

RESEARCH ARTICLE

Open Access

# Evaluation of an automated safety surveillance system using risk adjusted sequential probability ratio testing

Michael E Matheny<sup>1,2,3\*</sup>, Sharon-Lise T Normand<sup>4</sup>, Thomas P Gross<sup>5</sup>, Danica Marinac-Dabic<sup>5</sup>, Nilsa Loyo-Berrios<sup>5</sup>, Venkatesan D Vidi<sup>6</sup>, Sharon Donnelly<sup>6</sup> and Frederic S Resnic<sup>6</sup>

## Abstract

**Background:** Automated adverse outcome surveillance tools and methods have potential utility in quality improvement and medical product surveillance activities. Their use for assessing hospital performance on the basis of patient outcomes has received little attention. We compared risk-adjusted sequential probability ratio testing (RA-SPRT) implemented in an automated tool to Massachusetts public reports of 30-day mortality after isolated coronary artery bypass graft surgery.

**Methods:** A total of 23,020 isolated adult coronary artery bypass surgery admissions performed in Massachusetts hospitals between January 1, 2002 and September 30, 2007 were retrospectively re-evaluated. The RA-SPRT method was implemented within an automated surveillance tool to identify hospital outliers in yearly increments. We used an overall type I error rate of 0.05, an overall type II error rate of 0.10, and a threshold that signaled if the odds of dying 30-days after surgery was at least twice than expected. Annual hospital outlier status, based on the state-reported classification, was considered the gold standard. An event was defined as at least one occurrence of a higher-than-expected hospital mortality rate during a given year.

**Results:** We examined a total of 83 hospital-year observations. The RA-SPRT method alerted 6 events among three hospitals for 30-day mortality compared with 5 events among two hospitals using the state public reports, yielding a sensitivity of 100% (5/5) and specificity of 98.8% (79/80).

**Conclusions:** The automated RA-SPRT method performed well, detecting all of the true institutional outliers with a small false positive alerting rate. Such a system could provide confidential automated notification to local institutions in advance of public reporting providing opportunities for earlier quality improvement interventions.

## Background

Public reporting of risk adjusted mortality rates following cardiac surgery has become an important tool in the evaluation and improvement of quality of patient care [1]. Several states, including Massachusetts, have enacted legislation requiring public reporting of cardiac surgical outcome data [2]. Beginning in 2002, all Massachusetts hospitals performing cardiac surgery were required to submit cardiac surgical outcomes data to the Massachusetts Data Analysis Center (Mass-DAC), a

data coordinating center for the Massachusetts Department of Public Health (MA DPH), and to the Society of Thoracic Surgeons (STS) using the STS National Cardiac Database collection tool. Since then, Mass-DAC has published annual public reports on 30-day, all-cause, risk-adjusted mortality rates for isolated coronary artery bypass surgery (CABG) for institutions, and beginning in 2004, for individual cardiac surgeons [3].

Though public reports are intended to provide transparency and public accountability, and to inform consumer choice, there are other consequences to public reporting that have potentially significant long-term, financial and reputational impact on both the institutions and providers. In Massachusetts, two centers were

\* Correspondence: michael.matheny@vanderbilt.edu

<sup>1</sup>GRECC and Center for Health Services Research, Tennessee Valley Healthcare System, Veterans Administration, Nashville, TN, USA  
Full list of author information is available at the end of the article

identified as statistically significant high mortality outliers, one of which was identified as such in multiple, consecutive years [4,5]. As a result, this cardiac surgery program was temporarily suspended while quality improvement initiatives were undertaken [6]. Notification of the first year of outlier status for this program was not available publicly until 2 to 2.75 years after data collection (relative to calendar quarter) [5,6].

