

## Research Paper ■

## Monitoring Device Safety in Interventional Cardiology

MICHAEL E. MATHENY, MD, LUCILA OHNO-MACHADO, MD, PhD, FREDERIC S. RESNIC, MD, MS

**Abstract Objective:** A variety of postmarketing surveillance strategies to monitor the safety of medical devices have been supported by the U.S. Food and Drug Administration, but there are few systems to automate surveillance. Our objective was to develop a system to perform real-time monitoring of safety data using a variety of process control techniques.

**Design:** The Web-based Data Extraction and Longitudinal Time Analysis (DELTA) system imports clinical data in real-time from an electronic database and generates alerts for potentially unsafe devices or procedures. The statistical techniques used are statistical process control (SPC), logistic regression (LR), and Bayesian updating statistics (BUS).

**Measurements:** We selected in-patient mortality following implantation of the Cypher drug-eluting coronary stent to evaluate our system. Data from the University of Michigan Consortium Bare-Metal Stent Study was used to calculate the event rate alerting boundaries. Data analysis was performed on local catheterization data from Brigham and Women's Hospital from July 1, 2003, shortly after the Cypher release, to December 31, 2004, including 2,270 cases with 27 observed deaths.

**Results:** The single-stratum SPC had alerts in months 4 and 10. The multistrata SPC had alerts in months 5, 10, and 18 in the moderate-risk stratum, and months 1, 4, 7, and 10 in the high-risk stratum. The only cumulative alerts were in the first month for the high-risk stratum of the multistrata SPC. The LR method showed no monthly or cumulative alerts. The BUS method showed an alert in the first month for the high-risk stratum.

**Conclusion:** The system performed adequately within the Brigham and Women's Hospital Intranet environment based on the design goals. All three cumulative methods agreed that the overall observed event rates were not significantly higher for the new medical device than for a closely related medical device and were consistent with the observation that the initial concerns about this device dissipated as more data accumulated.

■ *J Am Med Inform Assoc.* 2006;13:180–187. DOI 10.1197/jamia.M1908.

Minimizing harm to patients and ensuring their safety are cornerstones of any clinical research effort. Safety monitoring is important in every stage of research related to a new drug, new medical device, or new therapeutic procedure. This type of monitoring of medical devices, under the auspices of the U.S. Food and Drug Administration (FDA), has undergone major changes over the past several decades.<sup>1–4</sup> These changes have largely been due to a small number of highly publicized adverse events.<sup>5–13</sup> The FDA's task is complex; the agency regulates more than 1,700 types of devices,

500,000 medical device models, and 23,000 manufacturers.<sup>3,6,14–18</sup> In premarketing clinical trials, rare adverse events may not be discovered due to small sample sizes and biases toward healthier subjects.<sup>19</sup> The FDA must balance this concern with the need to deliver important medical advances to the public in a timely fashion. In response to this, the FDA has shifted some of its device evaluation to the post-market period, allowing new devices to reach the market sooner.<sup>20</sup> This creates the potential for large numbers of patients to be exposed to a new product in the absence of long-term follow-up data and emphasizes the need for careful and thorough postmarketing surveillance.<sup>21</sup>

The current FDA policies in this area include a heterogeneous mix of voluntary and mandatory reporting.<sup>1,6,14,15,17,19,22–24</sup>

Voluntary reporting of adverse events creates limitations in significant event-rate recognition through underreporting bias, and highly variable reporting quality.<sup>25</sup> Several state and federal agencies have implemented mandatory reporting for medical devices for specific clinical areas, and national medical societies are making strides to standardize data element definitions and data collection methods within their respective domains.<sup>26,27</sup> Continued improvements in the quality and volume of reported data have created opportunities for timely and efficient analysis and reporting of alarming trends in patient outcomes.

Nonmedical industries (Toronado, HGL Dynamics, Inc., Surrey, GU; WinTA, Tensor PLC, Great Yarmouth, NR) have been using a variety of automated statistical process control

Affiliations of the authors: Decision Systems Group, Department of Radiology, Harvard Medical School, Boston, MA (MEM, LOM, FSR), Harvard-MIT Division of Health Sciences & Technology, Massachusetts Institute of Technology, Cambridge, MA (MEM, LOM, FSR).

This research was supported in part by grants 1-T15-LM-07092 and R01-LM08142-01 from the National Library of Medicine of the National Institutes of Health.

The authors thank Richard Cope and Barry Coflan for their excellent programming work on the project, Dr. Robert Greenes for editing assistance, and Anne Fladger and her staff for literature procurement assistance.

The authors have no conflict of interests to disclose.

Correspondence and reprints: Frederic S. Resnic, MD, MS, Brigham & Women's Hospital, 75 Francis Street, Boston, MA 02115; e-mail: <fresnic@partners.org>, Michael E. Matheny, MD; e-mail: <mmatheny@dsgh.harvard.edu>.

Received for review: 07/07/05; accepted for publication: 10/16/05.

