the machine could not possibly overtreat a patient (Leveson and Turner, 1993).

Since then, the number of devices using software and the number of devices recalled for software-related issues have been increasing. Fu reported that 6% of all device recalls issued by the Food and Drug Administration (FDA) from 1983 to 1997 cited software as the reason. The proportion nearly doubled from 1999 to 2005: 11.3% of device recalls were attributed to software. In 1983–1997, 24% of recalled devices relied on software in some way, and this increased to 49% during 1999–2005. In 2006, it was reported that over half the medical devices on the US market involved software in their function. In 2002–2010, there were more than 537 recalls of devices that used software, which affected over 1.5 million devices being used in the United States.

Software in a device is different from the hardware for two reasons, Fu asserted. First, software is discrete, rather than continuous. For example, there would be little concern if a manufacturer of 1-inch nails produced a product ranging from 0.9999 inch to 1.0001 inch. That small error is usually tolerable. However a single error in a computer system, changing a 20-mL entry for an infusion pump to a 200-mL entry, can have potentially catastrophic consequences. There is generally no analogous notion of a safety margin for software. Second, software is extremely difficult to test for every possible complication.

Fu noted that software itself can constitute a device itself, for example, an electronic health record. Electronic health records, if designed correctly, could reduce errors substantially, especially errors of patient misidentification. But electronic health records will need to have very strong integrity guarantees and strong security and privacy, and there is an issue of interoperability among hospitals and systems. System complexities involving the collation of vast amounts of information could introduce risks. When asked whether a paper medical record would be a reasonable predicate, especially in considering security and privacy, Fu stated that software behaves differ-

ently from paper. Paper is static, but an electronic health record is dynamic, more like a "living record." Fu asserted that there need to be different standards for electronic health records—not necessarily higher, but appropriate to the technology and the situation.

Mitigating Software Risks

Many of the risks associated with software design are preventable, Fu said. A workshop report from the Networking and Information Technology Research and Development Program found that the risks associated with software use are not peculiar to medical devices. Many of the standards and practices used in the development of software for other critical systems (such as avionics and nuclear systems) could be used to ensure confidence in medical devices, but they appear to be ignored by developers (NITRD, 2009). According to the report, "perhaps the most striking [difference] is the almost complete lack of regard, in the medical-device software domain, for the specification of requirements."

Fu suggested that the committee consider recommending that device software developers follow good systems-engineering practices. Systems engineering, he said, is a much more encompassing technique than simply testing software in isolation. A systems approach could address, for example, whether a ventilator for oxygen works when the software is integrated in an ambulance.

Implementation Errors

Although implementation errors are often the subject of news stories, they are not actually the primary source of problems with medical devices, Fu said. He used an example of an implementation error that involved a Baxter infusion pump. Underdosing of a patient with three drugs led to increased intracranial pressure and then brain death. A message appeared on the device screen indicating a "buffer overflow," and the pump shut down. In simple terms, a buffer overflow occurs when the buffer has too little memory space to hold the information that the program is attempting to place in it. A problem with buffer-overflow errors is that they are difficult to reproduce, especially during service. In this particular case, the manufacturer was eventually able to reproduce the problem outside the clinical setting and found that a pump-software upgrade had resulted in a slight coding implementation error that caused the device to fail (and ultimately led to the patient's death).

Human Factors

How do human factors come into play in errors associated with user interfaces? Infusion pump–user interfaces and software are used effectively and safely every day in medical practice. However, infusion pumps in general have been linked to over 500 deaths and over 56,000 adverse-event reports. Fu discussed an April 2010 *New York Time*s report on infusion-pump problems, which stated that 710 patient deaths were linked to a health-care provider's entering an incorrect dosage or to a malfunction in the software (Meier, 2010).

Implantable pumps are used to treat for some diseases. Computer control systems for them are used by health-care professionals to set dosage. In the user interface, there are spaces to enter dosage, bolus size, and the duration (hours, minutes, and seconds) for which to administer the bolus.

Fu cited one example: an adverse-event resulting in a patient death reported in Manufacturer and User Facility Device Experience (MAUDE). A bolus was given in 20 min rather than the intended 20 h (that is, at 60 times the intended rate). The patient who had the implanted drug pump lost consciousness while driving because of the overdose, was involved in a collision, and died. The FDA recall notice stated that the software did not provide a label for the hours, minutes, and seconds data-entry fields and that the new software has such labeling.

Fu also cited the more recent radiation-therapy accidents involving linear accelerators. Nothing on the machine, Fu said, warns technicians that they may have entered an inappropriate dose, and there is no way to know that they have entered the right data. It was reported that failures in computer software led a technologist to think that they had set the correct radiation exposure when the machine was actually administering a much larger amount of radiation, which led to several injuries and deaths.

Better analysis of human factors and how they interact with software could help to prevent injury and death, Fu said.

Software Maintenance

Software users are familiar with the dialogue boxes that appear on a computer screen and advise that a software update is available and should be downloaded and installed. Consumers of commercial, off-the-shelf software are in effect treated as beta testers. Traditionally, for noncritical systems, developers seem to believe that if there is a "bug" in the system they can just send out a patch later.

Problems with computer maintenance can have far-reaching consequences that affect the availability of care in hospitals and other infrastructure. In one example from April 2010, health-information technology

devices were globally rendered unavailable by a single point of failure, a software update gone awry. Many diagnostic machines use a particular antivirus product; to protect them from the latest computer viruses, the product automatically updates software. In one of the updates, however, one critical core component of the Windows operating system was classified as a virus and quarantined by the antivirus software. Computers later entered an endless loop of reboot cycles. Numerous computers and hospitals were affected. One-third of the hospitals in Rhode Island, for example, were forced to postpone elective surgery and to stop treating nontrauma patients in their emergency rooms. At Upstate University Hospital in New York, 2,500 of 6,000 computers were affected.

Fu pointed out that although technical difficulties with software are not unexpected (and might even be amusing) when someone is giving a presentation at a meeting, the same difficulties can occur during a mammography or radiation treatment. In one case, a magnetic resonance imaging machine entered an endless reboot cycle while a patient was in the device. The patient experienced cardiac arrest and died, but the health-care professionals did not notice, because they were focusing on determining why the computer was rebooting.

User factors that are ignored in noncritical systems are important in injury and death in health-care–related systems. Discussions on technology support-group Web sites suggest that end users are often helpless when it comes to dealing with these systems. Fu cited an online forum discussion in which a user was seeking information on how to downgrade the Windows operating system to a prior version (for example, from version SP3 to SP2). The user was setting up an electronic picture archiving and communication system (PACS) for recording medical images, such as x-ray pictures. The PACS was compatible only with the earlier version of Windows (SP2), but the new computers that he ordered to be used with it came with SP3 preinstalled. He had already invested substantially in products to be used with the PACS, many of which also could not be used with the SP3 operating system. Later in the chat, he received various advice on how to downgrade. That may solve the immediate problem of computer compatibility with the PACS, but it creates serious new problems. Specifically, Microsoft Windows ended its support of SP2 and is no longer providing security updates. Thus, for the PACS to work, the system is being run on new computers with obsolete, unsupported operating systems that are subject to security vulnerabilities.

Users share responsibility for keeping their software up to date, Fu said. A manufacturer might produce applications that run on commodity, off-the-shelf software, but then users (hospital health-care professionals) have to maintain it, and sometimes they have conflicting requirements. Fu said that shared responsibility results in no responsibility, especially in the case of software. A single platform may have components from many

manufacturers, and there can be many kinds of failures. As a result, it can be hard to assign responsibility to any kind of failure that might happen in a computer system.

Problems on the Horizon

The FDA Center for Devices and Radiological Health (CDRH) director, Jeffrey Shuren, has stated that health information technology (HIT) software is considered a medical device. FDA has largely refrained from enforcing regulatory requirements for HIT devices, but the agency has received 260 reports of HIT-related malfunctions, and the reports noted 44 injuries and six deaths.

Computer viruses are a continuing concern. Fu noted that in May 2010, Roger Baker, chief information officer for the US Department of Veterans Affairs (VA), testified before a House of Representatives subcommittee that over 122 medical devices in the VA network had been compromised by malware during the preceding 14 months.

A computer virus does not discriminate between a home computer and a piece of radiology equipment, Fu said. All are at risk. But what about intentional malfunctions in software beyond viruses? Fu asked. He reminded participants of how the security of drug packaging in the United States was improved substantially after deaths from cyanide-laced Tylenol in 1982. As a result of that malicious act, there is now regulatory guidance on secure packaging of medicines. Under 21 CFR 211.132, FDA has the authority to establish a uniform national requirement for tamper-evident packaging that will improve the security of over-the-counter drug packaging. Perhaps security needs to be considered more carefully in looking at the safety and efficacy of medical-device software, Fu said.

In an effort to improve patient safety and device security, Fu and colleagues analyzed an implantable cardiac defibrillator and, through reverse engineering, were able to develop a software radio with which they could wirelessly induce the device to cause ventricular fibrillation (Halperin et al., 2008). Fu said that that was possible because of a problem with requirement specification: an unauthorized person should not have been able to manipulate the device.

Together, emerging technologies, HIT software, wireless capabilities, interoperability of devices, patient mobility (which implies the use of devices outside the health-care setting), and the Internet lead to substantial security and privacy risks, Fu said.

510(k) Substantial Equivalence: What Is a Predicate Device for Medical-Device Software?

Fu reiterated a comment made by former CDRH Director David Feigal, at the June 14-15, 2010, workshop, who wondered what the first predicate for software had been. In considering substantial equivalence, Fu said, is a hardware implementation the predicate for the software? Hardware and software are very different entities, he stressed, with very different risks. He suggested that any kind of software device that cites hardware as its predicate deserves careful assessment from a risk perspective. If there were more meaningful requirement specifications, he noted, it would be easier to determine substantial equivalence.

Summary

Software in medical devices, Fu summarized,

- Breeds overconfidence ("the computer can't be wrong").
- Is not thoroughly testable (devices do not operate in isolation, and not all possible interactions can be tested).
- Is flooding into medical devices at an increasing rate.
- Is not equivalent to hardware from a risk perspective.

Many of the risks associated with medical-device software could be mitigated with known technology, Fu said. The software-engineering, systems-engineering, and safety-engineering communities have many techniques that could address problems that have led to software-associated device adverse events.

Fu highlighted several subjects for the committee to consider. They included the idea that device manufacturers need to be incentivized to adopt modern software-engineering and systems-engineering technologies—from static analysis to programming languages that are more easily able to integrate with requirement specifications. Better analysis of human factors that come into play in the use of device software could help to prevent injury and death. Attention should be paid to developing a safety net for security and privacy.

Outcome measures are needed, Fu said, and manufacturers should be able to state the outcome measures for the safe and effective use of medical devices more openly.

He also noted that statistics are needed and that most of the available data are on failures of device software and far fewer on successes. The software-engineering community has techniques that are known to work well in other critical systems, but the medical-device community has not, in general, used them.

There is a call for more open research and open test beds, Fu said. Researchers note that it is difficult to contribute their software technology to the medical-device community because of the proprietary nature of device design. If there were more open test beds, there would be more innovation, he suggested.

Fu reiterated that shared responsibility for a product has a tendency to mean that no party takes responsibility. There needs to be a single authority that is responsible for the safety and effectiveness of the software in a device.

Finally, he noted that research in other fields may be useful in considering an approach for devices. In avionics, for example, the National Aeronautics and Space Administration maintains a safety culture that includes technical and managerial approaches for mitigating the complex risk that arises when taking something in isolation and connecting it with other pieces. The avionics community is also using databases to try to understand successes, failures, and near-misses.

STRENGTH OF STUDY EVIDENCE EXAMINED BY THE FOOD AND DRUG ADMINISTRATION IN PREMARKET APPROVAL OF CARDIOVASCULAR DEVICES

Rita Redberg, professor of medicine at the University of California, San Francisco Medical Center and editor of *Archives of Internal Medicine,* presented an overview of a study of the strength of study evidence examined by FDA in premarket approval (PMA) of cardiovascular devices (Dhruva et al., 2009), and proposes some opportunities for improvement in the use of clinical evidence.

Cardiovascular devices are increasing in complexity and are an important part of the medical economy, with over 1 million stents, 350,000 pacemakers, and 140,000 implantable cardioverter–defibrillators (ICDs) implanted in 2008. Redberg noted that patients are increasingly exposed to direct-to-consumer advertising of specific devices with claims about how the devices can improve quality of life. The claims are not always consistent with the medical literature.

Background and Objectives

Class III devices are defined by FDA as devices that support or sustain human life, that are of substantial importance in preventing impairment of human health, or that present an unreasonable risk of illness or injury. The PMA process is the most stringent type of device-marketing application required by FDA and is required for most class III devices. Of the approximate 8,000 devices that are marketed yearly, 50–80 of these devices go through

the PMA process. Examples of class III cardiovascular devices are stents, heart valves, and ICDs.

Redberg noted that many high-risk devices are not going through the PMA process. A 2009 Government Accountability Office (GAO) report, *FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved Through the Most Stringent Premarket Review Process*, examined all the high-risk device approvals from 2003 through 2007 (GAO, 2009). GAO found that 78% of the 217 original and 85% of the 784 supplemental PMA submissions for class III devices were approved through the PMA process . However, GAO also found that more class III, or high-risk, devices were given a 510(k) exemption during that period than went through a PMA process. That means, Redberg said, that clinical-trial data on many high-risk devices are not being collected.

For the study, Redberg and colleagues reviewed the summary of safety and effectiveness data (SSED) on 78 high-risk cardiovascular devices that received PMA from January 2000 to December 2007. As would be the case in analyzing the quality of a clinical trial, study data in each SSED for devices that received PMA were assessed for randomization, blinding, primary end points, active controls, analysis, and followup time. Written by the device sponsor and reviewed by FDA, an SSED is "intended to present a reasoned, objective, and balanced critique of the scientific evidence which served as the basis of the decision to approve or deny the PMA" (FDA, 2010). After device approval, the SSED is made publicly available by FDA with the device's approval order and labeling.

Findings

The 78 devices that received PMA in 2000–2007 had undergone a total of 123 clinical studies (mean, 1.6 studies per device; range, of 1–5 studies per device). Most of the approvals (51 of 78 PMAs, or 65%) were based on a single study. Of the 123 studies, 33 (27%) were randomized studies, and 17 (14%) were blinded studies (either single-blind or double-blind).

The 123 studies encompassed 213 primary end points. Seventeen (14%) of the studies did not list a primary end point. For the rest, the number of primary end points ranged from 1 to 10. (Redberg acknowledged that normally one would have a single primary end point, but they recorded whatever was noted in the SSED, and many of the device studies listed more than one primary end point.)

The number of patients enrolled per study ranged from 23 to 1,548 and averaged 308. As is typical in cardiology studies, Redberg said, the mean age was about 62 years (only 87 of the 123 studies reported age). A little more than two-thirds of the patients were male, which is also fairly typical for cardiology studies (80 of the 123 studies reported the sex of participants).

Of those participants that reported race and ethnicity, 87% were white, 6% black, 5% Hispanic, and 3% other minorities (only 11 of the 123 studies reported race).

Of the 83 studies that listed location, 43 had all US sites, 22 had no US sites (that is, the PMA was achieved without any studies conducted in the United States), and the rest had a mixture of US and other sites.

Followup time varied by device. On the average, stents were approved on the basis of a 6-month followup, and implantable electrophysiology devices were approved on the basis of a 3-month followup. The longest followup periods were 1 year, for endovascular grafts and intracardiac devices.

