Early Detection of an Underperforming Implantable Cardiovascular Device Using an Automated Safety Surveillance Tool

Robert G. Hauser, MD; Andrew S. Mugglin, PhD; Paul A. Friedman, MD; Daniel B. Kramer, MD; Linda M. Kallinen, BS; Deepa M. McGriff, BS; David L. Hayes, MD

Background—Postmarket medical device surveillance in the United States depends largely on voluntary reporting of adverse events. Consequently, early safety signals may be missed, exposing patients to potentially hazardous products. The aim of this study was to assess the feasibility of using an automated safety surveillance tool to detect early signals that a marketed implantable cardiac device was underperforming.

Methods and Results—For this purpose, we performed simulated prospective monthly full-cohort and propensity-matched comparative survival analyses on our 3-center database of Sprint Fidelis and Quattro Secure implantable cardioverter-defibrillator leads, using a commercially available automated surveillance tool that was preset to trigger an alert if the log rank probability value was $<0.05$. During the study, 84 of 1035 Fidelis (8.1%) and 23 of 1675 Quattro (1.4%) leads failed. The simulated full-cohort analysis triggered a sustained alert for Fidelis leads beginning 13 months after the first implant and 2 years before Fidelis leads were removed from the market. Of the 1035 patients who had Fidelis leads, up to 969 (93.6%) were successfully matched to Quattro patients. In the propensity-matched analysis, the alert triggered 22 months after the first Fidelis implant and more than 1 year before the lead was recalled.

Conclusions—An active automated safety surveillance system could have identified this implantable cardiovascular device problem substantially sooner than was achieved through existing postmarket surveillance methods. Such a tool, when applied to clinical registries and remote monitoring databases, may limit the exposure of large populations to underperforming and potentially hazardous cardiovascular devices. *(Circ Cardiovasc Qual Outcomes. 2012;5:189-196.)*

Key Words: defibrillation ■ pacemakers ■ complications ■ follow-up studies

The goal of postmarket surveillance is to enhance the public health by reducing the incidence of medical device adverse experiences.1 However, medical device surveillance in the United States depends largely on voluntary reporting of adverse events, and, consequently, early safety signals may be missed.2-7 Federally mandated medical device reports submitted by manufacturers to the United States Food and Drug Administration (FDA) lack denominator data and patient-specific information.8 Determining the significance of a suspected medical device performance issue may be problematic because there are no data for comparable products. Moreover, the influence of clinical or operator variables on device failure rates may not be assessed during premarket clinical trials or evaluated by postmarket studies. This has resulted in a reactive surveillance system that fails to detect significant device defects before large patient populations are exposed to potentially hazardous products.8

The advent of detailed clinical registries and remote monitoring databases, for example, CareLink (Medtronic, Inc, Minneapolis, MN), has created avenues for improving medical device surveillance. Opportunities have emerged for automated tools that prospectively monitor large device databases for early, low-frequency adverse events while implementing statistical techniques to compare suspect devices with established products that are performing well. Recently, such a tool, the Data Extraction and Longitudinal Trend Analysis (DELTA) system (Coping Systems, Inc, Cambridge, MA), was applied to the Massachusetts coronary intervention registry.9 DELTA accumulates data for multiple devices simultaneously and monitors independent data sets, using bayesian and frequentist inferential methods that have been validated on clinical trial and outcomes databases.10-12 Alerts are generated when the specified outcome (rate or value) exceeds the expectation or a probability value meets a
prespecified level. The results of the coronary registry study demonstrated the feasibility of using DELTA for identifying in-hospital medical device safety issues that were not found during premarket approval studies. However, the ability of DELTA or a similar tool to detect a postmarket medical device’s tendency to fail long before it is removed from the market (recalled) has not been shown.