(SPC) techniques for quality control purposes for many years.<sup>28-30</sup> These systems rely on automated data collection and use standard SPC methods of varying rigor.<sup>31</sup> However, automated SPC monitoring has not been widely deployed in the medical domain due to a number of constraints: (a) historically, automated data collection could usually only be obtained for objective data such as laboratory results and vital signs; (b) much of the needed information about a patient's condition is subjective and may be available only in free text in the medical record; and (c) medical source data, due to heterogeneity of clinical factors, typically has more noise than industrial data, and standard industrial SPC metrics may not be directly applicable to medical safety monitoring.

Within the medical domain, the most closely related clinical systems that have been developed to date are those in clinical trial monitoring for new pharmaceuticals. A variety of software solutions (Clinitrace, Phase Forward, Waltham, MA; Oracle Adverse Event Reporting System [AERS], Oracle, Redwood Shores, CA; Trialex, Meta-Xceed, Inc., Fremont, CA; Netregulus, Netregulus, Inc., Centennial, CO) have been created to monitor patient data relevant to these trials. These systems rely on standard SPC methodologies and can provide real-time data monitoring and analysis through internal data standardization and collection for the trial. However, the focus of these systems is on real-time data aggregation and reporting to the FDA.

The increasing availability of detailed electronic medical records and structured clinical outcomes data repositories may provide new opportunities to perform real-time surveillance and monitoring of adverse outcomes for new devices and therapeutics beyond the clinical trial environment. However, the specific monitoring methodologies that balance appropriate adverse event detection sensitivity and specificity remain unclear.

In response to this opportunity, we have developed the Data Extraction and Longitudinal Time Analysis (DELTA) system and explored both standard and experimental statistical techniques for real-time safety monitoring. A clinical example was chosen to highlight the functionality of DELTA and to provide an overview of its potential uses. Interventional cardiology was chosen because the domain has a national data field standard,<sup>32</sup> a recent increase in mandatory case reporting from state and federal agencies, and recent device safety concerns publicized by the FDA.

## Methods

### System General Requirements

The DELTA system was developed to provide real-time monitoring of clinical data during the course of evaluating a new medical device, medication, or intervention. The system was designed to satisfy five principal requirements. First, the system should accept a generic data set, represented as a flat data table, to enable compatibility with the broadest possible range of sources. Second, the system should perform both prospective and retrospective analyses. Third, the system should support a variety of classical and experimental statistical methods to monitor trends in the data, configured as analytic modules within the system, allowing both unadjusted and risk-adjusted safety monitoring. In addition, the system should support different methodologies for alerting

the user. Finally, DELTA should support an arbitrary number of simultaneous data sets and an arbitrary number of ongoing analyses within each data set. That is, DELTA should "track" multiple outcomes from multiple data sources simultaneously, thus making it possible for DELTA to serve as a single portal for safety monitoring for multiple simultaneous analyses in an institution.

### Source Data and Internal Data Structure

A flat file representation of the covariates and clinical outcomes serves as the basis for all analyses. In addition, a static data dictionary must be provided to DELTA to allow for parsing and display of the source data in the user interface. Necessary information includes whether each field is going to be treated as a covariate or an outcome and whether it is discrete or continuous.

The system uses an SQL 2000 server (Microsoft Corp., Redmond, WA) for internal data storage, importing all clinical data and data dictionaries from source databases at regular time intervals. This database also stores system configurations, analysis configurations, and results that are generated by DELTA at the conclusion of a given time period. The user interface is Web-based and uses a standard tree menu format for navigation. DELTA's infrastructure and external linkages are shown in Figure 1.

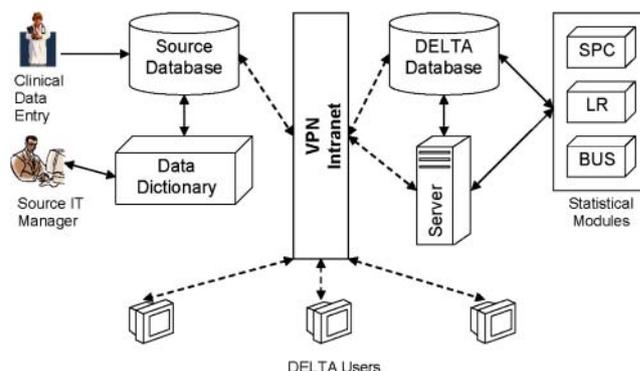
Security of patient data is currently addressed through record de-identification steps<sup>32</sup> performed to the fullest extent possible while maintaining the necessary data set granularity for the risk adjustment models. The system is hosted on the Partners Healthcare Intranet, a secure multihospital network, accessible at member sites or remotely through a virtual private network.

### Statistical Methods

DELTA uses a modular approach to statistical analysis that facilitates further expansion. DELTA currently supports three statistical methodologies: statistical process control (SPC), logistic regression (LR), and Bayesian updating statistics (BUS). Discrete risk stratification is supported by both SPC and BUS. Periodic and cumulative analysis of data is supported by SPC and LR, and only cumulative analysis is supported by BUS.

### Risk Stratification

Risk stratification is a process by which a given sample is subdivided into discrete groups based on predefined criteria.



**Figure 1.** Overall DELTA infrastructure and an example external data source. IT, information technology; VPN, virtual private network; SPC, statistical process control; LR, logistic regression; BUS, Bayesian updating statistics.