For about half the primary end points, patients who received the intervention were compared with controls who did not. However, in about one-third of the groups that were randomized, the controls were not enrolled as part of the study; instead, those who received the intervention were compared with "retrospective controls" from a previous study. The other half of the studies were single-arm studies (they had no randomization or control group); objective performance criteria were established by FDA in conjunction with the sponsor. Redberg pointed out that 187 of the 213 primary end points (88%) were surrogate end points, and she noted that there is concern about nonvalidated surrogate end points. It is not clear how well a surrogate end point represents the clinical end point. There is a difference, for example, between an angiographic finding that a stent is open or has closed and a clinical end point of presence or absence of chest pain, and patients who have open stents may still experience chest pain.

Another measure that Redberg analyzed was the number of patients who were enrolled vs the number included in the data analysis. For 122 of the 213 primary end points, there was a difference between the number of patients enrolled and the number analyzed. For all 213 primary end points, over 10,000 patients (27%) were listed as enrolled in the studies but not included in the SSEDs.

Redberg and colleagues could not interpret 15% of the primary end points, because no target goals for device performance were stated in the SSEDs. In several cases, a stated target end point was not met, but the device received PMA. One example cited by Redberg was the NaviStar Thermocool ablation catheter, which has a target stated in the SSED of 50% chronic success but was approved on the basis of achieving 47%.

Redberg noted several limitations of the study. Data were abstracted from the SSEDs that were available on the FDA Web site. Although an SSED is considered to be a summary of all the data on which FDA based its decision, the agency often has additional company-confidential documents that it does not post on its Web site. And it is possible that FDA required followup studies as a condition of approval, but such studies were not available to be included in the analysis.

Food and Drug Administration Response to the Study

After the publication of the study in *JAMA,* FDA issued a response stating, Redberg said, that a single pivotal study is adequate for approval of a device; that a randomized, double-blind, placebo-controlled trial is not always the best way to look at device data; and that the length of a study is not related to the quality of the data that it produces. Device trials, the agency said, must incorporate the practical realities of devices, which are different from drugs. According to Redberg, FDA also disagreed with the study conclusions regarding clinical vs surrogate end points, and she said that FDA noted that an SSED is not a "surrogate" for the full confidential data review (which is not publicly posted on the Web site).

Authors' Response to Food and Drug Administration Criticisms

In response to FDA, Redberg said that the study authors feel that a single clinical study is often not adequate and that two studies are preferred for high-risk cardiovascular devices. She added that randomized, controlled, blinded trials with complete followup provide the highest-quality data. Those devices are often implanted permanently, she stressed, and removal entails substantial risk. She expressed concern about the use of retrospective controls, noting the opportunity for bias when one can pick and choose to form a control group. With regard to blinded device trials, Redberg noted that sham controls are used in surgical trials. It is important to discern whether the effects are from the device's working as expected, she said, or from the invasive procedure of implanting the device. She also spoke of the need for clinical end points that directly measure how a patient feels, functions, or survives.

With regard to the additional confidential data that may be part of a PMA application, Redberg said that she was able to request access to the confidential files for a number of the PMAs that were part of the study, and her review of the additional data did not change the overall findings of the review of the SSED data described in the paper.

Opportunities for Improvement

On the basis of her study results, Redberg offered several recommendations for improving the FDA PMA process for high-risk devices:

- Require at least one randomized and blinded study for each device. Ideally, the clinical-trial population should be representative of the intended patients with respect to age, sex, race, and comorbidity; and most of the clinical sites should be in the United States.

- Require longer followup time and the use of clinical (not surrogate) end points.
- Require an intent-to-treat analysis of all enrolled patients. All patients enrolled should be reported in the SSED, and controls should be active, not retrospective. Adverse outcomes and poor efficacy may be missed if not all data on all enrolled patients are analyzed.
- Provide more public access to the raw data examined by FDA in an easy-to-navigate fashion.
- Make postmarketing studies available on the clinicaltrials.gov Web site.

In closing, Redberg expressed support for the recent FDA Transparency Initiative and the agency's plans to improve the quality of clinical trials and for the IOM committee's current assessment of the 510(k) process.

CONCERNS REGARDING CONSISTENCY OF DECISION MAKING IN THE 510(k) CLEARANCE PROCESS

Robert E. Fischell, founder and chief technology officer of Neuralieve Inc., described his company's recent experience in working to bring a product to market through the 510(k) clearance process as an example of industry concerns with a perceived lack of transparency and consistency in decision making. His company's 510(k) submission for a device which uses transcranial magnetic stimulation to relieve migraine headaches was rejected by FDA. He noted that the agency had previously cleared, via the de novo 510(k) clearance process, another company's device which uses similar technology to treat depression.

Fischell noted that FDA did not respond to his company's de novo 510(k) submission within the 60-day target and said that there is no recourse for the device sponsors when such response deadlines are not adhered to. At one point, company officials met with a FDA branch chief, who advised, Fischell said, that there should be a panel meeting. According to Fischell, a panel meeting for a 510(k) submission is unprecedented. Company officials were later told that there would be no panel meeting, and that the company should apply for a PMA with additional data to address concerns about efficacy and safety.

Fischell expressed concern that there is no appeal process that works in FDA to overrule the lowest level of review. He expressed further concerns about how FDA reviewers were assigned to particular devices and if their experience and training were appropriate for the device in question.

REFERENCES

Dhruva, S. S., L. A. Bero, and R. F. Redberg. 2009. Strength of study evidence examined by the FDA in premarket approval of cardiovascular devices. *JAMA* 302(24):2679-2685.

Dodick, D. W., C. T. Schembri, M. Helmuth, and S. K. Aurora. 2010. Transcranial magnetic stimulation for migraine: a safety review. *Headache* 50(7):1153-1163.

FDA (Food and Drug Administration). 2010. *PMA Application Contents: Summary of Safety and Effectiveness Data (§814.44)*. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/PremarketSubmissions/PremarketApprovalPMA/ucm050289.htm (accessed October 20, 2010).

GAO (General Accountability Office). 2009. *FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved Through the Most Stringent Premarket Review Process*. http://www.gao.gov/new.items/d09190.pdf (accessed August 16, 2020).

Halperin, D., T. S. Heydt-Benjamin, B. Ransford, S. S. Clark, B. Defend, W. Morgan, K. Fu, T. Kohno, and W. H. Maisel. 2008. Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses. In the proceedings of the *2008 IEEE Symposium on Security and Privacy*.

Leveson, N. G., and C. S. Turner. 1993. An Investigation of the Therac-25 Accidents. *Computer* 26:18-41.

Meier, B. 2010. FDA Steps Up Oversight of Infusion Pumps. *The New York Times*, April 23, B1. http://www.nytimes.com/2010/04/24/business/24pump.html (accessed August 11, 2010).

NITRD (Networking and Information Technology Research and Development). 2009. *High-Confidence Medical Devices: Cyber-Physical Systems for 21st Century Health Care*. http://www.nitrd.gov/About/MedDevice-FINAL1-web.pdf (accessed August 11, 2010).

A

Workshop Agenda

Wednesday, July 28, 2010
Room 100
Keck Center of the National Academies
500 Fifth Street, NW
Washington, DC

8:30 AM	**Welcome and Opening Remarks** David Challoner, Chair, IOM Committee on the Public Health Effectiveness of the FDA 510(k) Clearance Process
8:40	**Trustworthy Medical Device Software** Kevin Fu, University of Massachusetts Amherst
9:20	**Strength of Study Evidence Examined by the FDA in Premarket Approval of Cardiovascular Devices** Rita Redberg, Professor of Medicine UCSF Medical Center and Editor, Archives of Internal Medicine
10:00	**Issues with the Present FDA on the Matter of FDA 510(k) Clearance** Robert E. Fischell, Founder and Chief Technology Officer, Neuralieve Inc.
10:20	**Break**
	FDA Postmarket Surveillance
10:30	**Monitoring Device Safety: CDRH's Current System and Vision for the Future** Susan Gardner, Director of the Center for Devices and Radiological Health's (CDRH's) Office of Surveillance and Biometrics

11:10 **Premarket Notification: Analysis of FDA Recall Data**
William H. Maisel, Director, Medical Device Safety Institute

11:50 **FDA Recall Data Study**
Ralph Hall, Distinguished Visiting Professor of Law, University of Minnesota Law School

12:30 PM **Lunch**

Non-FDA Sources of Adverse Event Data

1:30 **The National Cardiovascular Data Registries: Opportunities and Challenges in Postmarket Surveillance**
Frederick A. Masoudi, MD, MSPH, Associate Professor of Medicine, Denver Health Medical Center & University of Colorado; Senior Medical Officer, National Cardiovascular Data Registries

2:00 **The VA-CART Program: Integration of Real-Time Data Collection into the Process of Clinical Care**
Paul D. Varosy, Director of Cardiac Electrophysiology, VA Eastern Colorado Health Care System, Project Director, CART-EP, and Assistant Professor of Medicine, University of Colorado Denver

2:30 **The Centers for Education and Research on Therapeutics (CERTs) Program**
Eric D. Peterson, Fred Cobb MD Distinguished Professor of Medicine and Associate Director of the Duke University Medical Center, and Director CV Research of the Duke Clinical Research Institute

3:00 **Automated Postmarket Safety Surveillance: The DELTA Surveillance Project**
Frederic S. Resnic, Director, Cardiac Catheterization Laboratory Brigham and Women's Hospital, and Assistant Professor of Medicine, Harvard Medical School

3:30 **Break**

3:40 **Postmarket Surveillance of Medical Devices**
Panel Discussion
Moderated by Lazar Greenfield, committee member
Panelists:

- Workshop speakers: Susan Gardner, Ralph Hall, William H. Maisel, Frederick A. Masoudi, Eric D. Peterson, Frederic S. Resnic, and Paul D. Varosy
- Susan Alpert, Senior Vice President–Chief Regulatory Officer, Medtronic Inc.
- Larry Kessler, Professor and Chair, Department of Health Services, School of Public Health, University of Washington

5:30 **Adjourn**

B

Biographic Information on Invited Speakers, Panelists, and Authors of Commissioned Papers

Susan Alpert, PhD, MD, joined Medtronic in July 2003 as vice president of regulatory affairs and compliance. She is now senior vice president and chief regulatory officer and is responsible for all Medtronic global regulatory efforts. Before joining Medtronic, she served C.R. Bard Inc., as vice president of regulatory sciences. She also previously worked at the Food and Drug Administration (FDA), where she held a variety of positions in the centers dealing with drugs, devices and radiologic health, and foods, including 6 years as the director of the Office of Device Evaluation. She is a microbiologist and pediatrician with a specialty in infectious diseases and has practical experience in laboratory research and clinical trials. Dr. Alpert has served on the board of the Food and Drug Law Institute. She serves on the board of advisers for the Medical Technology Leadership Forum, an educational organization focused on policy-makers, the general public, and the mass media regarding critical issues affecting the development and adoption of advanced medical technology. She also serves on the board of Women Business Leaders, an organization of women leaders in the health-care sector, and on the board of the Minnesota International Center. She is a past chair of the Regulatory Affairs Professional Society and a Fellow in that society. She serves on the Executive Committee of the Clinical Trials Transformation Initiative, one of the public–private partnerships working with FDA to streamline the development of medical products. Dr. Alpert received her undergraduate degree at Barnard College and holds a master's degree and a PhD in biomedical sciences from New York University. She received her medical degree from the University of Miami (Florida) and completed her

clinical training at Montefiore Medical Center in the Bronx, New York, and at Children's National Medical Center in Washington, DC.

Robert E. Fischell, ScD, was employed at the Johns Hopkins University Applied Physics Laboratory full-time for 25 years and part-time for an additional 13 years. At Johns Hopkins, Dr. Fischell was the chief engineer of the Space Department, where he worked on more than 50 spacecraft. His interests at Johns Hopkins then turned to the invention of new medical devices, such as pacemakers and implantable heart defibrillators. Starting in 1969, Dr. Fischell began the formation of 14 private companies that licensed his patents on medical devices. The companies included Pacesetter Systems Inc. (now called St. Jude Medical), IsoStent Inc., NeuroPace Inc., Neuralieve Inc., Angel Medical Systems Inc., and Svelte Medical Systems Inc. Dr. Fischell has over 150 issued US and foreign patents for medical devices. Dr. Fischell is a trustee of the University of Maryland College Park Foundation and a member of the Board of Visitors for the College of Engineering and the College of Computer, Mathematical, and Physical Sciences and the University of Maryland School of Medicine. Dr. Fischell's was elected to the National Academy of Engineering in 1989. He has received numerous awards for his contributions including the Inventor of the Year for the USA in 1984; Distinguished Physics Alumnus Award of the University of Maryland; the 2004 *Discover* magazine award for Technology for Humanity; the 2005 TED award; the 2007 Master Inventor award from the Applied Physics Laboratory, and the Woodrow Wilson Prize for Public Service from the Woodrow Wilson Society for Scholars, as well as several other medals for his accomplishments in science, engineering, and innovation. In 2005, he gave $30 million to create and fund the Fischell Department of Bioengineering in the Clark School of Engineering. In that same year, the University of Maryland created the Robert E. Fischell Institute for Medical Devices to further the pioneering work that Dr. Fischell has created. Dr. Fischell received his BSME from Duke University and MS and ScD degrees from the University of Maryland.

Kevin Fu, PhD, is an assistant professor in the Department of Computer Science of the University of Massachusetts Amherst. Prof. Fu investigates how to achieve trustworthy computing for embedded devices that must withstand both unintentional interference and determined, malicious intent. Prof. Fu's research contributions range from the design and implementation of cryptographic systems to the security-risk analysis of computer systems, such as implantable cardiac defibrillators, automated software updates, contactless no-swipe credit cards, and Web site log-in systems. He is an Alfred P. Sloan Research Fellow, MIT Technology Review TR35 Innovator of the Year, and recipient of the National Science Foundation CAREER award.

His research appears in computer-science conferences and medical journals and has been featured in the mass media, such as the *New York Times*, the *Wall Street Journal*, and various news programs. He served on numerous program committees of leading conferences in secure systems and has given dozens of invited talks worldwide to industry, government, and academe. He is a member of the Institute of Electrical and Electronic Engineers, the Association for Computing Machinery, and USENIX. Prof. Fu leads the University of Massachusetts Amherst Security and Privacy Research (SPQR) Laboratory. He serves as codirector of the Medical Device Security Center and director of the RFID Consortium on Security and Privacy. Prof. Fu is a frequent visiting faculty member at Microsoft Research and Beth Israel Deaconess Medical Center. He received three degrees from the Massachusetts Institute of Technology, including a PhD in electrical engineering and computer science.

Susan Gardner, PhD, is the director of the Office of Surveillance and Biometrics (OSB) in the Center for Devices and Radiological Health (CDRH) of the Food and Drug Administration (FDA). She joined FDA in 1995 as deputy director of OSB and became director in 2002. The responsibilities of OSB include providing statistical support to CDRH, development of innovative statistical and epidemiologic techniques, design and oversight of postmarket studies, postmarket problem signal detection through monitoring of adverse events databases and other data sources, interpretation of the Medical Device Reporting (MDR) regulation, and CDRH coordination of data-management standards. Before joining FDA, Dr. Gardner was the associate director of health studies at Westat, a social-science research firm. After graduating from the Johns Hopkins School of Nursing, Dr. Gardner attended Boston University, where she received a BA in sociology. She received her PhD in medical sociology from Catholic University of America.