The aim of this multicenter study was to determine if DELTA could detect early signals that a marketed implantable cardiovascular device was underperforming. For this purpose, we retrospectively applied DELTA to our multicenter database of Sprint Fidelis (Medtronic, Inc, Minneapolis, MN) implantable cardioverter-defibrillator (ICD) leads.13 This lead was released in September 2004 and recalled in October 2007 after 268,000 implants because it was failing at a higher than expected rate.14 The DELTA analysis recreated Fidelis survival data as they would have existed at monthly intervals and compared them with similar data for a control lead, the Quattro Secure (Medtronic, Inc, Minneapolis, MN), which has performed well since its introduction in 2001. We hypothesized that DELTA would have identified the Fidelis problem earlier than traditional surveillance techniques.

WHAT IS KNOWN

- Postmarket medical device surveillance in the United States often fails to detect low frequency adverse clinical events caused by device failure or malfunction. Thus, it fails to improve public health by reducing the incidence of medical device adverse experiences.

WHAT THE STUDY ADDS

- We show that an active automated safety surveillance tool can detect underperforming cardiovascular devices earlier than existing postmarket surveillance methods. The application of such tools to clinical registries and databases may limit the exposure of large patient populations to potentially hazardous devices.

Methods

Study Design
Patient information and lead data were collected prospectively by the 3 centers and pooled and analyzed retrospectively for the purposes of this study.

Study Population
Patients included had received a Sprint Fidelis (models 6931, 6948, and 6949) or Quattro Secure lead (model 6947) and were followed at the Minneapolis Heart Institute (Minneapolis, MN), Mayo Clinic (Rochester, MN), and the Beth Israel Deaconess Medical Center (Boston, MA). The leads were implanted between November 2001 and December 2008, and follow-up was complete through November 2009.

Techniques and Follow-Up
Lead implant techniques and the definitions used for this study have been described.13 Briefly, leads were inserted into the right ventricular apex or septum, using standard micropuncture and introducer sheath techniques. Defibrillation safety margins and pacing threshold and sensing measurements were obtained according to each center’s protocol. Patients were followed every 3–4 months in the clinic and/or by remote monitoring (Carelink). Remote monitoring was implemented with the use of a home monitor that interrogated the ICD and transmitted a rhythm electrogram and all data within the device memory by telephone to a secure Carelink network server. The information was subsequently reviewed by our centers on the Carelink website. Remote monitoring using the Carelink methodology has been shown to be comparable to clinic visits for identifying clinically important device issues.15

Definitions
As specified previously,13 a lead was implanted after the incision was closed. A lead malfunctioned if it did not meet its performance specifications or otherwise perform as intended. A lead failed if it exhibited (1) abnormal impedance; (2) electric noise as manifested by nonphysiologic signals on the electrogram or by pulse generator diagnostic data suggesting rapid oversensing, for example, nonphysiologic short intervals and/or recurrent nonsustained ventricular tachycardia with intervals usually <220 ms; or (3) if it could not sense R waves and/or provide effective electric therapy as the result of an apparent structural defect. Pseudocontacts caused by header or connector problems were identified at the time of lead revision and were not considered failures. Functional abnormalities, including exit block and physiological oversensing, were not failures. Lead dislodgment was not a lead failure unless a defect was identified. Lead failures were reviewed and adjudicated by a coinvestigator at each center. Leads removed from service were classified according to the Heart Rhythm Society’s recommendations.16

Automated Safety Surveillance System
The DELTA system is capable of supporting multiple simultaneous device-specific analyses that track the real-time cumulative performance of multiple devices from large remote databases. DELTA contains tools that join multiple related data sets and performs prospective analyses, including propensity matching, risk-adjusted cumulative outcomes, and survival analysis methods such as log rank tests and Kaplan-Meier survival curve estimation. The system is flexible and can be configured to trigger alerts at various levels of statistical significance, based on the desired sensitivity for a particular device or outcome. DELTA can be configured to automatically generate an e-mail notification to the analyst or data center when a prespecified alert occurs.