This process is used to allow providers to quickly estimate the probability of an outcome for a patient. Statistically, the goal of this process is to create a meaningful separation in the data to allow concurrent and potentially different analyses to be performed on each subset. Criteria are selected based on prior data, typically derived from a logistic regression predictive model, and the relative success of this stratification can be determined by a stepwise increase in the incidence of the outcome in each risk group. The LR method does not offer discrete risk stratification because it incorporates risk stratification on a case level.

#### *Data Aggregation*

Retrospective data analyses traditionally use the entire data set for all calculations. However, in real-time data analysis, it is of interest to monitor both recent trends and overall trends in event rates. Evaluation of recent trends will intrinsically have reduced power, because of the reduced sample size, to detect true, significant shifts in event rates. However, such monitoring may serve as a very useful "first warning" indicator when the cumulative event rate may not yet cross the alerting threshold. This type of alert is not considered definitive but can be used to encourage increased monitoring of the intervention of interest and heighten awareness of a potential problem. In DELTA, these recent data analyses are termed *periodic* and can be configured to be performed on a monthly, quarterly, or yearly basis.

#### *Statistical Process Control*

Statistical process control is a standard quality control method in nonmedical industrial domains. This method compares observed event rates to static alerting boundaries developed from previously published or observed empirical data. Each industry typically requires different levels of rigor in alerting, and selection of confidence intervals (CIs) (or number of standard errors) establishes this benchmark. In the medical industry, the 95% CI is considered to be the threshold of statistical improbability to establish a *true* difference. In DELTA, the 95% CI of proportions by the Wilson method is used to calculate the alerting boundaries for all statistical methods.<sup>33</sup> The proportion of observed events is then compared to these static boundaries, and alerts are generated if they exceed the upper CI boundary. DELTA's SPC module is capable of performing event rate monitoring on multiple-risk strata provided that criteria for stratification and benchmark event rates are included for each risk stratum. This method supports comparison of benchmark expected event rates with cumulative and periodic observed event rates.

While simple and intuitive, the SPC methodology does not support case-level risk adjustment. It is also dependent on accurate benchmark data, which may be limited for new procedures or when existing therapies are applied to new clinical conditions.

#### *Logistic Regression*

Logistic regression<sup>34</sup> is a nonlinear modeling technique used to provide a probability of an outcome on a case-level basis. Within DELTA, the LR method allows for continuous risk-adjusted estimation of an outcome at the case level. The LR model must be developed prior to the initiation of an analysis within DELTA and is mostly commonly based on previously published and validated models.

Alerting thresholds are established by using the LR model's expected mortality probability for each case. These probabilities are then summated in both periodic and cumulative time frames to determine the 95% CI of the event rate proportion by the Wilson method. Alerts are generated if the observed event rate exceeds the upper bounds of the 95% CI of a given boundary. This method provides accommodation for high-risk patients by adjusting the alerting boundary based on the model's expected probability of death. This can be very useful when outcome event rates vary widely with patient comorbidities. A limitation of this method is that the alerts become dependent on the discrimination (measure of population prediction accuracy) and calibration (measure of small group or case prediction accuracy) of that model.

#### *Bayesian Updating Statistics*

Bayesian updating statistics is an experimental methodology pioneered in non-health care industries.<sup>35</sup> This method incorporates Bayes' theorem<sup>36</sup> into a traditional SPC framework by using prior observed data to evolve the estimates of risk. Alerting boundaries are calculated by two methods, both of which are considered cumulative analyses only. The first method includes previous current study data with the prior data used in the SPC method to calculate the 95% CI of the event rate proportion by the Wilson method. This means that the alerting boundary shifts during the course of real-time monitoring due to the influence of the earlier study data.

The other alerting method is based on the evolution of the updated risk estimates represented as probability density functions (PDFs). In each period, a new PDF is generated based on the cumulative study event rate and baseline event rate. Alerting thresholds are generated by the user specifying a minimum percentage of amount of overlap of the two distributions (by comparison of central posterior intervals).<sup>37</sup> The first comparison PDF is the initial prior PDF, and the second is the previous period's PDF. BUS supports discrete risk stratification.

This method was included in DELTA because it tends to reduce the impact of early outliers in data and complements the other monitoring methods used in the system. It also may be particularly helpful in situations in which limited preexisting data exist. However, the method is dependent on accurate risk strata development and on the methods used for weighting of the prior data in the analysis.

#### **User Interface**

The user interface is provided via a Web browser and was developed in the Microsoft.NET environment, running Microsoft IIS 5.0 Web Server (Microsoft Corp.). Each data set is represented as a separate folder on the main page, and all analyses for that set are nested under that folder (Fig. 2). At the initiation of an analysis, the user designates the analysis period and starting and stopping dates and selects the statistical module and the outcome of interest. Data filters can be applied to restrict the candidate cases for analysis. Covariates used for risk stratification are selected. Last, periodic and cumulative alerts for the statistical method selected can be activated or suppressed based on user preferences. An analysis configuration can be duplicated and modified for convenience in configuring multiple statistical methods to concurrently monitor a data source.