There are inherent delays between performance of procedures and public reporting of outcomes due to the rigorous data review and manual case adjudication required from both a regulatory and data quality standpoint. Massachusetts has steadily decreased the delay between data collection and public reporting since the program's inception in 2002, currently reduced for fiscal year 2008 to 1.25 to 2 years (relative to calendar quarter) [7]. However, any temporal delays between the performance of the procedure and the analysis and public reporting of results is undesirable and could expose some patients to increased risk of morbidity and mortality as procedures continue to be performed while a quality issue exists. An active, automated prospective surveillance system, as an adjunct to the existing rigorous regulatory approach, could provide institutions or physicians timely internal feedback and provide opportunities to mitigate these risks well in advance of public release of the data.

We designed the Data Extraction and Longitudinal Trend Analysis (DELTA) system to provide real-time monitoring of clinical data to support continuous quality or safety monitoring of newly approved medical devices, medications, or therapeutic interventions [8]. This system supports a variety of frequentist and Bayesian statistical methods, which can be configured to provide unadjusted and risk-adjusted safety monitoring among prospective and retrospective cohorts [8-11]. DELTA also incorporates de-identification and encryption algorithms to guard protected health information and control data ownership, and employs flexible alerting mechanisms to trigger notifications via e-mail or through the web interface when an observed event rate exceeds boundaries of risk-adjusted expectations for the event of interest.

Risk-adjusted SPRT, a method for observational cohort safety surveillance, was first proposed by Spiegelhalter and colleagues [12]. This method has been used to evaluate hospital and physician performance among retrospective cohorts for coronary artery bypass patients and percutaneous coronary intervention patients [12-14]. While this method has not achieved widespread use, it is well suited for implementation in a prospective, automated system because it analyzes each sequential case and incorporates adjustment for repeated measures on the same subject with explicit type I and II error rates.

In this study, we sought to assess the utility of Risk-Adjusted Sequential Probability Ratio testing when imbedded in an automated surveillance tool as compared to the gold standard of retrospective annual quality reports used by the Massachusetts Department of Public Health (MA-DPH). The primary outcomes of this analysis were the sensitivity and specificity of the automated implementation as compared with the public reporting methods of Massachusetts Data Analysis Center (MASS-DAC) and the MA-DPH, assessed for hospital 30-day mortality after isolated coronary artery bypass graft surgery.

## Methods

### Study Setting

Massachusetts regulations require all acute care non-federal hospitals that provide cardiac surgery to collect data using a standardized data collection instrument based on the Society of Thoracic Surgery (STS) registry [15]. Each institution is required to submit data on a quarterly basis to Mass-DAC, and participating centers collect the data using a variety of point-of-care collection tools, chart review, and patient follow-up. Mass-DAC performs manual adjudication of all cases with adverse outcomes as well as a sample of all other case records. Yearly reports of hospital and surgeon 30-day mortality performance are published. Additional information and annual public reports are available online [3].

A total of 23,020 isolated adult coronary artery bypass surgery admissions were conducted from January 1, 2002 to September 30, 2007. The surgeries did not involve valve replacement or other associated cardiac surgical procedures. We selected these surgeries for our study because the state uses them as the primary index of institution and surgeon quality for cardiovascular surgery. In 2006, Mass-DAC changed reporting from a calendar year basis to a fiscal year basis that runs from October 1 through September 30. Consequently, the 2006 fiscal year analysis included the last three months of the 2005 calendar year. The primary patient outcome of the registry is the 30-day all-cause mortality after isolated coronary artery bypass graft surgery. We focused on 30-day all-cause hospital-specific risk-standardized mortality rates. The current study was approved by the Brigham & Women's and Harvard Medical School's Institutional Review Boards.

### Gold Standard Statistical Analysis

Mass-DAC reports the data annually utilizing Bayesian hierarchical logistic regression [1]. The model assumes that the log-odds of mortality is linearly related to a set of patient risk factors and permits baseline risk to vary across hospitals through the inclusion of a hospital-