For this study, Kaplan-Meier survival estimates were used to create survival curves for Fidelis and Quattro leads, which served as the control group. The Mantel-Cox log rank test was then applied to compare Fidelis and Quattro lead survivals. The DELTA alert was preset to trigger when the log rank probability value was <0.05. Repeated tests at significance level of 0.05 can lead to an inflated chance of false-positive alerts, but this was judged to be less consequential than false-negatives for a lifesaving device. Further, appropriate statistical methods to control false-positive rates generally require prespecifying a maximal sample size or event count, after which the null hypothesis of equally safe devices will be accepted if no boundary has been crossed; in the monitoring setting where these are unknown, the appropriate alpha adjustment is unclear.

Simulated Prospective Analysis
To determine if and when an alert based on lead failures would have been triggered for Fidelis leads, data sets were recreated as they would have existed on the last day of each month from Fidelis release in September 2004 to December 2009, a period including Fidelis removal from the market in October 2007. All patients implanted on or before the monthly cutoff date were included in the data set, as were all follow-up visits that had occurred by that time. Monthly survival curves were then created and compared.
Risk Adjustment Through Propensity Matching

Propensity score matching was used to control for possible differences in Fidelis and Quattro patient populations. At every given month’s analysis, a propensity-matched control (Quattro lead patient) was sought for each exposure (Fidelis lead patient). Any subject implanted before the end of the month was considered in the analysis data set and was eligible to be matched. Each patient’s propensity to be implanted with a Fidelis lead was calculated through logistic regression on the basis of the following clinical predictors: age and left ventricular ejection fraction as continuous variables; implant center, sex, previous cardiac surgery (yes or no), ICD indication (primary or secondary), and underlying cardiac disease (ischemic heart disease, dilated cardiomyopathy, hypertrophic cardiomyopathy, arrhythmogenic right ventricular dysplasia and channelopathies, and ventricular tachycardia/ventricular fibrillation) as categorical variables. These predictors appeared linearly in the regression model, without interactions. Values for these predictors were known at baseline and thus did not vary during the study period; however, because the available data on which propensity scores were calculated tended to grow over time, the propensity scores themselves changed from month to month as they were reestimated from the accruing data.

Each Fidelis patient was matched to the control Quattro patient with the nearest propensity score within a maximum caliper of 0.05 by means of a greedy matching algorithm. To continue matching to the nearest propensity score as the monthly data sets grew, at the end of each month’s analysis all matches were released, and a completely new matching was conducted at the next month’s analysis. Control subjects were not allowed to match to more than 1 exposure. Such matching without replacement is not invariant to the ordering of patients in the data set, so the ordering of patients was randomized before each month’s match procedure. Propensity scores could not be calculated for 87 (3.2%) of 2710 patients because baseline covariates were missing; these subjects were not eligible for matching in any match-based analysis. These comprised 28 Fidelis leads, including 6 failures, and 59 Quattro leads, including 1 failure.

Match quality was assessed by visually inspecting histograms of propensity scores as well as the prematch and postmatch standard difference calculations available in DELTA. To verify software accuracy, monthly data set creation and statistical analysis were duplicated by the authors in the R statistical language, with propensity matching implemented by means of the “Matching” add-on package of Sokhon.25

Because propensity matching is not invariant to the ordering of patients in the data set, so the entire analysis from January 2005 to December 2009 was iterated 1000 times, each time with different randomized orderings of patients. Items tracked included the first month in which P<0.05 and the first time 3 consecutive months resulted in P<0.05. The impact of repeated significance testing on the rate of false-positive findings was similarly explored by randomly permuting the Quattro/Fidelis “assignments” within our data (creating data with no systematic differences between treatment groups) and rerunning the entire 60-month sequence of analyses 1000 times, similarly tracking the months when P<0.05.

Results

A total of 2710 patients received a Fidelis (n=1035) or Quattro (n=1675) lead (Table). Compared with Fidelis patients, more Quattro patients were women and more had ischemic heart disease and received an ICD for secondary prevention. The 3 centers differed significantly in the proportion of Fidelis and Quattro leads they implanted. During the study, 84 Fidelis (8.1%) and 23 Quattro leads (1.4%) failed.