**Figure 2.** DELTA screenshot showing the results menu screen of the statistical process control (SPC) clinical example described in text. The main menu is displayed on the left, and the analysis menu is displayed above the viewing area.

The results screen of DELTA serves as the primary portal to all tables, alerts, and graphs generated from an analysis. Tabular and graphical outputs of the data and specific alerting thresholds by risk strata are available, and an export function is included to allow researchers to perform further evaluation of the data.

**Clinical Example**

As an example of the application of DELTA to real-world data, an analysis of the in-hospital mortality following the implantation of a drug-eluting stent was performed. The cardiac catheterization laboratory of Brigham and Women’s Hospital has maintained a detailed clinical outcomes database since 1997 for all patients undergoing percutaneous coronary intervention, based on the American College of Cardiology National Cardiovascular Data Repository data elements.<sup>32</sup>

For risk stratification, the University of Michigan risk prediction model<sup>38</sup> was used since it provides a concise method of comparing all three of DELTA’s statistical methods using one reference for previous experience. The previous experience of event rates for all risk strata from this work is listed in Appendix 1. A logistic regression model with risk stratification scores is listed in Appendix 2. The logistic regression model developed from the data was used to create a discrete risk scoring model. Based on the mortality of patients in the study sample at various risk scores, these data were divided into three discrete risk categories, and the compositions of those categories are listed in Appendix 2.

A total of 2,270 drug-eluting stent cases were performed from July 1, 2003, to December 31, 2004, at our institution, and the outcome in terms of in-hospital mortality was analyzed. These data were retrospectively evaluated in monthly periods for each of the three statistical methodologies. There was a total of 27 observed deaths (unadjusted mortality rate of 1.19%) during the study. Local institutional review board approval was obtained. Risk stratification of these cases by the University of Michigan model is listed in Table 1 and demonstrates increasing in-hospital mortality risk with 0%, 0.9%, and 23% mortality risk in the low-, medium-, and high-risk strata, respectively.

An alternative data set was generated by taking the clinical data above and changing the procedure date from the eight cases with the outcome of interest in the last five periods.

**Table 1** ■ Multiple Risk Strata SPC Statistical Process Control

Risk Stratum	Sample	Events	Events Rate
Low	641	0	00.00%
Moderate	1573	14	00.89%
High	56	13	23.21%

The procedure dates were changed by random allocation into one of the first 13 periods. The duration of the monitoring was then shortened to 13 periods. This was done to illustrate alerts when cumulative event rates clearly exceeded established thresholds. The overall event rate for this data set is 1.71% (27/1,583), and the risk stratified event rates were 0% (0/446), 1.3% (14/1,095), and 31% (13/42) for the low-, medium-, and high-risk strata, respectively.

**Results**

**Statistical Process Control**

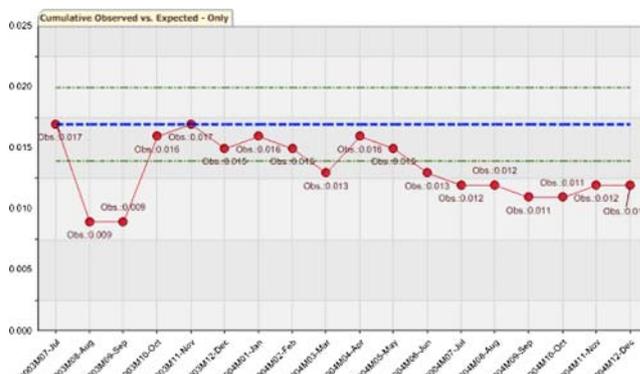
*Single Risk Stratum*

The single risk stratum SPC was configured with no risk stratification covariates. The static alert boundary was a 2.07% (upper 95% CI of 100/5,863). Periodic evaluations ranged from 0% to 4.5%. Period 4 exceeded the boundary with a 3.4% (5/148) event rate and period 10 with a 4.5% (5/110) event rate. Cumulative event rates ranged from 0.9% (2/213) to 1.7% (10/587). No cumulative evaluations had an event rate that exceeded the boundary. The cumulative evaluation is depicted graphically in Figure 3.

Periodic evaluations of the alternative data set ranged from 0% to 4.7%. Period 4 exceeded the boundary with a 4.7% (7/150) event rate, period 7 with a 2.6% (3/117) event rate, period 8 with a 2.6% (3/117) event rate, and period 10 with a 4.5% (5/110) event rate. Cumulative event rates ranged from 0.9% (2/213) to 2.4% (12/490). Periods 4 through 11 had event rates exceeding the 2.07% threshold and generated alerts and ranged from 2.1% to 2.4%.

*Multiple Risk Strata*

Alerting thresholds were calculated for the low-, medium-, and high-risk strata by using the upper 95% CI of the proportion of the event rates of each stratum in the University of



**Figure 3.** Single-stratum statistical process control (SPC) graph showing the cumulative observed event rates versus the static alerting threshold (expected rates) with 95% confidence intervals.

Michigan data. The thresholds were 0.3% (1/1,820), 1.7% (50/3,907), and 44% (49/136), respectively.