Ralph F. Hall, JD, serves as Distinguished Visiting Professor of Law at the University of Minnesota Law School. He is also counsel to the Indianapolis, Indiana, law firm of Baker & Daniels, where he counsels clients in drug and medical-device regulation. He serves as CEO of MR3 Medical LLC, a startup medical-device company. Before his association with the university, Prof. Hall served in various capacities with Guidant Corporation, including senior vice president and deputy general counsel for litigation and compliance, general counsel of the Cardiac Rhythm Management group, special counsel to the Guidant Board of Directors Compliance Committee, and counsel to the Guidant chief compliance officer. Before joining Guidant, he was with Eli Lilly, including serving as the head of Lilly's worldwide environmental law group. Prof. Hall received his BA from Indiana University in 1974 and his JD from the University of Michigan, where he was a

Weymouth Kirkland Scholar. Prof. Hall's interests include Food and Drug Administration regulation, negotiations and alternative dispute resolution, intellectual-asset management, and the interface between corporate practice and the academic world.

Larry Kessler, ScD, was appointed professor and chair of the Department of Health Services at the University of Washington (UW) School of Public Health in January 2009. The department also contains four centers; three are concerned with different aspects of public-health research, and the fourth is the Northwest Center for Public Health Practice. Before joining the faculty at UW, he spent 30 years working for the federal government, first at the National Institute of Mental Health, then at the National Cancer Institute, and most recently at the Food and Drug Administration (FDA) Center for Devices and Radiological Health (CDRH). In September 2002, Dr. Kessler was appointed director of the Office of Science and Technology of CDRH. In that position, he directed the efforts of the laboratories of CDRH and the Standards Coordination Program. The office became the Office of Science and Engineering Laboratories (OSEL) in a reorganization effort designed to improve integration into the function and mission of CDRH. In June 1995, he first joined CDRH as the director of the Office of Surveillance and Biometrics. From 1996 through 2001, he served as chair of Study Group 2 of the Global Harmonization Task Force (GHTF), concentrating on postmarket vigilance and surveillance. In 2007, Dr. Kessler became chair of the GHTF for 1-1/2 years. In the period September 2001–August 2002, Dr. Kessler took a position as a visiting scientist at the Fred Hutchinson Cancer Research Center. From 1984 to June 1995, Dr. Kessler served as chief of the Applied Research Branch at the National Cancer Institute (NCI). Dr. Kessler has published over 100 peer-reviewed journal articles and numerous book chapters and government reports. His research has concentrated on applications of quantitative methods and health-services research to problems in surveillance and public health. Dr. Kessler obtained his degree in operations research from the Johns Hopkins School of Public Health in 1978 and received his ScD in the same year.

William H. Maisel, MD, MPH, is director of the Medical Device Safety Institute, a nonprofit, industry-independent organization dedicated to improving the safety of medical devices. He is also a practicing cardiologist at Beth Israel Deaconess Medical Center and associate professor of medicine at Harvard Medical School, both in Boston, MA. He has served as a consultant to the Food and Drug Administration (FDA) Center for Devices and Radiological Health since 2003 and previously chaired FDA's Post Market and Heart Device Advisory Panels. Dr. Maisel's research interests involve the

safe and effective use of medical devices, and he has published extensively on medical-device safety and innovation and on consumer protection.

Frederick A. Masoudi, MD, MSPH, is a practicing cardiologist and director of echocardiography at the Denver Health Medical Center. He received his medical degree from the Johns Hopkins University School of Medicine and served as a resident and chief resident in medicine at the University of California, San Francisco. After completing his fellowship in cardiology and receiving a master of science in public health from the University of Colorado Denver (UCD), Dr. Masoudi joined the UCD faculty; he is now an associate professor. Dr. Masoudi is an expert in clinical registries and quality measurement. He served for 2 years on the Research and Publications Committee of the National Cardiovascular Data Registry (NCDR) Implantable Cardioverter Defibrillator (ICD) Registry. He is the senior medical officer and chair of the NCDR Science Oversight Committee. He served as the clinical coordinator of the National Heart Care Projects sponsored by the Centers for Medicare and Medicaid Services (CMS) from 1999 to 2005 and is the clinical coordinator of the CMS Hospital Measures Special Study for acute myocardial infarction and heart failure. Dr. Masoudi has published more than 100 peer-reviewed papers. His most recent research has focused primarily on patterns of care and effectiveness of ICDs in community practice in the multicenter Cardiovascular Research Network. Dr. Masoudi holds positions in national organizations focused on quality of care and outcomes research. He is the chair of the American College of Cardiology–American Heart Association Task Force on Performance Measures, a member of the American Society of Echocardiography Quality Task Force, and an associate editor of *Circulation: Cardiovascular Quality and Outcomes*.

Eric D. Peterson, MD, MPH, FAHA, FACC, is a professor of medicine in the Division of Cardiology of Duke University Medical Center. He is also an associate director and director of cardiovascular research at the Duke Clinical Research Institute. His formal research training includes an MPH from Harvard University with emphasis in biostatistics, health economics, and decision analysis. Dr. Peterson is a leader in quality research, with 480 peer-reviewed publications in the field. He is also the principal investigator for the National Institutes of Health–Agency for Healthcare Research and Quality Duke Centers for Education and Research on Therapeutics, the Society of Thoracic Surgeons National Cardiac Surgery Database, and the Data Coordinating Center for the American College of Cardiology (ACC) National Cardiovascular Data Registry (NCDR) and the American Heart Association (AHA) Get With the Guidelines (GWTG) database. He is also principal investigator and center director for one of the four AHA Pharmaceutical Roundtable Outcomes Centers nationwide. Dr. Peterson

participates on multiple national committees, including the AHA Quality of Care and Outcomes Research Interdisciplinary Working Group and the AHA Strategic Planning Committee and the National Quality Forum Outpatient Imaging Efficiency Project Steering Committee and the ACC–AHA Performance Measures Task Force; the American College of Cardiology Foundation Appropriateness Criteria Implementation Working Group; the Department of Veterans Affairs Quality Enhancement Research Initiative Executive Committee; the oversight board of the Massachusetts Data Analysis Center; the National Quality Forum Technical Advisory Panel for Priorities, Goals and a Measurement Framework: Efficiency and Episodes of Care; the Institute of Medicine (IOM) Committee on Redesigning Insurance Benefits, Provider Payments, and Accountability Programs to Promote Quality of Health Care Delivery; and the IOM Committee on Secondhand Smoke Exposure and Acute Coronary Events. Dr. Peterson is a member of the American Society for Clinical Investigation Council. He received the DukeMed Scholar Award in 2007. In April 2010, he became the Fred Cobb, MD Distinguished Professor. He is also a contributing editor of the *Journal of the American Medical Association*.

Rita F. Redberg, FACC, MD, MSc, graduated from Cornell University and the University of Pennsylvania Medical School and received a master's of science and health policy and administration from the London School of Economics. She is a professor of medicine in the Division of Cardiology of the University of California, San Francisco School of Medicine. She is the editor of *Archives of Internal Medicine*. Dr. Redberg has had long experience and training in health-policy and technology assessment. She served on the Medicare Evidence, Development and Coverage Advisory Commission from 2003 to 2006. She is a member of the California Technology Assessment Forum, the Medical Policy Technology and Advisory Committee, and the Food and Drug Administration (FDA) Cardiovascular Devices Expert Panel and is a consultant to the Center for Medical Technology for medical technology policy. She has recently completed an extensive review of the FDA cardiovascular-device premarket approval process, which was published in the *Journal of the American Medical Association*. In addition, Dr. Redberg has gained extensive knowledge of health-care and health-care–financing legislation through her experience in working with the Senate Judiciary Committee as a Robert Wood Johnson Health Policy Fellow in 2003–2006. In 2008, she was invited to speak at the Institute of Medicine (IOM) Evidence Based Medicine Roundtable on Engineering a Learning Healthcare System: A Look at the Future. Dr. Redberg is a member of the American College of Cardiology (ACC) Clinical Quality Committee and serves on the Quality in Technology Work Group. She does comparative-effectiveness research and serves on the ACC Comparative Effectiveness Work Group and several other ACC committees, including several on appropriate use of

cardiac imaging. She is the ACC representative to the Institute of Clinical and Economic Review Advisory Board.

Frederic S. Resnic, MD, received his undergraduate degree in electrical engineering from Duke University and his MD from Mount Sinai School of Medicine and completed his residency, cardiovascular fellowship, and interventional cardiology fellowship at Brigham and Women's Hospital (BWH) and Harvard Medical School. In addition, he has a master of science degree in medical informatics from the Massachusetts Institute of Technology and has completed a fellowship in medical informatics at BWH. In 2006, Dr. Resnic was appointed the director of the BWH Cardiac Catheterization Laboratory. His research interests have focused on the development of informatics tools to monitor medical device and procedural safety, and he leads a research program funded by the National Institutes of Health and the Food and Drug Administration (FDA) to explore the automated surveillance of medical-device safety in a network of Massachusetts hospitals. In addition to his clinical and research work at BWH, Dr. Resnic serves as a senior medical adviser for the Massachusetts Data Analysis Center and the Massachusetts Department of Public Health. In 2007, Dr. Resnic was appointed to the medical advisory panel for circulatory devices for FDA; in 2008, he became the chairperson of the Quality Oversight Committee for Massachusetts of the Massachusetts chapter of the American College of Cardiology (ACC). Dr. Resnic was elected president of the Massachusetts chapter of ACC in 2009.

Paul D. Varosy, MD, is the director of cardiac electrophysiology in the Department of Veterans Affairs (VA) Eastern Colorado Health Care System and assistant professor of medicine at the University of Colorado Denver. He is a recipient of a research career development award from the VA Office of Health Services Research and Development (HSR&D), evaluating real-world outcomes in veterans who have implantable cardioverter defibrillators (ICDs) and are enrolled in the VA National ICD Surveillance Center. He also serves in national roles as a member of the Science and Publications Committee for the National Cardiovascular Data Registry ICD Registry and as a member of the development team for the Safety of Atrial Fibrillation Ablation Registry Initiative. As a member of VA's Denver-based Cardiovascular Assessment, Reporting, and Tracking (CART) program (the VA national program under the leadership of John Rumsfeld, MD, PhD, by which all cardiology procedures are documented in the VA system), Dr. Varosy is leading the development of CART-EP, a new comprehensive VA-wide implantable arrhythmia-device monitoring system spanning preimplantation evaluation, documentation of the implantation procedure, in-clinic followup, and remote monitoring followup for veterans who have pacemakers, implantable defibrillators, and cardiac resynchronization therapy devices.

C

510(k) Premarket Notification Analysis of FDA Recall Data

William H. Maisel, MD, MPH

EXECUTIVE SUMMARY

The 510(k) process requires a device manufacturer to notify the Food and Drug Administration (FDA) before it intends to market a device and to establish that the device is "substantially equivalent" to a legally marketed "predicate" device that does not require premarket approval (PMA). A recall is an action taken to address a problem with a medical device that violates FDA law. Recalls occur when a medical device is defective and/or when it could be a risk to health.

Analysis of FDA's 510(k) Database (1996–2009) and Recall Database (2003–2009) revealed the following:

1. 48,402 510(k)s were cleared by FDA between January 1, 1996, and December 31, 2009, and were available for analysis.
2. From 2003 to 2009, 3,132 unique 510(k)s were subject to recall. Among 510(k)s affected by recall, 73.9% were recalled a single time and 26.1% were recalled more than once, including nearly 2% that were recalled more than five times.
3. Among 510(k)s cleared in 2003–2009, 98.4% remained free of recall 1 year following the decision. Longer-term follow-up shows that 92.6% and 91.5% of 510(k)s remain free of recall 5 and 6 years, respectively, following regulatory clearance.
4. The annual 510(k) recall rate is highest in the first 3 years following clearance (1.6–1.9%/year). Lower recall rates are observed in years 5 and 6 post clearance (0.9–1.1%/year).
5. More than half the 510(k) recalls are due to manufacturing process

errors (28.8%) or device design issues (28.4%). Materials and component issues (16.3%) and change control processes (11.9%) account for the majority of the remaining 510(k) recalls.

6. Compared to 510(k)s unaffected by recall, recalled 510(k)s are more likely to have been reviewed by a third party or submitted as a Special application (rather than Traditional or Abbreviated). Recalls are also more likely to affect 510(k)s involving life sustaining devices and Class III devices.

OVERVIEW OF 510(k) PREMARKET NOTIFICATION PROGRAM AND FDA RECALLS

The 510(k) process requires a device manufacturer to notify FDA before it intends to market a device and to establish that the device is "substantially equivalent" to a legally marketed "predicate" device that does not require a PMA.[1]

One measure of the success of the 510(k) premarket notification process would be to assess the performance and reliability of the thousands of individual devices that have been cleared via this pathway. This is both impractical and impossible. Recalls and Medical Device Reports represent surrogate markers of device reliability.

A recall is an action taken to address a problem with a medical device that violates FDA law.[2] Recalls occur when a medical device is defective and/or when it could be a risk to health. Recalls may be conducted on a firm's own initiative, by FDA request, or by FDA order under statutory authority. Importantly, recalls may be issued when there is only a small chance (sometimes <1%) of device malfunction or patient injury. Recalls are classified by FDA as follows:

- **Class I recall:** a situation in which there is a reasonable probability that the use of or exposure to a violative product will cause serious adverse health consequences or death.
- **Class II recall:** a situation in which use of or exposure to a violative product may cause temporary or medically reversible adverse health consequences or where the probability of serious adverse health consequences is remote.
- **Class III recall:** a situation in which use of or exposure to a violative product is not likely to cause adverse health consequences.

[1]Government Accountability Office. Medical Devices: FDA Should Take Steps to Ensure That High-Risk Device Types Are Approved Through the Most Stringent Premarket Review Process. January 2009. Accessed June 16, 2009 at http://www.gao.gov/new.items/d09190.pdf.

[2]U.S. Food and Drug Administration. Recalls, Market Withdrawals, and Safety Alerts. Accessed at http://www.fda.gov/Safety/Recalls/ucm165546.htm.

Medical Device Reporting (MDR) is a mechanism by which FDA receives information on significant medical device adverse events from manufacturers, importers, and user facilities (hospitals, nursing homes, etc.). Because of incomplete adverse event descriptions, significant underreporting of device-related adverse events, and the absence of denominator data, these data have significant limitations. Nevertheless, MDR analysis can provide some insights into the performance of the 510(k) program.

METHODOLOGY

Methods 2.1

Two primary databases—510(k) database and recall database—were used to conduct the analysis presented in this report. A third database, containing MDR data, was also analyzed. Each database was provided by FDA to the Institute of Medicine. Data analysis was conducted independently of FDA.

Methods 2.1.1 510(k) Database for Years 1996–2009
Consisted of all 510(k) applications that were submitted between January 1, 1996, and December 31, 2009, and were found to be substantially equivalent (SE).

Methods 2.1.2 FDA Recall Database for Years 2003–2009
Consisted of all medical device recalls from January 1, 2003, to December 31, 2009.

Methods 2.1.3 Medical Device Reporting (MDR) Data
Consisted of a subset of all submitted MDRs. A subset of the MDRs occurring between January 1, 2003, and December 31, 2009, were analyzed.

Methods 2.2

All statistical calculations were performed using SAS (Version 9.1, Cary, NC). A two-sided P value of <0.05 was considered to be statistically significant. Chi-square tests, Mantel–Haenszel chi-square tests, and log-rank tests were used where appropriate.

DATA ANALYSIS

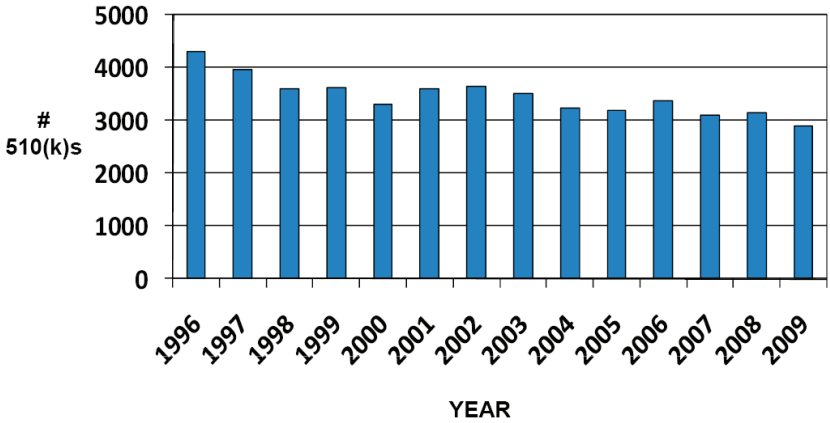

FIGURE C-1 Annual number of 510(k) applications.