Simulated Prospective Full-Cohort Survival Analyses

Simulated sequential unmatched monthly survival analyses on the full cohort of patients were performed beginning in January 2005, or 4 months after the first Fidelis was implanted at one of our centers in September 2004 and 4 years after the first Quattro implant in November 2001 (Figure 1). By January 2005, 95 Fidelis implants had occurred without a failure, and 848 Quattro leads had been implanted with a single failure at 18 months. One year after the first Fidelis implant (September 30, 2005; Figure 1), none of 382 Fidelis leads had failed and the second of 950 Quattro leads had failed at 36 months. In the following month (October 31, 2005; Figure 1), a single Fidelis failure occurred 13 months after implant; this event triggered a DELTA alert (P<0.001). The DELTA alert was sustained for all subsequent monthly survival analyses.

Simulated Prospective Propensity-Matched Survival Analyses

Simulated sequential propensity-matched monthly survival analyses were then performed by matching a patient who received a Fidelis lead to a Quattro patient who was similar in age, sex, left ventricular ejection fraction, indication, underlying heart disease, prior heart surgery, and implant center (Figure 2). Of the 1035 patients who had Fidelis leads, up to 969 (93.6%) were successfully matched to Quattro patients (matching percentage varied by month). In this propensity-matched analysis, the DELTA alert triggered in July 2006 and persisted until January 2007. During 2007, the alert returned in June and persisted during the remaining months of the study.

Because Quattro patients were not matched to more than 1 Fidelis patient for any given month’s analysis (matching without replacement), the order in which matching occurred could affect the results that are shown in Figure 2. Thus, the
entire simulation was performed 1000 times, each time using different randomized orderings of patients before matching. The results of the 1000 simulations are plotted in Figure 3. During the 4th quarter of 2005 and the first 5 months of 2006, the DELTA alert occurred in 12–26% of simulations. Beginning in June 2006, the DELTA alert was triggered in all subsequent months in >60% of simulations except for December 2006 and January 2007, when the alert occurred in 43% and 53%, respectively.

The distributions of the timing of the DELTA propensity-matched alerts are displayed in Figure 4A. The first alerts are clustered during the last quarter of 2005 and the 2nd quarter of 2006. The second quarter of 2006 was also the time when the alert was most likely to be activated for 3 consecutive months.

We investigated the impact of different DELTA alert probability value thresholds (0.001, 0.005, and 0.01–0.10). Among these 1000 instances, the median month of first alert is constant (January 2006), regardless of our choice of threshold.

To evaluate false-positive alerts, we randomly reassigned lead models to observations and conducted unadjusted analyses to determine whether and when alerts were triggered. The results are displayed in Figure 4B. Using the

**Figure 1.** Simulated monthly prospective unmatched full-cohort survival analyses comparing Sprint Fidelis with the control Quattro Secure leads. QS indicates Quattro Secure; SF, Sprint Fidelis.
study threshold of 0.05, 46% of 1000 simulations yielded at least 1 first alert, but only 8.2% yielded a triple-consecutive alert. For a threshold of 0.01, the results were 11.2% for at least 1 alert and 1.4% for a triple-consecutive alert.

Discussion
The results of this multicenter study suggest that an active automated surveillance system could have identified this cardiovascular device problem substantially sooner than was achieved through existing postmarket surveillance methods. The initial report of Fidelis lead failure was published in July 2007, and the manufacturer did not remove the device from the market until October 15, 2007, when the FDA announced a voluntary recall. At that time, there had been approximately 268,000 Fidelis leads implanted worldwide. Thus, by implication, an alert in 2006 or early 2007 could have spared thousands of patients the consequences of receiving a device prone to failure, including the risks and costs of lead replacement. This finding is novel and important because it demonstrates the potential value of using robust analytic techniques to detect low-frequency events involving newly approved medical devices as they are being introduced into large-scale use. If our simulation is supported by pro-

Figure 2. Simulated monthly prospective propensity-matched survival analyses comparing Sprint Fidelis with the control Quattro Secure leads. QS indicates Quatto Secure; SF, Sprint Fidelis.
spective studies using active databases, a significant gap in the postmarket surveillance of medical devices can be closed by the application of tools such as DELTA to national clinical registries and data warehouses.