There were no events in the low-risk stratum, and no alerts were generated. In the moderate-risk stratum, the periodic observed event rates ranged from 0% to 2.7%. The alerting boundary was exceeded with rates of 2.7% (2/75) in period 5, 2.6% (2/78) in period 10, and 1.9% (2/108) in period 18. The cumulative observed event rates ranged from 0.7% to 1.3% and never exceeded the upper alert boundary. In the high-risk stratum, the periodic observed event rates ranged from 0% to 100%. The alerting boundary was exceeded with rates of 100% in periods 1 (1/1), 7 (1/1), and 10 (3/3), and by a rate of 50% (4/8) in period 4. The cumulative observed event rates ranged from 16.7% to 100%. The alerting boundary was exceeded by a rate of 100% (1/1) in period 1.

Evaluation of the alternative data set was performed periodic and cumulative alerts. There were no events in the low-risk stratum, and no alerts were generated. In the moderate-risk stratum, the periodic observed event rates ranged from 0% to 3.6%. The alerting boundary was exceeded with rates of 3.6% (3/84) in period 3, 2.1% (2/95) in period 4, 2.7% (2/75) in period 5, 2.5% (2/81) in period 7, and 2.6% (2/78) in period 10. The cumulative observed event rates ranged from 0.7% to 2.0% and exceeded the alerting boundary in periods 3 through 11. In the high-risk stratum, the periodic event rates ranged from 0% to 100%. The alerting boundary was exceeded with rates of 100% in periods 1 (1/1), 7 (1/1), and 10 (3/3) and by a rate of 55.6% (5/9) in period 4. The cumulative observed event rates ranged from 16.7% to 100%. The alerting boundary was exceeded by a rate of 100% (1/1) in period 1.

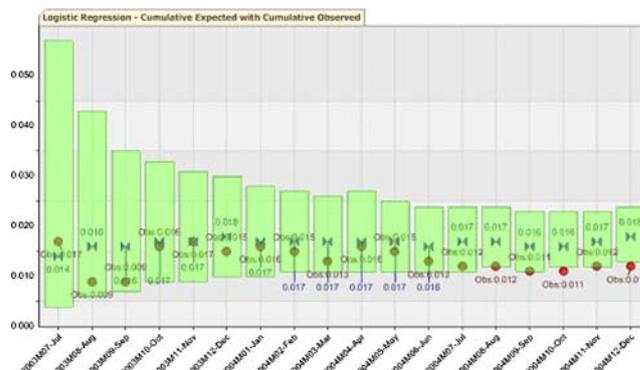
### Logistic Regression

Alerting thresholds were calculated on a periodic basis using the expected probability of death for the cases in their respective periods, and the 95% upper CI ranged from 4.9% (1.02/112) to 7.1% (2.51/110). Cumulative-based upper alerting boundaries ranged from 2.3% (29.86/1,835) to 5.7% (1.66/115). The overall expected cumulative event rate was 1.75% (39.7/2,270).

Periodic event rates ranged from 0% to 4.5%, and no alerts were generated. The two highest periodic event rates of 3.4% (5/148) in period 4 and 4.5% (5/110) in period 10 had upper alerting boundaries of 5.8% (2.98/148) and 7.1% (2.51/110), respectively. Cumulative event rates ranged from a 0.9% (2/213) to 1.7% (10/587) event rate, and the cumulative upper 95% CI was well above the observed event rate throughout the evaluation and shown in Figure 4.

Alerting thresholds for the alternative data set based on the upper 95% CI ranged from 4.9% (1.02/112) to 7.5% (3.2/117) in the periodic analysis and from 2.6% (28.17/1,583) to 5.7% (1.66/115) in the cumulative analysis. The overall expected event rate was 1.78% (28.17/1,583).

Periodic event rates ranged from 0% to 4.7%, and no alerts were generated. The two highest periodic event rates of 4.7% (7/150) in period 4 and 4.5% (5/110) in period 10 had upper alerting boundaries of 6.2% (3.56/150) and 5.1% (2.51/110). Cumulative event rates ranged from 0.9% (2/213) to 2.4% (12/490) and were well below the alert boundaries through the evaluation.



**Figure 4.** Logistic regression graph showing the cumulative observed event rate versus the cumulative expected event rate with 95% confidence interval.

### Bayesian Updating Statistics

The upper alert boundary varied from 0.2% to 0.3%, from 1.5% to 1.7%, and from 40.2% to 44.4% in the low-, medium-, and high-risk strata, respectively. From all strata, the only alert generated was in the high-risk stratum period 1 with an observed event rate of 100% and an upper alert boundary of 44.4% (49/136).

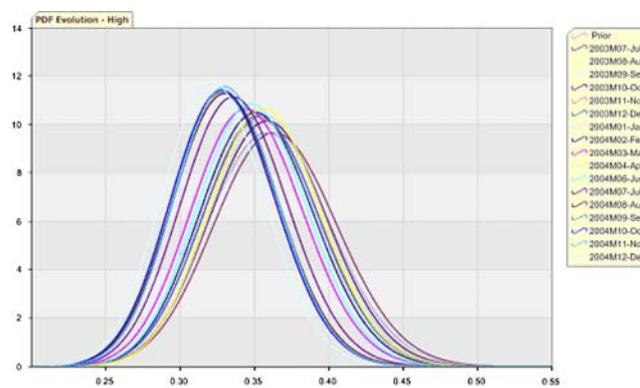
There was a trend toward lower event rates in the PDFs of all risk strata, as illustrated for the high-risk stratum cases in Figure 5. At no time in any strata did the posterior CI overlap fall below the user-specified 80% criteria.