- From 1996 to 2009, 48,402 510(k)s were available for analysis.
- The annual number of applications ranged from a high of 4,286 in 1996 to a low of 2,890 in 2009.

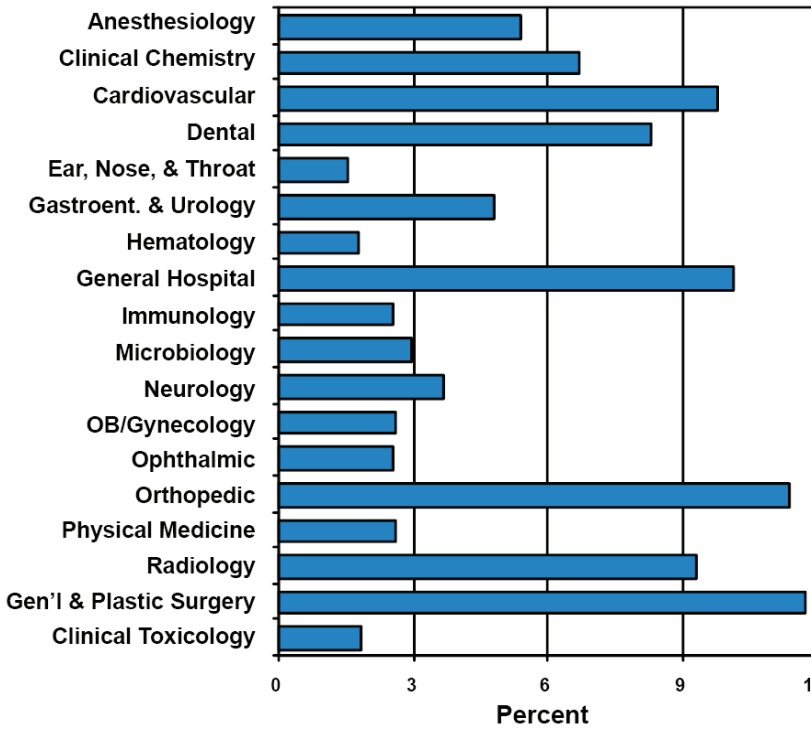

FIGURE C-2 Advisory committee assignments for submitted 510(k)s.

- All devices submitted for clearance under the 510(k) program are given an Advisory Committee assignment.
- Devices cleared via the 510(k) program are most often assigned to the General & Plastic Surgery (11.8%), Orthopedic (11.4%), General Hospital (10.2%), and Cardiovascular (9.8%) Advisory Committees.
- 510(k) devices are least often assigned to the Ear, Nose, & Throat (1.5%), Hematology (1.8%), and Clinical Toxicology (1.9%) Advisory Committees.
- While Advisory Committee assignments inform about the type of devices cleared, the committees are rarely involved in the premarket 510(k) process.

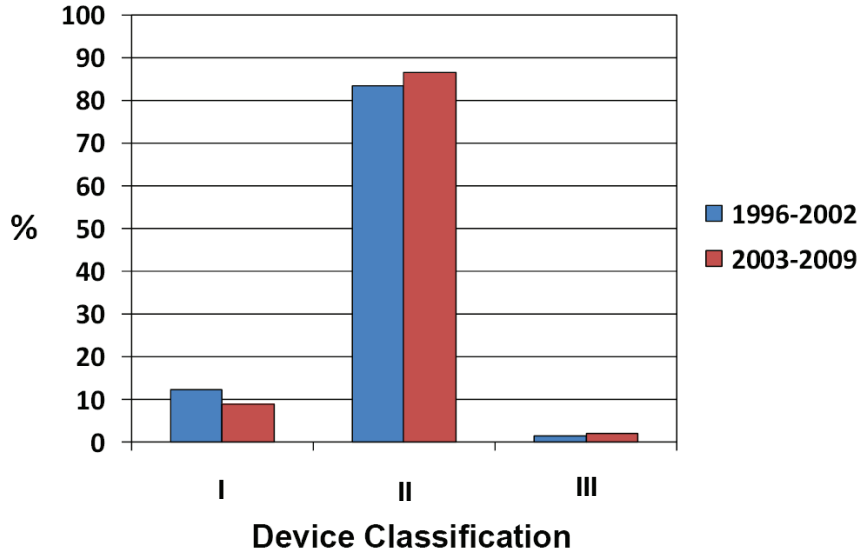

FIGURE C-3 510(k) device classification.

- More than 80% of 510(k) devices are classified as Class II, ~10% as Class I, and <2% as Class III.
- Minor changes in the distribution of device classification have occurred from 1996 to 2002 compared to 2003–2009. Specifically, there has been a small reduction in the number of 510(k) devices classified as Class I, and a small increase in the number of Class II and Class III designations.

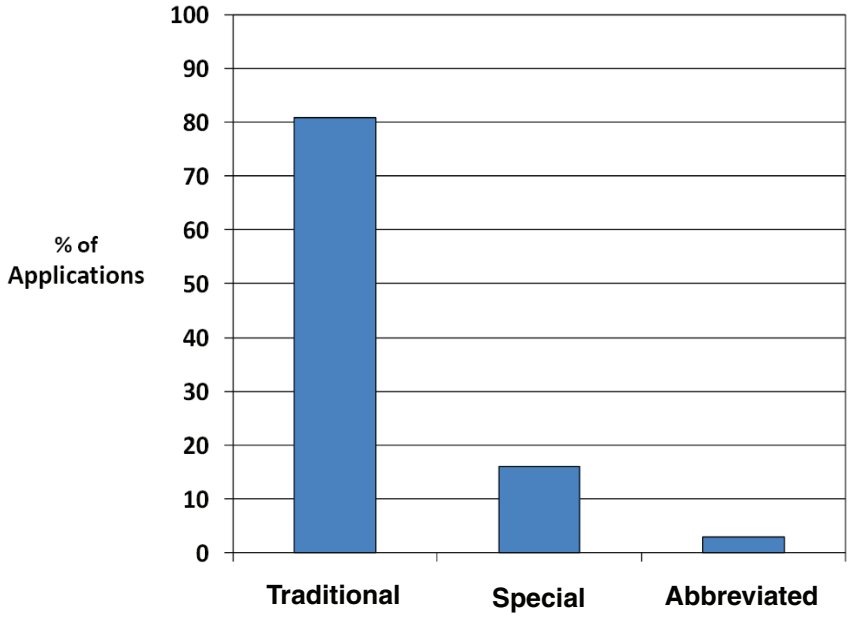

FIGURE C-4 510(k) type, 1996–2009.

- The majority of 510(k)s are Traditional (80.9%). Fewer are Special (16.0%) or Abbreviated (3.0%).
- Special 510(k)s are submitted when a manufacturer makes modifications to its own device, design control processes are appropriate, and design validation is performed. Abbreviated 510(k)s are submitted when FDA guidance, a special control, or recognized standard exists and the manufacturer intends to use it. All other 510(k) devices utilize the traditional pathway.

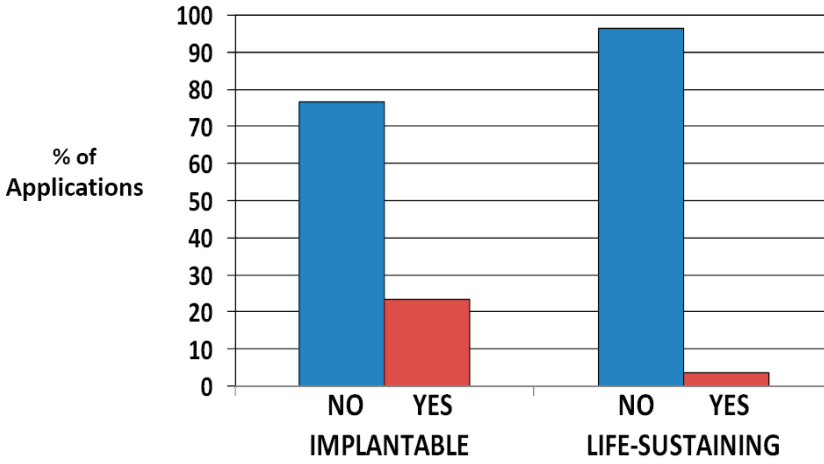

FIGURE C-5 510(k) implantable and life-sustaining features, 2003–2009.

- Nearly one-quarter (23.4%) of 510(k)s are implantable devices.
- The vast majority of 510(k) devices are NOT considered life-sustaining. However, a small percentage (3.5%) are life-sustaining devices.

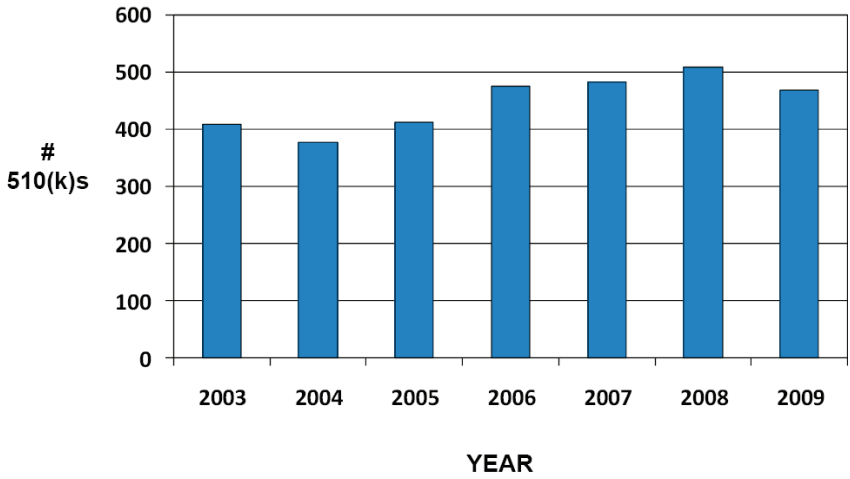

FIGURE C-6 Annual number of unique 510(k)s affected by recall.

- From 2003 to 2009, 3,132 unique 510(k)s were subject to FDA recall.
- The annual number of recalled 510(k)s ranged from a low of 377 in 2004 to a high of 508 in 2008.

TABLE C-1 Number of Recalls per 510(k)

Recalls per 510(k)	# Occurrences	Percentage of 510(k) Recalls
1	2,298	73.9
2	492	15.8
3	150	4.8
4	74	2.4
5	37	1.2
6	16	0.5
7	15	0.5
8	10	0.3
9	8	0.3
10	2	0.1
>10	7	0.2

- Among 510(k)s affected by recall, 73.9% were recalled a single time.
- 26.1% of recalled 510(k)s were recalled more than once, including nearly 2% that were recalled more than five times.
- Importantly, multiple recalls may be due to expansion of an initial recall to additional products with the same potential defect and do not necessarily represent multiple modes of product failure. Additionally, because some 510(k)s contain more than one device, multiple recalls do not necessarily represent product defects repeatedly affecting the same device.

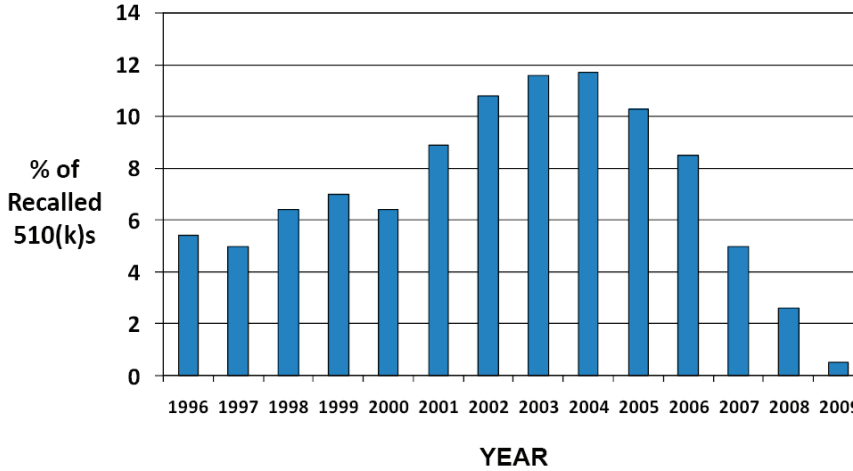

FIGURE C-7 Year of 510(k) decision for recalls occurring in 2003–2009.

- The original decision year for 510(k)s affected by recall in 2003–2009 is displayed.
- Nearly half (49.9%) of the 510(k) recalls occurring in 2003–2009 were for products cleared in 1996–2002.
- 40.6% of recalls were for products that were cleared in 2003–2009.

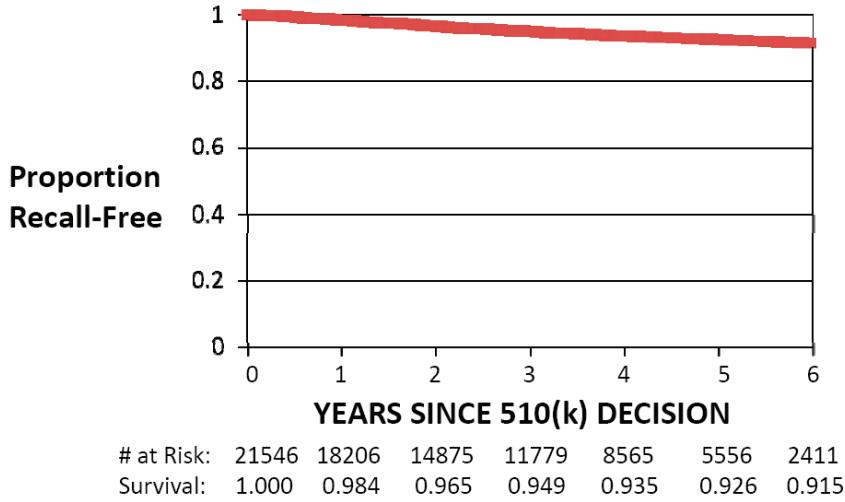

FIGURE C-8 Recall-free 510(k) "survival," 2003–2009.

- Only 510(k)s cleared from January 1, 2003, to December 31, 2009, are included in this graph. Kaplan–Meier "survival" estimates were calculated using standard methodology.
- Recall-free survival estimates demonstrate that 98.4% of 510(k)s cleared in 2003–2009 remained free of recall 1 year following the decision.
- Longer term follow-up shows that 92.6% and 91.5% of 510(k)s remain free of recall 5 and 6 years, respectively, following regulatory clearance.

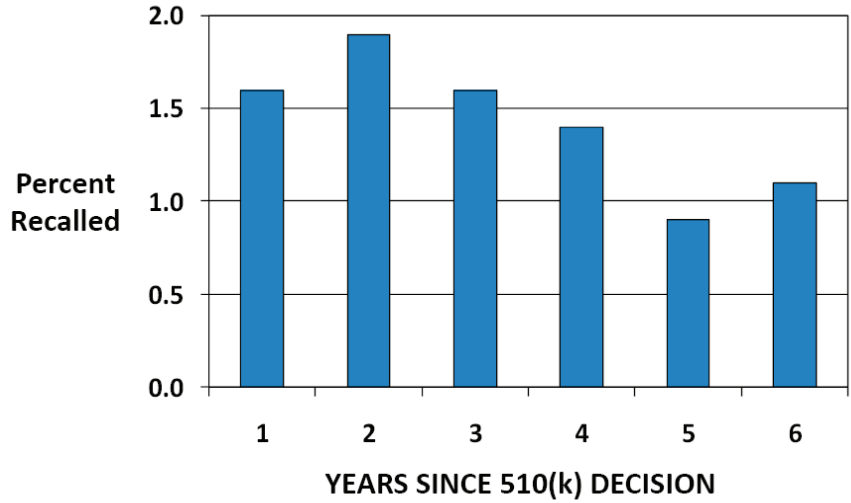

FIGURE C-9 Annual 510(k) recall rate based on years since decision.