Our study used a multicenter database, which was created to assess the long-term performance of the Fidelis ICD lead and the clinical risk factors associated with its failure. Recently, we reported that the risk of Fidelis fracture was higher in younger patients, women, and individuals with hereditary heart disease and that the hazard of Fidelis fracture increased with every percent increase in left ventricular ejection fraction.\textsuperscript{13} None of these variables were associated with the risk of Fidelis fracture.

---

**Figure 3.** Frequency of Data Extraction and Longitudinal Trend Analysis alerts ($P<0.05$) by month during 1000 simulations of propensity-matched survival analyses comparing Sprint Fidelis leads with control Quattro Secure leads.

**Figure 4.** A, Timing of Data Extraction and Longitudinal Trend Analysis (DELTA) alerts during 1000 propensity-matched simulations of Fidelis and Quattro lead survivals. B, Timing of DELTA alerts during 1000 propensity-matched simulations of Fidelis and Quattro lead survivals when leads were deidentified.
with Quattro Secure failure, and the low failure rate of Quattro leads remained consistent over 7 years of follow-up. Thus, we used the Quattro as the control lead in this study and calculated propensity scores based on clinical variables, which were then matched to patients in the Fidelis population. Propensity score matching was used to reduce the impact of treatment selection bias in this multicenter database. Of the various matching approaches available, we chose to match a Fidelis lead patient to a similar Quattro lead patient once, that is, matching without replacement, to preserve the independence of the control Quattro lead population. After 1000 simulations, we found that the frequency of the DELTA alerts increased progressively from late 2005 to mid-2007.

The first DELTA alert in our full-cohort unmatched simulation occurred in October 2005, or 13 months after the Fidelis was market released. At this juncture, there was 1 Fidelis failure among 419 implants. Notably, all subsequent monthly unmatched survival analyses triggered the DELTA alert. However, the likely reaction to the early alerts would have been additional studies, for example, propensity-matched analyses with multiple randomized simulations using a larger cohort. We speculate that by early 2007, Fidelis leads would either have been removed from the market or infrequently used. Indeed, as the Fidelis failure trend emerged in 2006, it is probable that many physicians would have opted to limit their use of Fidelis leads. This proactive approach to postmarket surveillance has the key attributes of timeliness and scientific rigor while minimizing the chance that clinical or operator variables could account for the observed failures.

Nevertheless, caution should be exercised when interpreting the alerts; because the trigger may be preset to be highly sensitive, for example, $P<0.05$, false-positive alerts could occur as we demonstrated. For this reason, an initial alert should not precipitate an immediate regulatory action or recall or even a change in practice while confirmatory studies and analyses are underway, assessing for example different triggering thresholds or potential confounding influences such as differential follow-up or competing risks. Indeed, the first trigger may be regarded as hypothesis-generating.

Although we chose a high-sensitivity alert for this critical device, other investigators and stakeholders can certainly debate and explore different balances of alert sensitivity and specificity tailored to the device or outcome in question. Certain situations (such as life-threatening acute complications) may call for highly sensitive approaches (with alerts issued for even very small differences between the devices being investigated), whereas other events may warrant far more specific thresholds (such as an increase in minor complications for medium- to low-risk devices). Methods that strictly control false-positive rates are available if investigators are willing to prespecify the maximal amount of information and (in some methods) an $\alpha$-spending plan.

The feasibility of performing computerized automated prospective surveillance of large clinical outcomes registries was demonstrated by Resnic et al. who applied DELTA to a database of >74,000 interventional coronary procedures contained in the Massachusetts angioplasty registry. Using a variety of statistical methods, including propensity score matching, DELTA triggered a sustained alert for a specific drug-eluting stent because patients receiving this particular model were having a higher rate of periprocedural myocardial infarction than matched control patients treated with another market-released drug-eluting stent. However, our study is unique because it is based on multicenter long-term follow-up and survival analyses, and the stent identified in their report has not been recalled.