The upper alert boundaries in the alternative data set were the same as the real data set. However, in the moderate-risk stratum, the observed event rates of 1.73% (4/231) in period 3, 1.84% (6/325) in period 4, 2.0% (8/400) in period 5, 1.73% (10/576) in period 7, and 1.71% (14/818) in period 10 exceeded the alert boundaries that ranged from 1.65% to 1.74% for those periods.

The trend toward lower event rates in the PDFs of all risk strata in the real data set was not found in the alternative data set. At no time in any strata did the posterior CI overlap fall below the user-specified 80% criteria.

### Discussion

The DELTA system satisfied all prespecified design requirements and performed all analyses and graphical renderings within two seconds each on the hospital Intranet.



**Figure 5.** Bayesian updating statistics probability density function (PDF) evolution for high-risk cases by period.

The SPC method triggered periodic alerts in both single- and multiple-risk strata analyses. This method also triggered the first period's cumulative alert in the multiple-risk strata, but this can be considered a periodic equivalent alert. Otherwise, there were no cumulative event rate alerts detected by the SPC method. The LR method generated no alarms in either the periodic or cumulative evaluations. The BUS method generated an alert only in the first period of the high-risk stratum. While all BUS alerts are considered cumulative, the alert was generated from one case with a positive outcome for that period.

The alternative data set event rate was elevated manually to generate alarms. The single stratum SPC method alerted to event rates exceeding the threshold for periods 4 through 11. The multistrata SPC method revealed that the event rate increase of concern was in the moderate-risk group, alerting from periods 3 through 11. The LR method generated no periodic or cumulative alerts in the alternative data set. The BUS method agreed that the elevation was primarily of concern in the moderate-risk stratum by generating cumulative alerts in periods 3, 4, 5, 7, and 10.

Periodic alerts are very sensitive measures of elevated event rates but generally lack the statistical power to make a conclusive decision about the safety of a device. These alerts would serve to heighten surveillance and possibly reduce the interval of evaluation for the new device but would not, in and of themselves, be sufficient to recommend withdrawal of the device. The discrepancy between SPC and LR periodic alerting was because LR attempts to adjust the alerting threshold based on the expected outcome of a given case. If there were an increase in the event rate for a period, SPC would trigger an alert as the rate exceeds the static threshold. However, if the LR model expected the cases to have that outcome, then the method would likely not alert because the alert threshold would be adjusted based on that expectation. The cumulative alerts for this analysis were consistent across statistical methods, and the alerts in the first period were due to a very low number of examined cases. In the alternative set, the LR method has no cumulative alerts, and this could be due to the fact that the events were expected by the model.

In phase 3 randomized controlled trials, there are no previous data to use as a benchmark, and a common method of determining the threshold of stopping the trial is to initially place the threshold at a very statistically improbable number (such as five or six standard errors from an estimated allowable rate) and gradually reduce the allowable error as the volume of data grows. The allowable rates are generally established by expert consensus and are manually generated on a trial-by-trial basis.

The benefits of incorporating previous information into the development of alerting thresholds include the ability to develop and establish explicit rules for alerting thresholds. This methodology could then be applied in an objective manner to a wide variety of monitoring applications. This removes the need for an expert consensus to develop the thresholds.

However, this objective methodology has limitations. The accuracy of the alerting boundaries is dependent on the source data. In the case of this clinical example, the University of

Michigan Bare-metal Stent Study mortality data and model were established as the benchmark. DELTA then considered mortality event rates statistically significantly above that baseline to be abnormal and of concern. This becomes important when assessing the external validity of the benchmark data with regards to applicability in a different patient population. In addition, applying multiple concurrent statistical methodologies to a monitoring process is meant to guard against specific vulnerabilities that one methodology might have to these types of confounding.

Statistical process control is only concerned with the overall event rate in the benchmark source population to establish alerting boundaries, and these are static throughout the analysis. This is the least sensitive to subpopulation variations between the study and baseline populations. Including multiple-risk strata in the analysis increases the sensitivity to finding problems in a specific risk group but requires the user to ensure that the study subpopulations using the risk stratification criteria are representative of the source subpopulations. Similar proportions and relative event rate risks between the source data and study data support the use of stratification in this clinical example.

Logistic regression is the most susceptible method to population differences because it provides a case-level estimation based on a number of risk factors. In a number of studies, these models have degraded predictive ability at the case level in disparate populations and as the time from the model's development increases.<sup>39</sup> In the example, the population's event rate was 1.19% and the LR model's expected event rate was 1.75%. This shows that the LR model overpredicted mortality for this population.