- The annual rate of recall for 510(k)s cleared January 1, 2003, to December 31, 2009, is displayed.
- The annual recall rate is highest in the first 3 years following clearance (1.6–1.9%).
- The lowest recall rates are observed in years 5 and 6 post clearance (0.9–1.1%).

TABLE C-2 Causes of 510(k) Recalls

Cause	Description	Recalls (%)
Manufacturing Process	Manufacturing process inadequately controlled, inadequate environmental controls, storage, packaging, labeling, equipment maintenance, material removal, etc.	28.8
Device Design	Failure of device to perform as intended despite meeting design specifications.	28.4
Materials/ Components	Materials/components that are non-conforming, contaminated, degraded, counterfeit, or inadequately tested	16.3
Change Control	A change made to a specification, program, procedure, vendor, etc. that adversely affects a component, finished device, packaging, leveling, etc.	11.9
Employee Error	Employee error (not a systematic problem). Usually corrected by retraining.	7.1
Miscellaneous		7.5
TOTAL		100

- FDA classifies the cause of each recall based on the available information.
- More than half the 510(k) recalls are due to manufacturing process errors (28.8%) or device design issues (28.4%).
- Materials and component issues (16.3%) and change control processes (11.9%) account for the majority of the remaining recalls.
- Employee error is a less common cause of 510(k) recalls (7.1%).

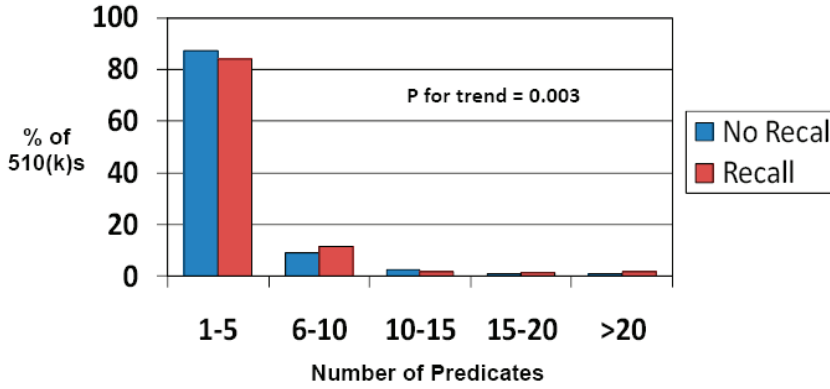

FIGURE C-10 Impact of predicate number on 510(k) recalls, 2004–2009.

- The number of predicates cited by each individual 510(k) is displayed.
- Most 510(k)s utilize 1-5 predicates (>80%). Fewer than 5% utilize more than 10 predicates.
- Fewer predicates (1-5) are associated with a lower rate of recall while a higher number of predicates (6-10, for example) are associated with an increased 510(k) recall rate.
- Many of the submissions that cite multiple predicates contain multiple products (for example, submissions to the Office of In Vitro Diagnostic Device Evaluation and Safety [OIVD] for diagnostic tests). Therefore, one cannot conclude from these data that multiple predicates are unsafe without indexing the recall rate to the number of unique products at risk. This latter analysis could not be performed based on the available data.

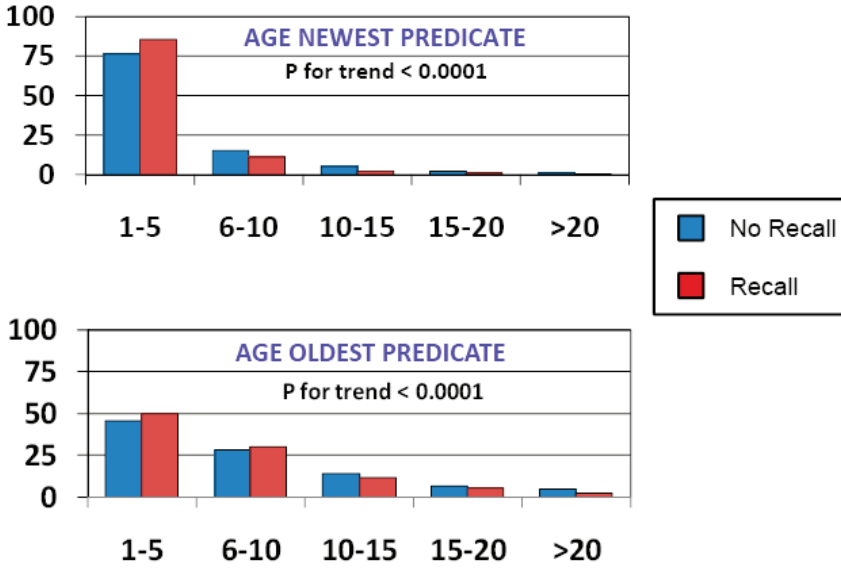

FIGURE C-11 Impact of predicate age on 510(k) recalls, 2004–2009.

- The age of the newest predicate cited for 510(k) clearance was <5 years in most (>75%) cases and <10 years in >90%.
- The age of the oldest predicate was <5 years in ~ 50%, 6-10 years in ~ 25%, and >10 years ~25% of the time.
- The newest predicate and the oldest predicate tended to be younger among recalled 510(k)s than among those 510(k)s that had been unaffected by recall.

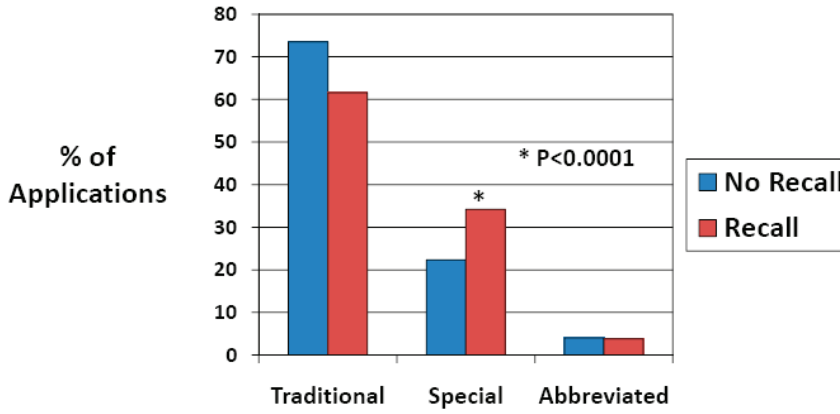

FIGURE C-12 Impact of 510(k) type on recall rate, 2003–2009.

- Traditional 510(k) were most common for both 510(k)s affected and unaffected by recalls.
- Special 510(k)s represented a higher percentage of 510(k) type among recalled 510(k)s than among 510(k)s unaffected by recall (34.2% vs 22.3%).
- Abbreviated 510(k)s were infrequent for both 510(k)s affected and unaffected by recall.

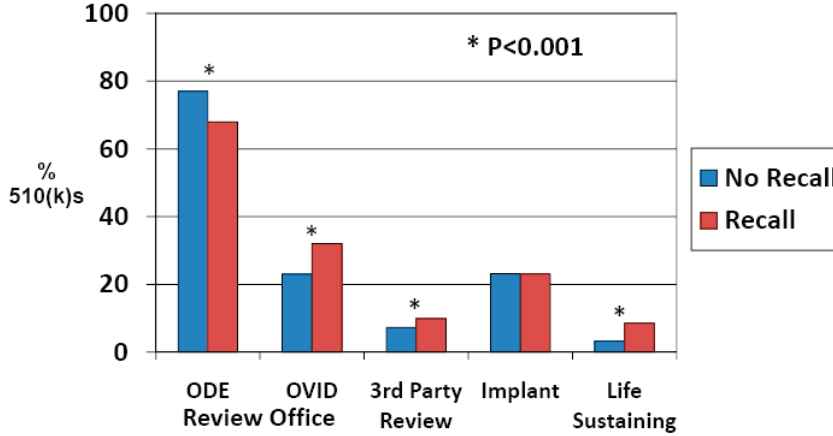

FIGURE C-13 Association of 510(k) features with recall rate, 2003–2009.

- 510(k)s affected by recall were less often cleared by the Office of Device Evaluation (67.9% vs 77.0%) and more often cleared by the OIVD (32.1% vs 23.0%) than devices unaffected by recall.
- 510(k)s were more often cleared via third party review among recalled 510(k)s than among those unaffected by recall (9.9% vs 7.3%).
- 510(k)s affected by recall were more often life-sustaining devices than 510(k)s unaffected by recall (8.5% vs 3.2%).

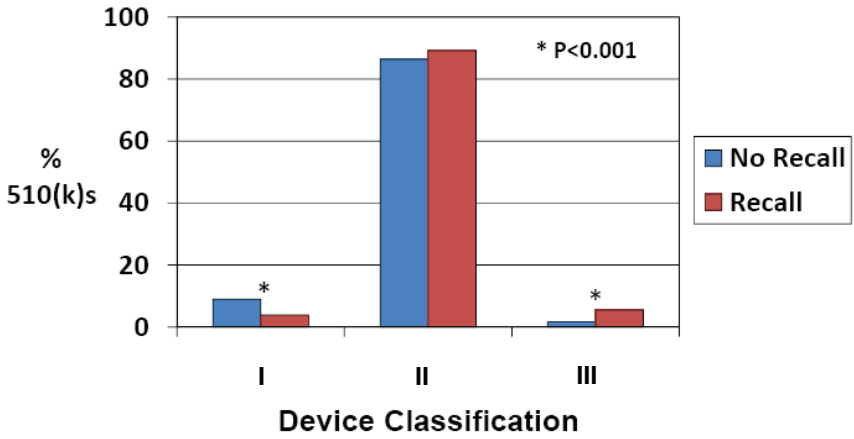

FIGURE C-14 Device classification and 510(k) recall rate, 2003–2009.

- More than 85% of 510(k)s were Class II devices for devices both affected and unaffected by recall.
- There was a more than 3-fold increase in the percentage of Class III devices among 510(k)s affected by recall compared to those unaffected (5.6% vs 1.7%). In contrast, there was a reduction in the percentage of Class I devices among recalled 510(k)s compared to 510(k)s unaffected by recall (5.6% vs 1.7%).

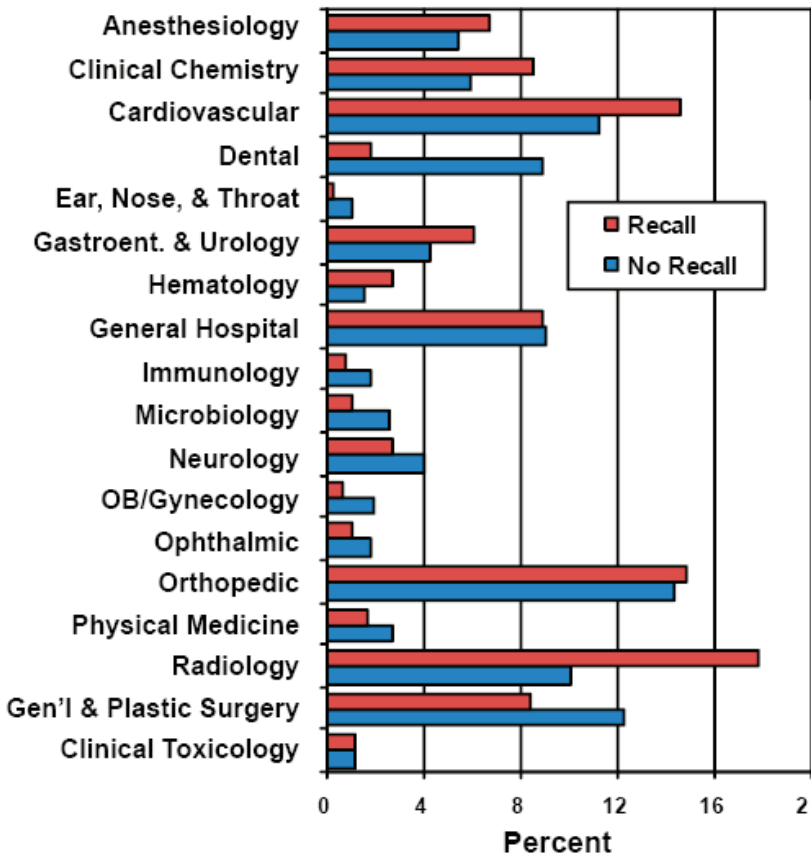

FIGURE C-15 Association of advisory committee assignments with 510(k) recall rate, 2003–2009.

- Devices associated with certain advisory committee assignments were more likely to be affected by recall.
- For example, Radiology (17.8% vs 10.1%), Cardiovascular (14.6% vs 11.2%), Anesthesiology (6.8% vs 5.4%), and Clinical Chemistry (8.6% vs 5.9%) represented a higher percentage of devices among recalled than unaffected 510(k)s.
- In contrast, Dental (1.8% vs 8.9%), Ear, Nose & Throat (0.3% vs 1.1%), and General & Plastic Surgery (8.4% vs 12.3%) represented a lower percentage among recalled devices compared with devices unaffected by recall.

TABLE C-3 Medical Device Reporting and 510(k)s

Description	Number
510(k)s with MDR	7823
Total MDRs	182,394
MDRs with Death	2361
MDRs with Injury	53,879
MDRs with Malfunction	12,110
MDRs Other	5051

- 182,394 MDRs associated with 7,823 510(k)s from 1996 to 2009 have been filed.
- Nearly two-thirds of the reports (66.4%) are associated with a device malfunction.
- 29.5% of the reports are associated with a patient injury, although <2% involve a patient death.
- **Data on MDRs are included in this report because they were specifically requested by the Institute of Medicine. These data are subject to incomplete reporting, insufficient information, and misclassifications and should be interpreted with caution.**

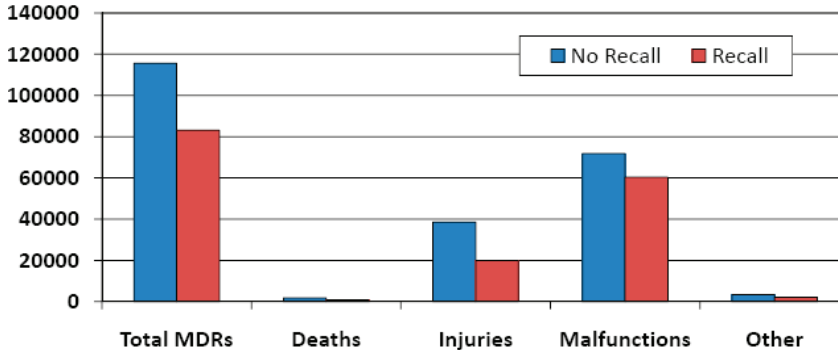

FIGURE C-16 Medical device reporting and recalls.

- 41.8% of 510(k) MDRs were associated with 510(k)s that were subject to recall.
- 30.0% of death adverse event reports, 34.0% of reported patient injuries, and 45.6% of device malfunction reports involving 510(k)s were associated with recalled 510(k)s.
- Data on MDRs are included in this report because they were specifically requested by the Institute of Medicine. These data are subject to incomplete reporting, insufficient information, and misclassifications and should be interpreted with caution.