Large registries and remote monitoring databases are accumulating patient- and device-specific data to improve outcomes. The National Cardiovascular Data Registry (NCDR) gathers data from >2200 hospitals in the United States. The NCDR databases include clinical information for a variety of cardiovascular procedures and devices, including ICDs and stents. CareLink is a proprietary remote monitoring service that acquires cardiac rhythm management device data that the manufacturer uses to report product performance information. Our study suggests that the implementation of real-time tools such as DELTA, which are capable of analyzing large data sets using sophisticated statistical techniques, could identify low-frequency adverse events in registries like those in the NCDR and large device-specific databases such as CareLink. If this were successful, it may be possible to automatically detect, interpret, and analyze safety signals effectively and efficiently. Such an accomplishment would remove a major deficiency in the current postmarket surveillance system. The implementation of DELTA or a similar safety surveillance tool has certain challenges that will need to be addressed, including statistical oversight to ensure, for example, propensity score balance. Moreover, although we used the Quattro lead as our gold standard for comparative purposes, real-time postmarket monitoring of multiple devices will require a comparative device or devices that are performing acceptably to physicians and patients.

In our simulation of 1000 iterations, the dependence of the timing of the first alert on the randomized ordering of patients raises the question of how best to conduct a propensity-matched analysis. We suggest that any automatic monitoring approach using propensity matching without replacement iterate each analysis as we did; such an approach is practical and inexpensive, and it may unmask a significant safety issue.

This study has certain limitations. The information contained in the lead database was collected retrospectively, although each center enters the data into their electronic medical records prospectively. In most cases, lead failure could not be verified by engineering analyses either because the complete lead was not removed and returned to the manufacturer or the analysis had not been completed. The DELTA alert was preset for a probability value $<0.05$; this or another arbitrary safety signal should not be viewed as a definitive warning sign but rather hypothesis-generating and warranting comprehensive examination. False-positive alerts can occur. Despite propensity matching, other confounding factors may have influenced our results. This method works best in multicenter data environments because overlap of patient covariates across different hardware and implanters is crucial to the robustness of the results. Moreover, a large sample size for follow-up is also required so that estimates do not become unstable with time.
In conclusion, an active automated safety surveillance tool such as DELTA can detect underperforming cardiovascular devices earlier than existing postmarket surveillance methods. This observation suggests that the application of such tools to clinical registries and databases may limit the exposure of large patient populations to potentially hazardous devices.

Sources of Funding

This study was funded by research grants from the Minneapolis Heart Institute Foundation, Minneapolis, MN, and the Abbott Northwestern Hospital Foundation, Minneapolis, MN.

Disclosures

Dr Mugglin is a statistical consultant for Medtronic and St Jude Medical. Dr Hayes is a speaker with honoraria for Medtronic, Boston Scientific, St Jude Medical, Biotronik, and Sorin; he serves on advisory boards for Boston Scientific, St Jude Medical, and Pixel Velocity; and he is on steering committee or study advisory committee for Medtronic and St Jude Medical. Dr Friedman is a speaker or consultant with honoraria for Medtronic, Boston Scientific, St Jude Medical, and Bard; he receives significant research support from Medtronic and Pfizer; and he has intellectual property rights with Bard EP, Hewlett Packard, Medical Positioning Inc, Aegis Medical, and NeoChord.

References

Early Detection of an Underperforming Implantable Cardiovascular Device Using an Automated Safety Surveillance Tool

Robert G. Hauser, Andrew S. Mugglin, Paul A. Friedman, Daniel B. Kramer, Linda Kallinen, Deepa McGriff and David L. Hayes

Circ Cardiovasc Qual Outcomes. 2012;5:189-196; originally published online March 6, 2012; doi: 10.1161/CIRCOUTCOMES.111.962621

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/5/2/189

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org/subscriptions/