specific intercept. Estimation of the model parameters, including the between-hospital variance, hospital overall mean log-odds, and regression coefficients of patient-level risk factors are obtained via Markov chain Monte Carlo (MCMC) methods. The MCMC method uses the Gibbs sampler to sequentially sample from probability distributions and produces a Markov chain with the joint posterior density as its stationary distribution [16]. This is accomplished by selecting a set of starting values, then performing a number of “burn-in” sampling iterations that are not recorded, followed by the collection and averaging of additional sampling iterations to form the final posterior estimates. The primary analysis used all of the data for the year and declared a hospital as an outlier if the lower limit of the 95% posterior interval of the risk-adjusted institutional standardized mortality rate exceeded the unadjusted statewide mortality rate. Because of the small number of cardiac surgery hospitals in Massachusetts, Mass-DAC also performs cross-validation analyses in which each hospital is eliminated and data from the remaining hospitals are used to assess hospital performance in the eliminated hospital. This strategy was developed to avoid one large center from having too great an influence on statewide risk expectations. A hospital was considered an outlier if either the 95% posterior interval from the statewide comparison exceeded the unadjusted statewide mortality rate or the posterior predictive p-value from the cross-validation analysis was 0.01 or smaller.

Hospital 9 was declared an outlier in the 2002-2005 reports, and hospital 8 was declared an outlier in the 2004 report. All of hospital outliers were detected with only the cross-validation evaluation by MASS-DAC. Although the original 2002 public report did not include the cross-validation analysis method, we have exactly reproduced this analysis protocol on this data to provide consistency across all years. A summary of the risk-adjusted standardized mortality incidence rate (with upper and lower limits of the 95% interval) by hospital and year are reported in Figure 1. The dotted line in the figure represents the statewide unadjusted mortality incidence rate, and values in red indicated an outlying hospital.

#### Automated Risk Adjusted Sequential Probability Ratio Testing (RA-SPRT)

The SPRT control chart methodology detects unacceptable event rates by evaluating each unit of analysis sequentially in time [12,13]. The risk hypothesis is whether the observed outcome event rate exceeds the accepted or baseline event rate given a specific odds ratio (OR) and Type I and II error [12]. This method accepts or rejects this hypothesis after each sequential case is evaluated. Risk adjustment is performed through the use of a risk prediction model whereby the

cumulative log likelihood ratio is adjusted by the probability of the outcome [17]. Repeated measurement (re-analysis after each additional case) error adjustments are incorporated explicitly in the framework. These features are uncommon in statistical process control methods and are strengths of this method.

The following describes the calculations necessary to construct risk-adjusted SPRT control charts as refined by Rogers and colleagues[13] based on Spiegelhalter’s work [12]. The control limits are defined by

$$h_0 = \frac{\ln\left(\frac{1-\alpha}{\beta}\right)}{\ln(OR)} \tag{1}$$

and

$$h_1 = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)}{\ln(OR)} \tag{2}$$

where  $h_0$  and  $h_1$  are the cumulative log-likelihood ratio values in which the null hypothesis ( $H_0$ ) or the alternate hypothesis ( $H_1$ ) is accepted (respectively), OR is the odds ratio,  $\alpha$  is the type I error rate, and  $\beta$  is the type II error rate.

The cumulative log-likelihood ratio value ( $T^{cum}$ ) is calculated in sequence for higher risk detection ( $OR > 1$ ) with