Bayesian user statistics carries many of the same benefits and drawbacks of using the aggregate source population's event rate to establish alerting thresholds but allows for the movement of these thresholds by changing study event rates. This method is the most capable in determining a significant shift in a short period of time.

Overall, the results of the example analysis support that the in-hospital mortality following implantation of a drug-eluting stent was acceptably low over the time period studied when compared with the University of Michigan Bare-metal Stent Study benchmark data. The prototype system currently in use at Brigham and Women's Hospital Cardiac Catheterization Laboratory is in a testing and evaluation phase, and as such, clinicians do not consult the system directly. An evaluation of the current user interface will be conducted to assess DELTA's acceptability in the clinical environment by different health care providers. However, the preliminary results of our testing are encouraging: the DELTA system shows promise in filling a need for automated real-time safety monitoring in the medical domain and may be applicable to routine safety monitoring for hospital quality assurance and monitoring of new drugs and devices.

#### References ■

1. Munsey RR. Trends and events in FDA regulation of medical devices over the last fifty years. *Food Drug Law J.* 1995;50:163-77.
2. Medical Device Amendments of 1976 to the Federal Food, Drug, and Cosmetic Act, Pub. L. No. 94-295, 90 Stat. 539 (1976).

3. Pritchard WF, Carey RF. U.S. Food and Drug Administration and regulation of medical devices in radiology. *Radiology*. 1997;205:27-36.
4. Safe Medical Devices Act of 1990, Pub. L. No. 101-629, 104 Stat. 4511 (1990).
5. Merrill RA. Modernizing the FDA: an incremental revolution. *Health Aff (Millwood)*. 1999;18:96-111.
6. Kessler L, Richter K. Technology assessment of medical devices at the Center for Devices and Radiological Health. *Am J Manag Care*. 1998;4, Spec no:SP129-35.
7. O'Neill WW, Chandler JG, Gordon RE, et al. Radiographic detection of strut separations in Bjork-Shiley convexo-concave mitral valves. *N Engl J Med*. 1995;333:414-9.
8. Brown SL, Morrison AE, Parmentier CM, Woo EK, Vishnavujala RL. Infusion pump adverse events: experience from medical device reports. *J Intraven Nurs*. 1997;20:41-9.
9. Fuller J, Parmentier C. Dental Device-associated problems: an analysis of FDA postmarket surveillance data. *J Am Dent Assoc*. 2001;132:1540-8.
10. White GG, Weick-Brady MD, Goldman SA, et al. Improving Patient care by reporting problems with medical devices. *CRNA*. 1998;9:139-56.
11. Dwyer D. Medical device adverse events and the temporary invasive cardiac pacemaker. *Int J Trauma Nurs*. 2001;7:70-3.
12. Dillard SF, Hefflin B, Kaczmarek RG, Petsonk EL, Gross TP. Health effects associated with medical glove use. *AORN J*. 2002;76:88-96.
13. Brown SL, Duggirala HJ, Pennello G. An association of silicone-gel breast implant rupture and fibromyalgia. *Curr Rheumatol Rep*. 2002;4:293-8.
14. Monsein LH. Primer on medical device regulation. Part I. History and background. *Radiology*. 1997;205:1-9.
15. Improving patient care by reporting problems with medical devices. *Med Watch*. Rockville, MD: Department of Health and Human Services, U.S. Food and Drug Administration, HF-2; September 1997.
16. Center for Devices and Radiologic Health. Available from: <http://www.fda.gov/cdrh/index.html/>. Accessed January 18, 2005.
17. Feigal DW, Gardner SN, McClellan M. Ensuring safe and effective medical devices. *N Engl J Med*. 2003;348:191-2.
18. Moss AJ, Hamburger S, Moore RM, Jeng LL, Howie LJ. Use of selected medical device implants in the United States, 1988. *Advance Data from Vital and Health Statistics of the National Center for Health Statistics* 191:1-24 Hyattsville, MD: National Center for Health Statistics, 26 February 1991.
19. Managing Risks from Medical Product Use: Creating a Risk Management Framework. Report to the FDA Commissioner from the Task Force on Risk Management. Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration; May 1999.
20. Monsein LH. Primer on medical device regulation. Part II. Regulation of medical devices by the U.S. Food and Drug Administration. *Radiology*. 1997;205:10-8.
21. Maisel WH. Medical device regulation: an introduction for the practicing physician. *Ann Intern Med*. 2004;140:296-302.
22. Medical device and user facility and manufacturer reporting, certification and registration; delegations of authority; medical device reporting procedures; final rules. *Fed Regist*. 1995;60:63577-606.
23. Postmarket surveillance. Final rule. *Fed Regist*. 2002;67:5943-52.
24. Center for Devices and Radiologic Health Annual Report Fiscal Year 2000. January 2001. Available from: <http://www.fda.gov/cdrh/annual/fy2000/annualreport-2000-5.html/>. Accessed January 18, 2005.
25. Managing Risks from Medical Product Use: Creating a Risk Management Framework. Report to the FDA Commissioner from the Task Force on Risk Management. Rockville, MD: U.S. Department of Health 26. 105 Code of Massachusetts Regulations: Department of Public Health. 2001. 130.1201-130.130.
26. Cannon CP, Battler A, Brindis RG, Cox LJ, Ellis SG, Every NR, et al. ACC key data elements and data definitions for measuring the clinical management and outcomes of patients with acute coronary syndromes: a report of the American College of Cardiology Task Force on Clinical Data Standards. *J Am Coll Cardiol*. 2001;38:2114-30.
27. New York Pub. Health L. No. 2805-1 Incident Reporting (added L. 1986, c.266).
28. Cook DF. Statistical process control for continuous forest products manufacturing operations. *Forest Prod J*. 1992;42:47-53.
29. Grigg N, Walls L. The use of statistical process control in food packaging: preliminary findings and future research agenda. *Br Food J*. 1999;101:763-84.
30. U.S. DOT Federal Railroad Administration. Developed Wheel and Axle Assembly Monitoring System for Improved Passenger Train Safety. RR00-02 March 2000.
31. Doble M. Six Sigma and chemical process safety. *Int J Six Sigma Compet Adv*. 2005;1:229-44.
32. Standards for Privacy of Individually Identifiable Health Information: final rule. *Fed Regist*. 2002;67:53182-273.
33. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med*. 1998;17:857-72.
34. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. San Francisco: Jossey-Bass, 2000.
35. Siu N, Apostolakis G. Modeling the detection rates of fires in nuclear plants. Development and application of a methodology for treating imprecise evidence. *Risk Anal*. 1986;6:43-59.
36. Bayes T. Essay towards solving a problem in the doctrine of chances. *Philos Trans R Soc Lond*. 1763;53:370-418.
37. Resnic FS, Zou KH, Do DV, Apostolakis G, Ohno-Machado L. Exploration of a Bayesian updating methodology to monitor the safety of interventional cardiovascular procedures. *Med Decis Making*. 2004;24:399-407.
38. Moscucci M, Kline-Rogers E, Share D, et al. Simple bedside additive tool for prediction of in-hospital mortality after percutaneous coronary interventions. *Circulation*. 2001;104:263-8.
39. Matheny ME, Ohno-Machado L, Resnic FS. Discrimination and calibration of mortality risk prediction models in interventional cardiology. *J Biomed Inform*. 2005;38:367-75.