SUMMARY OF KEY FINDINGS

- 48,402 510(k)s were cleared by FDA between January 1, 1996, and December 31, 2009, and were available for analysis.
- From 2003 to 2009, 3,132 unique 510(k)s were subject to recall. Among 510(k)s affected by recall, 73.9% were recalled a single time and 26.1% were recalled more than once, including nearly 2% that were recalled more than 5 times.
- Among 510(k)s cleared in 2003–2009, 98.4% remained free of recall 1 year following the decision. Longer term follow-up shows that 92.6% and 91.5% of 510(k)s remain free of recall 5 and 6 years, respectively, following regulatory clearance.
- The annual 510(k) recall rate is highest in the first 3 years following clearance (1.6–1.9%/year). Lower recall rates are observed in years 5 and 6 post clearance (0.9–1.1%/year).
- More than half the 510(k) recalls are due to manufacturing process errors (28.8%) or device design issues (28.4%). Materials and com-

ponent issues (16.3%) and change control processes (11.9%) account for the majority of the remaining 510(k) recalls.

- Compared to 510(k)s unaffected by recall, recalled 510(k)s are more likely to have been reviewed by a third party or submitted as a Special application (rather than Traditional or Abbreviated). Recalls are also more likely to affect 510(k)s involving life sustaining devices and Class III devices.

D

Trustworthy Medical Device Software

Kevin Fu, PhD

EXECUTIVE SUMMARY

This report summarizes what the computing research community knows about the role of trustworthy software for safety and effectiveness of medical devices. Research shows that problems in medical device software result largely from a failure to apply well-known systems engineering techniques, especially during specification of requirements and analysis of human factors. Recommendations to increase the trustworthiness of medical device software include (1) regulatory policies that specify outcome measures rather than technology, (2) collection of statistics on the role of software in medical devices, (3) establishment of open-research platforms for innovation, (4) clearer roles and responsibility for the shared burden of software, (5) clarification of the meaning of substantial equivalence for software, and (6) an increase in Food and Drug Administration (FDA) access to outside experts in software. This report draws upon material from research in software engineering and trustworthy computing, public FDA data, and accident reports to provide a high-level understanding of the issues surrounding the risks and benefits of medical device software.

INTRODUCTION

Software plays a significant and increasing role in the critical functions of medical devices. From 2002 to 2010, software-based medical devices resulted in over 537 recalls affecting more than 1,527,311 devices (Stewart and Fu, 2010). From 1999 to 2005, the number of recalls affecting devices containing software more than doubled from 118 to 273 (Bliznakov et al.,

2006). During this period, 11.3% of all recalls were attributable to software failures. This recall rate is nearly double compared to the period of 1983–1997 where only 6% of recalls were attributable to computer software (Wallace and Kuhn, 2001). For pacemakers and implantable cardioverter–defibrillators, the number of devices recalled due to software abnormalities more than doubled compared with 1991–2000 (Maisel et al., 2002). In 2006, Faris noted the milestone that over half the medical devices on the US market now involve software (Faris, 2006).

Yet, despite the lessons learned by tragic accidents, such as the radiation injuries and deaths caused by the Therac-25 linear accelerator over 20 years ago (Leveson and Turner, 1993), medical devices that depend on software continue to injure or kill patients in preventable ways. Problems in medical device software result largely from a failure to apply well-known systems engineering techniques, especially during specification of requirements and analysis of human factors.

"The ability of software to implement complex functionality that cannot be implemented at reasonable cost in hardware makes new kinds of medical devices possible. . ." (NRC, 2007).

Software Can Help and Hurt

Software can significantly affect patient safety in both positive and negative ways. Software helps to automatically detect dangerous glucose levels that could be fatal for a person using an insulin pump to treat diabetes. Medical linear accelerators use software to more precisely target the radiation dose. Remote monitoring of implanted devices may help to more quickly discover malfunctions and may lead to longer survival of patients (Kolata, 2010). However, medical device software contributes to the injury or death of patients. Problems ranging from poor user interfaces to overconfidence in software have led to accidents such as fatally incorrect dosages on infusion pumps (FDA, 2004a, 2010; Meier, 2010) and in radiation therapy (Leveson and Turner, 1993; Bogdanich, 2010b). A common trait for adverse events in medical device software is that the problems are often set in place before any implementation begins (see Table D-1).

Medical Devices Ought to Be Trustworthy

In the context of software, trustworthiness is inextricably linked with safety and effectiveness. There are several definitions of trustworthy software (see Sidebar 1) that vary by the specific contributions and terminology of various research subdisciplines. However, the fundamental idea is that software trustworthiness is a system property measuring how well a software system meets requirements such that stakeholders will trust in the

TABLE D-1 Examples of Adverse Events Where Medical Device Software Played a Significant Role

Eng. Stage	Adverse Event	Contributing Factor
Requirements Specification	Linear accelerator: Patients died from massive overdoses of radiation.	An FDA memo regarding the Corrective Action Plan (CAP) notes that "unfortunately, the AECL response also seems to point out an apparent lack of documentation on software specifications and a software test plan" (Leveson, 1995).
Design	Pacemakers and implantable defibrillators: Implant can be wirelessly tricked into inducing a fatal heart rhythm (Halperin et al., 2008).	Security and privacy need to be part of the early design process.
Human Factors	Infusion pump: Patients injured or killed by drug overdoses.	Software that did not prevent key bounce misinterpreted key presses of 20 mL as 200 mL (Flournoy, 2010).
Implementation	Infusion pump: Underdosed patient experienced increased intracranial pressure followed by brain death.	Buffer overflow (programming error) shut down pump (FDA, 2007).
Testing	Ambulance dispatch: Lost emergency calls.	An earlier system for the London Ambulance Service failed two major tests and was scuttled (Graham, 1992). Ambulance workers later accused the computer system of losing calls and said that "the number of deaths in north London became so acute that the computer system was withdrawn" (Tompsett, 1992). The ambulance company attributed the problems to "teething troubles" with a new computer system (Tompsett, 1992).
Maintenance	Health information technology (HIT) devices: Computer systems globally rendered unavailable.	An anti-virus update misclassified a core Windows operating system component as malware and quarantined the file, causing a continuous reboot cycle for any system that accepted the software update (Leyden, 2010). Numerous hospitals were affected. At Upstate University Hospital in New York, 2,500 of the 6,000 computers were affected (Tobin, 2010). In Rhode Island, a third of the hospitals were forced "to postpone elective surgeries and stop treating patients without traumas in emergency rooms" (Svensson, 2010).

Sidebar 1
The many definitions of trustworthiness.

One definition of trustworthy software is "software that is dependable (including but not limited to reliability, safety, security, availability, and maintainability) and customer-responsive. It can fulfill customer trust and meet the customer's stated, unstated, and even unanticipated needs" (Jayaswal and Patton, 2007). Another definition emphasizes the multidimensional, system-oriented nature that trustworthiness of a system implies that it is worthy of being trusted to satisfy its specified requirements (e.g., safety, effectiveness, dependability, reliability, security, privacy) with some [quantifiable] measures of assurance (Neumann, 2006). The National Science Foundation associates trustworthiness with properties of security, reliability, privacy, and usability—arguing that these "properties will lead to the levels of availability, dependability, confidentiality, and manageability that our systems, software and services must achieve in order to overcome the lack of trust people currently feel about computing and what computing enables" (NSF, 2010).

operation of the system. The requirements include overlapping and sometimes competing notions of safety, effectiveness, usability, dependability, reliability, security, privacy, availability, and maintainability.

Failure to meaningfully specify requirements, complacency, and lack of care for human factors can erode trustworthiness. The lack of trustworthy medical device software leads to shortfalls in properties such as safety, effectiveness, usability, dependability, reliability, security, and privacy. Good systems engineering (Ryschkewitsch et al., 2009) and the adoption of modern software engineering techniques can mitigate many of the risks of medical device software. Such techniques include a technical and managerial mindset that focuses on "design and development of the overall system" (Leveson, 1995) as opposed to focusing on optimization of components; meaningful specification of requirements such as intent specifications (Leveson, 2000); application of systems safety (Leveson, 1995); and static analysis (NITRD, 2009).

Although it is possible to create trustworthy medical device software under somewhat artificial constraints to achieve safety and effectiveness without satisfying other properties, in practice it is difficult to find environments where the properties are not linked. A medical device that works effectively in isolation may lose the effectiveness property if another component engineered separately joins the system, causing unanticipated inter-

actions. For example, a computer virus caused 300 patients to be turned away from radiation therapy because of shortfalls in security (BBC News, 2005). A security component can also reduce effectiveness if not designed in the context of system. For instance, a mammography imaging system may become ineffective if an automatic update of an anti-virus program designed to increase security causes the underlying operating system to instead fail (Leyden, 2010).

Innovations that combine computer technology with medical devices could greatly improve the quality of health care (Lee et al., 2006; NITRD, 2009), but the same life-saving technology could reduce safety because of the challenges of creating trustworthy medical device software. For instance, an implantable medical device with no physical means to wirelessly communicate over long distances may work safely and effectively for years. However, adding remote monitoring of telemetry to the device introduces an interface that fundamentally changes the properties of the overall system. The new system must require not only that any component designed to interact with the device be trustworthy, but also that any component capable of communicating with the device be trustworthy.

MEDICAL DEVICES, BUT WITH SOFTWARE: WHAT'S THE DIFFERENCE?

Patients benefit from software-based medical devices because "computers provide a level of power, speed, and control not otherwise possible" (Leveson, 1995). Without computer software, it would not be feasible to innovate a closed-loop, glucose-sensing insulin pump; a remotely monitored, implantable cardiac defibrillator; or a linear accelerator that calculates the radiation dose based on a patient's tissue density in each cross-section. However, the methodology used in practice to mitigate risks inherent in software have not kept pace with the deployment of software-based medical devices. For example, using techniques that work well to assure the safety and effectiveness of hardware or mechanical components will not mitigate the risks introduced by software. The following points use the writing of Pfleeger et al. (2001) with permission. There are several reasons why software requires a different set of tools to assure safety and effectiveness:

- The discrete (as opposed to continuous) nature of software (Lorge Parnas, 1985). Software is sensitive to small errors. Most engineered systems have large tolerances for error. For example, a 1-inch nail manufactured to be 1.0001 inch or 0.9999 inch can still be useful. Manufacturing is a continuous process, and small errors lead to results essentially the same as the exact, desired result. However, consider a slight error in entering a bolus dosage on an infusion pump. A

single key press error in selecting hours vs minutes could result in a bolus drip at 60 times the desired rate of drug delivery (FDA, 2004b). With some exceptions, small changes in continuous systems lead to small effects; small changes to discrete systems lead to large and often disastrous effects. The discrete nature of software also leads to limited ability to interpolate between test results. A system that correctly provides radiation doses of 20 centigrays (cGy) and 40 cGy does not on its own allow interpolation that would work correctly for 32 cGy. There is also seldom no direct equivalent to "over-engineering" safety margins for software systems in comparison to physical systems.

- The immaturity of software combined with rapid change. We keep running at an ever-faster pace to develop or use increasingly complex software systems that we do not fully understand, and we place such software in systems that are more and more critical. For example, a Networking and Information Technology Research and Development Program (NITRD) report of the High-Confidence Medical Devices, Software, and Systems (HCMDSS) Workshop (NITRD, 2009) notes that

> Many medical devices are, essentially, embedded systems. As such, software is often a fundamental, albeit not always obvious, part of a devices's functionality . . . . Devices and systems are becoming increasingly complicated and interconnected. We may already have reached the point where testing as the primary means to gain confidence in a system is impractical or ineffective.

The recent reporting of several radiation deaths stemming from medical linear accelerators (Bogdanich, 2010a) further highlights how complexity outpaces the maturity of present-day practices for creating trustworthy medical device software:

> "When it exceeds certain levels of complexity, there is not enough time and not enough resources to check the behavior of a complicated device to every possible, conceivable kind of input," said Dr. Williamson, the medical physicist from Virginia.
>
> But the technology introduces its own risks: it has created new avenues for error in software and operation, and those mistakes can be more difficult to detect. As a result, a single error that becomes embedded in a treatment plan can be repeated in multiple radiation sessions.

Despite these challenges, software has improved the effectiveness of critical systems in contexts such as avionics. Modern airplanes would be difficult to fly without the assistance of software, but airplanes have also introduced safety risks of software by using fly-by-wire (electronic) controls instead of pneumatics. However, there is a substantial belief among

software engineers that the medical device community (unlike the avionics community) does not take full advantage of well-known techniques for engineering software for critical systems. Many software engineers feel that that well-known technology not only lags, but is often ignored by medical device manufacturers. The safety culture of the avionics community does not appear to have universal appreciation in the medical device community.

TECHNIQUES TO CREATE TRUSTWORTHY MEDICAL DEVICE SOFTWARE

While the role of software in medical devices continues to increase in significance, deployment lags for well-known techniques that can mitigate many of the risks introduced by software. The following discussion draws from several technical documents on software engineering for critical systems.

The reader is strongly encouraged to read the full text of reports from NITRD on high-confidence medical devices (NITRD, 2009) and from the National Academies on software for dependable systems (NRC, 2007). Highly recommended reading on software engineering for critical systems includes Safeware: System Safety and Computers (Leveson, 1995) and Solid Software (Pfleeger et al., 2001) as well as evidence-based certification strategies such as the British Ministry of Defence Standard 00-56 (Ministry of Defence, 2007).

Adopt Modern Software Engineering Techniques

Medical device software lags in the adoption of modern software engineering techniques ranging from requirements specification to verification techniques. Fortunately, mature technology is already available to address common problems in medical device software, and that technology has been successful in other safety-critical industries such as avionics.

Programming languages that do not support software fault detection as comprehensively as possible should be avoided in medical device software. The C programming language, for example, has a very weak type system, and so the considerable benefits of strong type checking are lost. By contrast, the Ada programming language provides extensive support for software fault detection. Similarly, mechanical review of software using a technique known as static analysis is a mature technology that can identify possible faults quickly and efficiently. Static analysis supports the overall goal of developing trustworthy software and should be employed to the extent possible. Type checking and static analysis are two mature methods that guide software engineers toward safer and more effective medical device software by reducing or eliminating common sources of software errors.

Some programming systems permit a specification of software to be embedded into the software itself so that compliance of the code with the specification can be checked mechanically. A commercial system that provides this capability along with commercial support of both the language itself and the associated static analysis tools is SPARK Ada. Techniques such as these should be employed whenever possible to enable more effective testing and analysis of software.

A software specification is a statement of what the software has to do. Stating precisely what software has to do has proved extremely difficult, and specification is known to be a major source of software faults. Research over many years has yielded formal languages—i.e., languages with semantics defined in mathematics—that can help to avoid specification errors. Formal specification has been shown to be effective, and formal specifications for medical devices should be employed whenever possible.

Meaningfully Specify Requirements

Safety failures in software tend to stem from flaws during specification of requirements (Leveson, 1995). The first example in Table D-1 represents a failure of requirements specification in a 1980s linear accelerator that killed a number of patients, and some believe that the lack of meaningful systems-level specification of requirements contributed to the deaths in the recent radiation overdoses from a modern linear accelerator (Bogdanich, 2010a).

In critical systems, meaningful specification of requirements is crucial to properly anchor testing and analysis. Shortfalls in specification of requirements will lead to a false sense of safety and effectiveness during subsequent design, implementation, testing, etc. An example of meaningful specification of a requirement might be "stop delivery if dose exceeds patient's prescription" or "patient's received level of radiation must match level of radiation specified by operator." Such specification of requirements goes beyond purely functional descriptions such as "pressing start button begins primary infusion" or "delivered level of radiation adjusts to tissue density" that do not meaningfully capture the end-to-end system properties of a medical device.