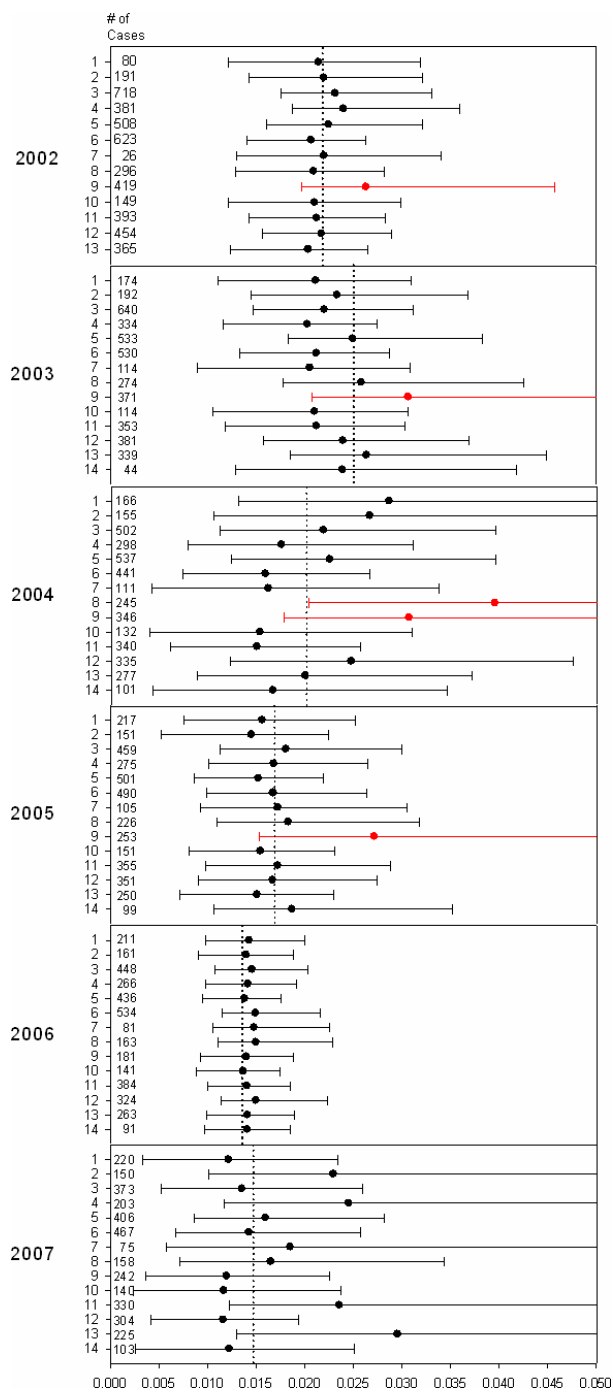
$$T_i^{Cum} = T_{i-1}^{Cum} + (O_i - s_i) \tag{3}$$

where  $T_0^{cum} = 0$ ,  $O_i$  is the observed outcome (0 or 1) for a binary procedure for  $i^{th}$  case, and

$$s_i = \frac{\ln((1 - p_i) + (OR * p_i))}{\ln(OR)} \tag{4}$$

where  $p_i$  is the calculated probability of the outcome for the  $i^{th}$  case as determined by the risk prediction model.

We have previously described an automated real-time safety monitoring tool, Data Extraction and Longitudinal Trend Analysis (DELTA), that is able to perform larger numbers of concurrent prospective analyses using a variety of statistical methodologies and alerting thresholds [18]. The system uses a SQL 2005 server (Microsoft Corp., Redmond, WA) to provide internal data storage and configuration information, as well as providing the capability to integrate with external databases. The user interface was developed in the Microsoft .NET programming environment and was displayed in a web browser from a Microsoft IIS 6.0 Web Server (Microsoft Corp., Redmond, WA). Security of patient data is further addressed by record de-identification steps and user login access restrictions [19].



**Figure 1 Gold Standard Results for MASSDAC CABG by year.** Each number on the y axis within each year represents a unique hospital. Circles represent posterior mean risk-standardized mortality rates for each institution; lines are corresponding 95% posterior intervals. The hospitals in red alerted from the cross-validation method only.

The RA-SPRT method was implemented directly within DELTA, and the statistical module evaluated the data after setting cohort inclusion and exclusion criteria as well as the necessary statistical parameters, such as the desired odds ratio and the type I and II error.

Parameters and risk variable selection for the logistic regression risk adjustment were then passed through a bi-directional interface to SAS (Version 9.1, Cary, NC) in order to develop the required logistic regression models.