#### Further Readings ■

1. Medical device and user facility and manufacturer reporting, certification and registration; delegations of authority; medical device reporting procedures; final rules. *Fed Regist*. 1995;60:63577-606.
2. Postmarket surveillance. Final Rule. *Fed Regist*. 2002;67:5943-42.
3. Center for Devices and Radiologic Health Annual Report Fiscal Year 2000. January 2001. Accessed at <http://www.fda.gov/cdrh/annual/fy2000/annualreport-2000-5.html> on January 18, 2005.
4. Cook DF. Statistical Process Control for continuous forest products manufacturing operations. *Forest Products Journal*. 1992; 42(7/8):47-53.
5. Grigg N, Walls L. The use of statistical process control in food packaging: preliminary findings and future research agenda. *British Food Journal*. 1999;101(9):763-84.
6. Tornado, HGL Dynamics, Inc. Surrey, GU.
7. US DOT Federal Railroad Administration. Developed Wheel and Axle Assembly Monitoring System for Improved Passenger Train Safety. RR00-02 March 2000.
8. WinTA, Tensor PLC, Great Yarmouth, NR.
9. Clinitrace, Phase Forward, Waltham, MA.
10. Oracle Adverse Event Reporting System (AERS), Oracle, Redwood Shores, CA.
11. Trialex, Meta-Xceed, Inc., Fremont, CA.
12. Netregulus, Netregulus, Inc., Centennial, CO.

## Appendix

### Appendix 1 ■ Summary of the Sample Population and Outcome of Interest (Death) per Risk Strata in the University of Michigan Data Sample<sup>39</sup>

Risk Stratum	Risk Score	Sample	Deaths	% Death	Upper 95% CI
Low	0–1.49	1,820	1	0.015	0.03
Moderate	1.5–5.49	3,907	50	0.28	0.017
High	5.50+	136	49	36.0	0.443
Total		5,863	100	1.71	0.0207

CI = confidence interval.

### Appendix 2 ■ University of Michigan Covariates with Beta Co-efficients for the Logistic Regression (LR) Model and Risk Scores for the Discrete Risk Stratification<sup>39</sup>

Covariate	LR $\beta$	Odds Ratio	Risk Score
Myocardial infarction within 24 hours	1.03	2.8	1
Cardiogenic shock	2.44	11.5	2.5
Creatinine >1.5 mg/dL	1.70	5.5	1.5
History of cardiac arrest	1.29	3.65	1.5
No. of diseased vessels	0.44	1.54	0.5
Age 70–79 yr	0.81	2.24	
Age $\geq$ 80 yr	0.97	2.65	
Age $\geq$ 70 yr			1.0
Left ventricular ejection fraction <50%	0.51	1.66	0.5
Thrombus	0.52	1.67	0.5
Peripheral vascular disease	0.46	1.57	0.5
Female gender	0.59	1.82	0.5
Intercept	-7.20		

Intercept is the logistic regression model equation intercept.