Leading software engineers believe that many medical device manufacturers have an opportunity to significantly improve specification of requirements. In comparing medical devices to avionics systems, researchers wrote in the NITRD report High-Confidence Medical Devices: Cyber-Physical Systems for 21st Century Health Care (NITRD, 2009) that "perhaps the most striking [difference] is the almost complete lack of regard, in the medical-device software domain, for the specification of requirements." A National Academies report (NRC, 2007) similarly noted that "at least in comparison with other domains (such as medical devices), avionics software

appears to have fared well inasmuch as major losses of life and severe injuries have been avoided."

The NITRD report emphasizes that business models and incentives in the medical device sector lead to highly proprietary technologies that have two detrimental side effects: (1) companies are less likely to perceive value from specification of requirements, and (2) academic researchers have a much harder time in participating in the innovation of medical device technology.

The National Academies report recommended a direct path to dependable software (Jackson, 2009) for critical systems such as found in medical devices. Under this philosophy, system designers focus on providing direct evidence to support claims about software dependability. The approach contrasts with prescriptive standards that may otherwise dictate the specific claims. System designers are given flexibility to innovate by selecting the claims deemed necessary for the specific application at hand. Designers are forced to think carefully about proving the claims, but a difficulty remains in that the results are only as meaningful as the chosen claims.

Apply a Systems Engineering Approach

Software adds such complexity to the design of medical devices that the device must be treated as a system rather than an isolated component. The behavior of medical device software depends on its context within a system. Whereas biocompatibility of material may lend itself to conventional testing (Kucklick, 2006), the complexity of software requires a systems engineering approach (Ryschkewitsch et al., 2009). At a recent workshop on infusion pumps, it was pointed out that the 510(k) process is mostly a checklist, but this checklist approach provides less assurance as devices increase in complex system behavior (Chapman, 2010). Shuren (2010) provides an example of software-based medical devices that may operate safely and effectively in isolation, but not when integrated as a system:

> Images produced by a CT scanner from one vendor were presented as a mirror image by another vendor's picture archiving and communication system (PACS) web software. The PACS software vendor stipulates that something in the interface between the two products causes some images to be randomly "flipped" when displayed.

The NITRD report of the HCMDSS workshop (NITRD, 2009) notes that

> Integrating technology into the clinical environment—which includes practitioners, workflows, and specific devices—often lacks a holistic, systems perspective. Many medical devices are designed, developed, and marketed largely as

individual systems or gadgets. Device integration, interoperability, and safety features are not considered during development, acquisition, or deployment.

The rapid push toward device interoperability, wireless communication, and Internet connectivity will likely improve the effectiveness of care but will also reinforce the notion of medical device software as systems rather than isolated devices. Because medical devices are no longer isolated devices, an effective strategy for increasing trustworthiness is to follow good systems engineering methodology.

Evaluation of medical device software should require independent, third-party review by experts who are not connected with the manufacturer. Third-party evaluation in combination with good systems engineering can mitigate many of the system-level risks of medical device software.

Mitigate Risks Due to Human Factors

Poor understanding of human factors can lead to the design of medical device software that reinforces risky behavior, which can result in injury or death. For instance, a software application card used in an implantable drug pump was recalled because of a user interface where the hours and minutes fields for a bolus rate were ambiguously labeled on the computer screen (FDA, 2004a). A patient with an implantable drug pump died from an overdose because the health care professional set the bolus interval to 20 minutes rather than 20 hours (FDA, 2004b). Thus, the drug was administered at 60 times the desired rate. The patient passed out while driving, experienced a motor vehicle accident, and later died after the family removed life support.

Unmitigated risks of human factors also contributed to the recent radiation overdoses of patients treated by linear accelerators. One report from the *New York Times* (Bogdanich and Ruiz, 2010) quotes Dr. James Thrall, professor of radiology at Harvard Medical School and chairman of the American College of Radiology, saying, "There is nothing on the machine that tells the technologist that they've dialed in a badly incorrect radiation exposure."

Medical device software must accommodate inevitable human errors without affecting patient safety. Moreover, the specification of requirements should take into account all the key stakeholders. For instance, it is believed that some infusion pump manufacturers specify requirements based mostly on interactions with physicians rather than the primary operators of the pump: nurses. When nurses become disoriented and frustrated using infusion pumps, operational problems can result. Inadequate attention to human factors during specification of requirements will promote hazardous situations.

Mitigate Low-Probability, High-Consequence Risks

Manufacturers, health care professionals, and users often put too much confidence in medical device software. "It can't happen here." "There are no reported problems." Such statements have only a shallow basis in fact, but lead to a false sense of security. The manufacturer of the Therac-25 linear accelerator, which killed and injured a number of patients with radiation overdoses, initially responded to complaints from treatment facilities by saying that "the machine could not possibly over treat a patient and that no similar complaints were submitted to them" (Leveson and Turner, 1993; Leveson, 1995; Faris, 2006). It is very difficult to reproduce problems in software—often leading to denial rather than discovery of root causes. This difficulty derives in part from the complexity of a device's system-of-systems architecture and from the embedded nature of the system.

Security and privacy fall into the category of low-probability, high-consequence risks that could lead to widespread problems with little or no warning. Problems range from downtime to intentional harm to patients. Because devices can easily connect with physically insecure infrastructure such as the Internet and because software vulnerabilities (see Sidebar 2) are

Sidebar 2
Medical devices are susceptible to malware.

Medical devices are no more immune to malware (e.g., viruses, botnets, and keystroke loggers) than any other computer. Computer viruses can delete files, change values, expose data, and spread to other devices. A computer virus does not distinguish between a home computer and a hospital computer. Yet in the health care setting, the consequences of malicious software could lead to less effective care (e.g., corrupted electronic medical records that necessitate retesting) and diminished safety (e.g., overdoses from infusion pumps, radiation therapy, or implantable medical devices).

For these reasons, vendors may advise health care providers to install antivirus software with automated Internet-based updates. However, these products introduce risks that can themselves reduce the trustworthiness of the medical device software. When McAfee released an automated update of its virus definition files, the antivirus product incorrectly flagged a critical piece of Windows software as malicious—and quarantined the software (Leyden, 2010). This disruption of a critical file caused a number of hospitals to suffer downtime. Medical systems were rendered unavailable.

often discovered with little or no warning before threats exploit the vulnerability (Staniford et al., 2002), security and privacy outcome measures should play a central role in all major aspects of software development of medical device software (specification, design, human factors, implementation, testing, and maintenance).

Patients who receive treatment from a potentially lethal medical device should have access to information about its evaluation just as they have access to information about the side effects and risks of medications (NRC, 2007).

Specification of requirements should address low-probability, high-consequence risks. If a high-consequence risk proves too difficult or costly to mitigate, health care professionals deserve to know about the risks, no matter how small.

Innovations in wireless communication and computer networking have led to great improvements in patient care ranging from remote, mobile monitoring of patients (e.g., at-home monitors for cardiac arrhythmias or diabetes) to reduced risks of infection as a result of removing computer equipment from the sterile zone of an operating room (e.g., wireless wands for pacemaker reprogramming). However, the increased interconnectedness of medical devices leads to security and privacy risks for medical devices both in the hospital and in the home (Kilbridge, 2003; Fu, 2009; Maisel and Kohno, 2010). For instance, there is no public record of a specification that requires a home monitoring device to be physically incapable of reprogramming an implanted cardiac device. Thus, a malicious piece of software could change the behavior of a home monitor to quietly disable therapies or even induce a fatal heart rhythm—without violating the public specification.

POLICY RECOMMENDATIONS FOR TRUSTWORTHY MEDICAL-DEVICE SOFTWARE

Regulatory and economic policies should promote innovation while incentivizing trustworthiness in a least burdensome manner. One study of medical device recalls concludes that the economic impact of poor quality does not in general have severe financial penalties on the affected company (Thirumalai and Sinha, 2010). The policy recommendations below focus on technical and managerial issues rather than financial penalties or incentives.

Specify Outcome Measures, Not Technology

The safety and effectiveness of software-based medical devices could be better regulated in terms of outcome measures rather than in terms of

specific technologies.[1] The regulatory infrastructure should aim at making industry meet meaningful goals and show proof of achieving such goals.

The push toward prescriptive standards leads to an oversimplification in that the trustworthiness of a device depends on context. For example, one FDA notice advises to "update your operating system and medical device software" (FDA, 2009). However, software updates themselves can carry risks that should be either accepted or mitigated depending on the situation specific to each medical device. On a desktop computer used to update a portable automated external defibrillator (AED), it might be reasonable to routinely update the operating system even if there is a risk that the update may fail in a manner that makes the desktop machine inoperable. However, updating the operating system on the defibrillator itself carries a risk that failure could render the AED inoperable. A hospital that updates all its devices simultaneously is vulnerable to systemwide inability to provide care.

Rather than prescribe specific technologies, regulatory policies should incentivize manufacturers to specify meaningful outcome measures in the context of the given device and be required to prove such claims. Lessons from evidence-based medicine (IOM, 2007) could assist in creating outcome measures for trustworthy medical device software.

Collect Better Statistics on the Role of Software in Medical Devices

Many questions about the trustworthiness of medical device software are difficult to answer because of lack of data and inadequate record keeping. Questions include the following:

- To what degree are critical device functions being performed by software (vs hardware)? Is the amount increasing? Decreasing?
- What effect does software have on reliability? Availability? Maintainability? Ease of use?
- How do these software characteristics compare with similar implementations in hardware? Does the software make the device safer or more effective?
- What do data on the predicate device reveal about the new device? Do predicate data save time in specification of the new device? Do predicate data save time in testing of the new device?

Many record-keeping tools are already in place (e.g., the MAUDE adverse events database and the recalls database at FDA). However, these tools are severely underutilized. Databases suffer from severe underreporting. For

[1]In Europe, the legal definition of a medical device explicitly mentions software (Fries, 2005). In the United States, the legal definition of a medical device is less specific.

example, in the same time period there are only 372 adverse event reports in MAUDE that cite "computer software issues" despite there being well over 500 entries in the recall database that cite software as a reason for the recall. In the Department of Veterans Affairs, "over 122 medical devices have been compromised by malware over the last 14 months," according to House testimony (Baker, 2010). But there are no records in MAUDE citing a "computer system security issue."

Scott Bolte of GE Healthcare emphasizes that for security problems, formal reporting is especially lacking (Bolte, 2005):

> Although there is a lot of anecdotal evidence that malicious software has compromised medical devices, there is a notable lack of formal evidence. So without this formal reporting, FDA is limited in its ability to act or intervene. Reporting is something providers and arguably the manufacturers themselves can and should start doing immediately.

Policies should encourage better reporting of adverse events and recalls. Otherwise it will only be possible to point out anecdotal failures rather than confidently point out trends for successful products that epitomize innovation of trustworthy medical device software.

Enable Open Research in Software-Based Medical Devices

The highly proprietary nature of the medical device software industry makes it difficult for innovators to build upon techniques of properly built systems. Some information may become public after an accident, but this information teaches about failure rather than success. More open access to success stories of engineering medical device software would lead to innovation of safer and more effective devices. The NITRD report (2009) explains:

- Today we have open-research platforms that provide highly effective support for the widespread dissemination of new technologies and even the development of classified applications. The platforms also provide test beds for collaborations involving both researchers and practitioners. One spectacular example is the Berkeley Motes system with the TinyOS operating system.
- The medical-device community could benefit from the existence of such open-research platforms. They would enable academic researchers to become engaged in directly relevant problems while preserving the need for proprietary development by the industry. (TinyOS facilitates academic input even on government-classified technology, which is an example of what is possible.)
- An open research community needs to be established comprising academics and medical device manufacturers to create strategies for

the development of end-to-end, principled, engineering-based design and development tools.

Clearly Specify Roles and Responsibility

In complex systems of systems that rely on software, it is difficult to pinpoint a single party responsible for ensuring trustworthiness of software because the property is of the system of systems rather than of individual components. A modern linear accelerator is an example of a complex system of systems because commercial off-the-shelf (COTS) software such as Windows may serve as the underlying operating system for a separately engineered software application for planning and calculation of dose distribution. An embedded software system then uses the treatment plan to control mechanical components that deliver radiation therapy to a patient. When different entities separately manage software components in complex systems of systems, system-level properties such as safety are more difficult to ensure because no single entity is responsible for overall safety.

The FDA notes that a key challenge is a shared responsibility for failures in software (FDA, 2009). If the user updates the software on a medical device, is the manufacturer truly at fault? If a medical device relies on third party software such as operating systems, who is responsible for maintaining the software?

Technology alone is unlikely to mitigate risks that stem from system-level interactions of complex software designed by different organizations with different agendas and outcome measures. The problem is probably intractable without a single authority responsible for the trustworthiness of interfaces between interacting systems. The interface between medical device application software and COTS software is a common battleground for disclaimers of responsibility (see Sidebar 3).

Leveson (1995) points out that diffusion of responsibility and authority is an ineffective organizational structure that can have disastrous effects when safety is involved. The British Ministry of Defence (2007) provides a good example of clear roles and responsibilities for safety management of military systems. The ideas apply broadly to critical systems and may work well for medical systems.

Clarify the Meaning of Substantial Equivalence for Software

In the context of the 510(k) pre-market notification process, demonstration of "substantial equivalence" to a previously approved "predicate" medical device allows a manufacturer to more quickly seek approval to market a medical device (see Sidebar 4).

Imagine if the predicate device has a function implemented in hardware, and the manufacturer claims that the new version is substantially equivalent

Sidebar 3
Take Service Pack 3 and see me in the morning.

Medical devices can outlast the underlying operating system software. Many medical devices rely on commercial off-the-shelf (COTS) software, but COTS software tends to have a shorter lifetime for expected use than a typical medical device. For instance, Microsoft mainstream support for Windows XP lasted for less than 8 years (December 2001–April 2009) (Microsoft, 2003), whereas an MR scanner may have an operational life of 10–20 years (Bolte, 2005).

It is not uncommon for a newly announced medical device to rely on an operating system no longer supported by its manufacturer. Microsoft ended support for security patches for Windows XP Service Pack 2 and advises vendors to upgrade products to Service Pack 3. But hospitals often receive conflicting advice on whether to update software. House testimony (Joffe, 2009) mentions that

> As a sobering side-note, over the last three weeks, in collaboration with a researcher from Georgia Tech in Atlanta who is involved with the Conficker Working group, I have identified at least 300 critical medical devices from a single manufacturer that have been infected with Conficker. These devices are used in large hospitals, and allow doctors to view and manipulate high-intensity scans (MRI, CT Scans etc), and are often found in or near ICU facilities, connected to local area networks that include other critical medical devices.

because the only difference is that the new version is implemented in software. Because hardware and software exhibit significantly different behavior it is important that the design, implementation, testing, human factors analysis, and maintenance of the new device mitigate the risks inherent in software. However, this difference casts doubt on substantial equivalence because of the different technological characteristics that raise different risks to safety and effectiveness. Furthermore, when does a software-related flaw in a recalled predicate device imply that the same flaw exists in the new device?

As was noted at the Institute of Medicine Workshop on Public Health Effectiveness of the FDA 510(k) Clearance Process held in June 2010, there is doubt as to whether hardware can act as a predicate for functions implemented in software. Dr. David Feigel, former director of the FDA Center for Devices and Radiological Health (CDRH), said that "one of the interesting

> Worse, after we notified the manufacturer and identified and contacted the hospitals involved, both in the US and abroad, we were told that because of regulatory requirements, 90 days notice was required before these systems could be modified to remove the infections and vulnerabilities.