### Statistical Analysis

The risk-adjusted sequential probability ratio testing (RA-SPRT) method was used to evaluate the data separately for each calendar or fiscal year. Although one of the strengths of this method is that it can accumulate data continuously until the alerting odds ratio hypothesis is accepted or rejected, analyses were terminated at the end of each calendar or fiscal year and the cumulative log-likelihood ratio was reset to 0. This was done in order to be directly comparable to the gold standard. Risk adjustment was performed using standard logistic regression with the same risk factors used in the source method by Mass-DAC. Each logistic regression model was developed from data in the prior 11 months and then applied sequentially to each case in the “current” month. This process was repeated throughout the entire range of the data. Data were not available prior to 2002, so the models developed prior to December 2002 used from one to ten months of data (depending on the analysis month). Data from January 2002 were not analyzed by RA-SPRT because they were required to build the first model. It should be noted that the risk models developed from the first two months of data showed regression coefficient instability and poor calibration, as would be expected with low sample sizes. Both of these measurements subsequently stabilized for the remainder of the data. A type I error of 0.05 and type II error level of 0.10 were used in each of the RA-SPRT analyses, and an OR threshold of 2.0 was defined as the reasonable thresholds for concern regarding the clinical quality of the institution evaluated. An outlier was declared if a hospital exceeded the log likelihood ratio threshold at any point during that calendar year.

### Results

A total of 23,020 isolated coronary artery bypass graft surgeries were evaluated from January 1<sup>st</sup>, 2002 to September 30<sup>th</sup>, 2007, and 925 cases in the last quarter of 2005 were included in both calendar year 2005 and fiscal year 2006 evaluations to be consistent with MASS-DAC reporting. The annual 30-day mortality rate ranged from 1.41 percent to 2.19 percent with a general decreasing trend. Most of the risk factors included in the risk models were stable across years, with the exceptions of declining prior CABG surgeries, increasing prior percutaneous coronary transluminal angiography procedures, and declining cardiogenic shock rates. A summary of patient risk factor prevalence and outcome event rates by year are shown in Table 1. The risk-adjusted SPRT method detected three hospitals with outlying 30-day mortality outcomes. Hospital 4 experienced a false positive alert in 2002 (Figure 2A), Hospital 8 experienced a true positive alert in 2004 (Figure 2B), and Hospital 9 experienced true positive alerts in 2002,

2003, 2004, and 2005 (Figure 2C). The remaining evaluations for each hospital were true negatives (for example, Hospital 6 shown in Figure 2D). This resulted in a sensitivity of RA-SPRT of 100% (5/5) and the specificity was 98.8% (79/80) compared to the publicly available reports.

### Discussion

The automated RA-SPRT method within DELTA performed well applied to a statewide clinical registry data over a number of years when compared with the method used by the state to produce the public reporting of institutional and physician performance reports. The method detected each of the true outliers and generated a single false positive, which is a desirable profile for an early detection system used for knowledge discovery and internal quality improvement initiatives. Due to the frequent periodic analysis that the method employs, the DELTA system would have reported these findings to the local institutions significantly earlier than they were available publicly.

In general, an automated outcomes surveillance system, whether used for post-marketing surveillance or institutional and physician profiling, should be tuned with appropriate error levels and alerting thresholds to be over-sensitive in a manner similar to a screening test, where it is highly desirable to capture all of the true signals and tolerance to false positive signals is dependent on the resources that are available to perform root cause analyses and further data exploration. The RA-SPRT method did perform in this manner in this data set using stock values for types I and II error rates, and a common odds ratio threshold, which is encouraging, but does require further validations in other clinical domains to increase the generalizability of these findings. Selection of a desirable false positive versus false negative threshold depends on the clinical domain's desire to avoid missing a true signal and the cost of performing more in depth analysis of each detected signal. In addition, processing time was very reasonable within the automated application, and a quarter of data was able to be analyzed in seconds, with the most time consuming step being that of logistic regression model generation in each month analyzed.

The RA-SPRT method has a number of strengths, including explicit alerting boundary thresholds and incorporation of repeated testing [12-14]. It should be noted that the RA-SPRT method uses a simple hypothesis rather than a composite hypothesis that tests all odds ratios greater than 1.0 for statistical significance. This could potentially result in a statistically significant odds ratio less than the selected threshold (such as 1.1 or 1.2) that is not detected as an outlier by the method. However, in many cases, such an alert may not be

**Table 1 Summary information for patient and institutions by year.**

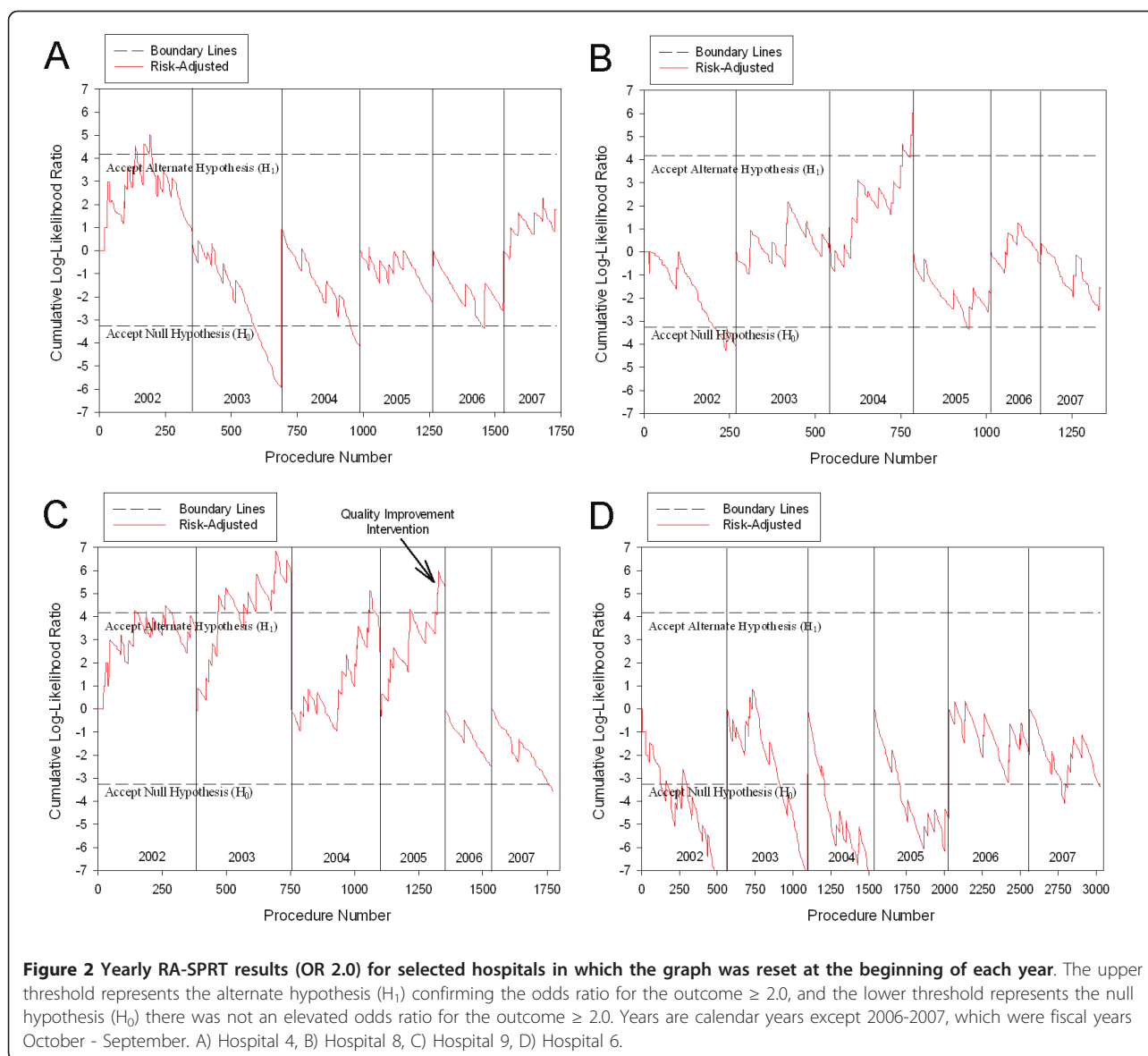
|                                  | 2002                | 2003                | 2004                | 2005                | 2006                | 2007                |
|----------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Number of Hospitals (count)      | 13                  | 14                  | 14                  | 14                  | 14                  | 14                  |
| Number of Admissions (count)     | 4604                | 4393                | 3986                | 3885                | 3684                | 3396                |
| 30-Day Crude Mortality           | 2.17<br>(1.77-2.63) | 2.25<br>(1.84-2.74) | 2.01<br>(1.60-2.49) | 1.65<br>(1.27-2.10) | 1.41<br>(1.06-1.85) | 1.47<br>(1.09-1.94) |
| Mean Age in Years (SD)           | 66.5<br>(10.7)      | 66.7<br>(10.6)      | 66.9<br>(10.7)      | 66.5<br>(10.8)      | 66.5<br>(10.7)      | 66.0<br>(10.9)      |
| Male                             | 74.5<br>(73.2-75.8) | 73.5<br>(72.2-74.8) | 74.5<br>(73.1-75.9) | 76.4<br>(75.1-77.8) | 75.2<br>(73.8-76.7) | 75.9<br>(74.4-77.3) |