Users of medical picture archiving and communication systems (PACS) struggle to meet conflicting requirements: medical device manufacturers who require health care facilities to use old, insecure operating systems and FDA guidelines that advise keeping operating systems up-to-date with security patches. One anonymous posting on a technical support Web site (Windows Client Tech Center, 2008) reads:

> I am setting up a medical imaging facility and I am trying to do the same thing as well. The PACS system we are integrating with is only compatible with SP2. I order 6 new Dell workstations and they came preloaded with SP3. There are "actual" versions of programs out there that require SP2. For instance, the $250,000 Kodak suite I am installing. Plus a $30,000/yr service contract. This holds true for the majority of the hospitals which have PACS systems.
>
> However, if what you are saying is true then I found something useful within your post. You stated "if you installed XP with integrated sp3, it is not possible to downgrade sp3 to sp2," is this true? Do you have any supporting documentation as this would be very helpful so that I can provide Dell with a reason why I need to order downgraded XP discs.

The plaintive quality of this call for help provides insight into how helpless some users feel because of the diffusion of responsibility for maintaining COTS software contained within medical devices.

classes is radiation equipment . . . even the software, which I wonder where they got the first predicate for software" (IOM, 2010). The interpretation of substantial equivalence needs clarification for software-based medical devices.

Increase FDA Access to Outside Experts in Software Engineering

The FDA should increase its ability to maintain safety and effectiveness of medical devices by developing a steady pipeline of human resources with expertise in software engineering for critical systems.

Various offices within FDA's CDRH employ a small number of software experts. FDA also has a number of successful fellowship programs including the Commissioner's Fellowship Program, the Medical Device Fellowship Program, and the Device Evaluation Intern Program to attract

Sidebar 4
Substantial equivalence: paper or plastic?

An interesting thought experiment is to ask how the trustworthiness of electronic health records differs from traditional paper records. FDA generally does not consider a paper medical record as a medical device. However, FDA may consider an electronic health record as a medical device. Adding automated algorithms to prioritize display of data from an electronic medical record would shift the system toward regulation as a medical device.

Paper records are subject to threats such as fire, flood, misplacement, incorrect entry, and theft. Paper records are cumbersome to back up and require large storage rooms. But electronic records introduce risks qualitatively different from paper records. Making changes to a paper record tends to leave behind physical evidence that is auditable, but making electronic records auditable requires intentional design. A single coding error or errant key press could lead to destruction of an entire collection of electronic records—especially for encrypted data. The speed of technology can make electronic record keeping easier, but can encourage bad habits that to lead to difficult to detect mistakes. For instance, a computer display that clears the screen following the completion of an operation makes it difficult to trace back a sequence of changes. Overconfidence in software for electronic medical records could lead to financially motivated decisions to discontinue paper-based backup systems. One full-scale failure of a clinical computing system at the Beth Israel Deaconess Medical Center lasted four days—forcing the hospital to revert to manual processing of paper records (Kilbridge, 2003). While paper-based backup procedures allowed care to continue, few of the medical interns had any experience with writing orders on paper. When health care professionals struggle with technology, patients are at risk.

Heated debates about paper versus electronic recording appears in other contexts such as voting. A National Academies report (NRC, 2005) provides context for the electronic voting debate with arguments applicable to the safety and effectiveness of electronic medical records.

students and experienced experts from medical and scientific communities. However, software experts are notably underrepresented in these programs. The Web page for the Medical Device Fellowship Program[2] targets health

[2]http://www.fda.gov/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyPrograms/MedicalDeviceFellowshipProgramCDRH/default.htm.

professionals, and other existing programs primarily target biomedical engineers rather than software engineers. Of the fifty Commissioner's Fellows selected in 2009, none had formal training in computer science.[3] In 2008, one of the fifty fellows had a computer science degree, but did not work in CDRH. A former FDA manager indicated that software experts rarely participate in these fellowship programs. Another person familiar with FDA processes noted that seldom does an FDA inspector assigned to review a 510(k) application have experience in software engineering—even though the majority of medical devices today rely on software.

The FDA should expand its access to outside experts for medical device software by creating fellowship programs that target software engineers. For instance, FDA could more aggressively recruit students and faculty from computer science and engineering—especially individuals with advanced training in software engineering topics such as system and software safety, dependable computing, formal methods, formal verification, and trustworthy computing.

SUMMARY

The lack of trustworthy medical device software leads to shortfalls in safety and effectiveness, which are inextricably linked with properties such as usability, dependability, reliability, security, privacy, availability, and maintainability. Many risks of medical device software could be mitigated by applying well-known systems engineering techniques, especially during specification of requirements and analysis of human factors. Today, the frequency of news reports on tragic, preventable accidents involving software-based medical devices falls somewhere between that of automobile accidents and airplane accidents. Event reporting on tragic medical device accidents is likely headed toward the frequency of the former given the continued increase in system complexity of medical device software and present-day regulatory policies that do not adequately encourage use of modern software engineering and systems engineering practices.

ACKNOWLEDGMENTS

Several individuals deserve thanks for assistance with this report. Shari Lawrence Pfleeger provided extensive feedback and contributed material on threats and on why software behaves differently from hardware. John Knight provided feedback and contributed material on modern software engineering techniques. Philip Henneman, Daniel Jackson, Leon Osterweil,

[3]http://www.fda.gov/downloads/AboutFDA/WorkingatFDA/FellowshipInternshipGraduateFacultyPrograms/CommissionersFellowshipProgram/UCM216921.pdf.

and Jerry Saltzer provided feedback on early drafts of this report. Paul L. Jones, Nancy Leveson, John F. Murray Jr., and Gideon Yuval provided further technical input. Quinn Stewart collected data pertaining to FDA medical device recalls. Hong Zhang assisted with formatting of the report. Heather Colvin and Abigail Mitchell at the Institute of Medicine deserve special thanks for organizing a successful series of workshops on the Public Health Effectiveness of the FDA 510(k) Clearance Process.

REFERENCES

Baker, R.W. 2010. Statement of the Honorable Roger W. Baker. House Committee on Veterans' Affairs, Subcommittee on Oversight and Investigations, *Hearing on Assessing Information Security at the US Department of Veterans' Affairs*.

BBC News. 2005. *Hospital struck by computer virus*. [cited 10/29/2010] Available from http://news.bbc.co.uk/2/hi/uk_news/england/merseyside/4174204.stm.

Bliznakov, Z., G. Mitalas, and N. Pallikarakis. 2006. Analysis and classification of medical device recalls. *World congress on medical physics and biomedical engineering 2006,* 14(6):3782–3785.

Bogdanich, W. 2010a. *As technology surges, radiation safeguards lag*. The New York Times [cited 10/29/2010] Available from http://www.nytimes.com/2010/01/27/us/27radiation.html.

Bogdanich, W. 2010b. *Radiation offers new cures, and ways to do harm*. The New York Times [cited 10/29/2010] Available from http://www.nytimes.com/2010/01/24/health/24radiation.html.

Bogdanich, W., and R. Ruiz. 2010. *FDA to increase oversight of medical radiation*. The New York Times [cited 10/29/1020] Available from http://www.nytimes.com/2010/02/10/health/policy/10radiation.html.

Bolte, S. 2005. *Cybersecurity of medical devices* [cited 10/29/2010] Available from http://www.fda.gov/MedicalDevices/Safety/MedSunMedicalProductSafetyNetwork/ucm127816.htm.

Chapman, R. 2010. *Assurance cases for external infusion pumps. FDA Infusion Pump Workshop* [cited 10/29/2010] Available from http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM217456.pdf.

Faris, T.H. 2006. *Safe and Sound Software: Creating an Efficient and Effective Quality System for Software Medical Device Organizations*. ASQ Quality Press.

FDA (US Food and Drug Administration). 2004a. *Medtronic announces nationwide, voluntary recall of model 8870 software application card*. [cited 10/29/2010] Available from http://www.fda.gov/MedicalDevices/Safety/RecallsCorrectionsRemovals/ListofRecalls/ucm133126.htm.

FDA. 2004b. *Neuro n'vision programmer*. MAUDE Adverse Event Report [cited 10/29/2010] Available from http://www.accessdata.fda.gov/SCRIPTs/cdrh/cfdocs/cfMAUDE/Detail.cfm?MDRFOI__ID=527622.

FDA. 2007. *Baxter healthcare pte. ltd. colleague 3 cxe volumetric infusion pump 80frn. MAUDE Adverse Event Report*. [cited 10/29/2010] Available from http://www.accessdata.fda.gov/SCRIPTs/cdrh/cfdocs/cfMAUDE/Detail.cfm?MDRFOI__ID=914443.

FDA. 2009. *Reminder from FDA: Cybersecurity for networked medical devices is a shared responsibility* [cited 10/29/2010] Available from http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm189111.htm.

FDA. 2010. *Infusion pump improvement initiative* [cited 10/29/2010] Available from http://www.fda.gov/medicaldevices/productsandmedicalprocedures/GeneralHospitalDevicesandSupplies/InfusionPumps/ucm205424.htm.

Flournoy, V.A. 2010. *Medical device recalls*. [cited 10/29/2010] Available from http://www.fda.gov/downloads/MedicalDevices/NewsEvents/WorkshopsConferences/UCM216992.pdf

Fries, R.C. 2005. *Reliable Design of Medical Devices, Second Edition*. CRC Press.

Fu, K. 2009. Inside risks, reducing risks of implantable medical devices. *Communications of the ACM*, 52(6).

Graham, D.R. 1992. London ambulance service computer system problems. *Risks Digest*, 13(38) [cited 10/29/2010] Available from http://catless.ncl.ac.uk/Risks/13.38.html.

Halperin, D., T.S Heydt-Benjamin, B. Ransford, S.S. Clark, B. Defend, M. Benessa, W. Morgan, K. Fu, K. Tadayoshi, and W.H. Maisel. 2008. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *Proceedings of the 29th Annual IEEE Symposium on Security and Privacy.*

IOM (Institute of Medicine). 2007. *The Learning Healthcare System: Workshop Summary*. IOM Roundtable on Evidence-Based Medicine. Washington, DC: The National Academies Press.

IOM. 2010. Public Health Effectiveness of the FDA 510(k) Clearance Process: Balancing Patient Safety and Innovation. Edited by T. Wizemann. Washington, DC: The National Academies Press.

Jackson, D. 2009. A direct path to dependable software. *Communications of the ACM,* 52(4):78–88.

Jayaswal, B.J., and P.C. Patton. 2007. *Design for Trustworthy Software*. Prentice Hall, Pearson Education.

Joffe, R. 2009. House Committee on Energy and Commerce Subcommittee on Communications, Technology and the Internet. *Testimony of Rodney Joffe*, May 1 2009, [cited 10/29/2010] Available from http://energycommerce.house.gov/Press_111/20090501/testimony_joffe.pdf.

Kilbridge, P. 2003. Computer crash–lessons from a system failure. *New England Journal of Medicine,* 348(10):881–882.

Kolata, G. 2010. *New tools for helping heart patients*. The New York Times [cited 10/29/2010] Available from http://www.nytimes.com/2010/06/22/health/22heart.html.

Kucklick, T.R. 2006. *The Medical Device R&D Handbook*. Taylor & Francis.

Lee, I., G. Pappas, R. Cleaveland, J. Hatcliff, B. Krogh, P. Lee, H. Rubin, and L. Sha. 2006. High-confidence medical device software and systems. *Computer*, 39(4):33–38.

Leveson, N.G., and C. Turner. 1993. An investigation of the Therac-25 accidents. *IEEE Computer*, 26(7):18–41.

Leveson, N.G. 1995. *Safeware: System Safety and Computers*. Addison-Wesley.

Leveson, N.G. 2000. Intent specifications: An approach to building human-centered specifications. *IEEE Transactions on Software Engineering*, 26:15–35.

Leyden, J. 2010. *Rogue McAfee update strikes police, hospitals and Intel*. [cited 10/29/2010] Available from http://www.theregister.co.uk/2010/04/22/mcafee_false_positive_analysis/.

Lorge Parnas, D. 1985. Software aspects of strategic defense systems. *Communications of the ACM*, 28(12):1326–1335.

Maisel, W. H. and T. Kohno. 2010. Improving the security and privacy of implantable medical devices. *New England Journal of Medicine*, 362(13):1164–1166.

Maisel, W. H., W.G. Stevenson, and L.M. Epstein. 2002. Changing trends in pacemaker and implantable cardioverter defibrillator generator advisories. *Pacing and Clinical Electrophysiology*, 25(12):1670–1678.

Meier, Barry. 2010. *FDA steps up oversight of infusion pumps*. The New York Times [cited 10/29/2010] Available from http://www.nytimes.com/2010/04/24/business/24pump.html.

Microsoft. 2003. *Windows XP history*. [cited 10/29/2010] Available from http://www.microsoft.com/windows/WinHistoryProGraphic.mspx.

Ministry of Defence. 2007. *Ministry of defence standard 00-56: Safety management requirements for defence systems*, June 2007. Issue 4.

Neumann, P.G. 2006. *Risks of untrustworthiness*. In 22nd Annual Computer Security Applications Conference (ACSAC).

NITRD (Networking and Information Technology Research and Development). 2009. High-confidence medical devices: Cyber-physical systems for 21st century health care. Networking and Information Technology Research and Development. *Networking and Information Technology Research and Development Program*.

NRC (National Research Council). 2005. *Asking the Right Questions About Electronic Voting*. Washington, DC: The National Academies Press.

NRC. 2007. *Software for Dependable Systems: Sufficient Evidence?* Washington, DC: The National Academies Press.

NSF (National Science Foundation). 2010. *Trustworthy computing*. The National Science Foundation. [cited 10/29/2010] Available from http://www.nsf.gov/funding/pgm_summ.jsp? pims_id=503326.

Pfleeger, C.P. and S.L. Pfleeger. 2007. *Security in Computing*. Prentice Hall, 4e edition.

Pfleeger, S.L., L. Hatton, C. Howell. 2001. *Solid Software*. Prentice Hall.

Ryschkewitsch, M., D. Schaible, W. Larson. 2009. [cited 10/29/2010] *The art and science of systems engineering*. *NASA Monograph*, Available from http://spacese.spacegrant.org/uploads/images/Art_and_Sci_of_SE.pdf.

Shuren, J. 2010. Testimony of Jeffrey Shuren, Director of FDA's Center for Devices and Radiological Health. *Health Information Technology (HIT) Policy Committee Adoption/Certification Workgroup*.

Staniford, S., V. Paxson, and N. Weaver. 2002. How to own the internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium*, pages 149–167.

Stewart, Q., and K. Fu. 2010. *A second look: Analysis of FDA recalls of software-based medical devices*. Manuscript, July 2010.

Svensson, P. 2010. *McAfee antivirus program goes berserk, freezes PCs*. Associated Press, [cited 10/24.2010] Available from http://abcnews.go.com/Technology/wireStory?id=10437730

Thirumalai, S., and K.K. Sinha. 2010. Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. Accepted to *Management Science*, May 2010.

Tobin, D. 2010. University hospital computers plagued by anti-virus glitch. *The Post-Standard*, [cited 10/29/2010] Available from http://www.syracuse.com/news/index.ssf/2010/04/university_hospital_plagued_by.html.

Tompsett, B. 1992. Long call wait for London ambulances. *Risks Digest*, Volume 13, Issue 42, [cited 10/29/2010] Available from http://catless.ncl.ac.uk/Risks/13.42.html#subj8.

Wallace, D. R., and D. Kuhn. 2001. Failure modes in medical device software: an analysis of 15 years of recall data. *International Journal of Reliability Quality and Safety Engineering*, 8:351–372.

Windows Client Tech Center. 2008. *Downgrade from SP3 to SP2*. http://social.technet.microsoft.com/forums/en-US/itproxpsp/thread/5a63cf6d-f249-4cbe-8cd2-620f301ec6fc/ (accessed 2010).