| Renal Failure                    | 7.3<br>(6.6-8.1)    | 6.9<br>(6.2-7.7)    | 5.8<br>(5.1-6.6)    | 6.4<br>(5.6-7.2)    | 6.6<br>(5.8-7.4)    | 7.5<br>(6.6-8.4)    |
| Diabetes Mellitus                | 38.0<br>(36.6-39.4) | 38.1<br>(36.6-39.5) | 37.0<br>(35.5-38.5) | 39.3<br>(37.8-40.8) | 39.4<br>(37.9-41.0) | 42.6<br>(40.9-44.3) |
| Hypertension                     | 77.0<br>(75.7-78.2) | 79.5<br>(78.3-80.7) | 82.7<br>(81.1-83.5) | 83.9<br>(82.7-85.1) | 84.1<br>(82.9-85.3) | 83.5<br>(82.2-84.8) |
| Peripheral Vascular Disease      | 18.0<br>(16.9-19.2) | 17.4<br>(16.3-18.5) | 17.7<br>(16.5-18.9) | 17.4<br>(16.3-18.7) | 17.4<br>(16.1-18.6) | 17.4<br>(16.1-18.7) |
| Prior CABG Surgery               | 3.8<br>(3.3-4.4)    | 3.1<br>(2.6-3.7)    | 2.6<br>(2.1-3.2)    | 2.5<br>(2.0-3.0)    | 1.8<br>(1.4-2.3)    | 2.1<br>(1.6-2.6)    |
| Prior PTCA/PCI                   | 18.6<br>(17.5-19.8) | 17.8<br>(16.7-18.9) | 19.1<br>(17.8-20.3) | 20.2<br>(19.0-21.6) | 21.0<br>(19.7-22.3) | 21.6<br>(20.2-23.0) |
| Cardiogenic Shock                | 2.2<br>(1.8-2.7)    | 1.6<br>(1.3-2.0)    | 1.1<br>(0.8-1.5)    | 1.0<br>(0.7-1.4)    | 0.9<br>(0.6-1.2)    | 0.8<br>(0.5-1.1)    |
| Ejection Fraction                |                     |                     |                     |                     |                     |                     |
| <30% or missing                  | 12.8<br>(11.9-13.9) | 12.6<br>(11.4-13.4) | 11.8<br>(10.8-12.8) | 11.6<br>(10.6-12.7) | 10.1<br>(9.1-11.1)  | 10.4<br>(9.5-11.6)  |
| 30 - 39                          | 11.7<br>(10.8-12.7) | 12.4<br>(11.4-13.4) | 11.0<br>(10.1-12.0) | 10.9<br>(9.9-11.9)  | 10.0<br>(9.0-11.0)  | 10.5<br>(9.5-11.6)  |
| Myocardial Infarction            |                     |                     |                     |                     |                     |                     |
| Within 24 Hours                  | 2.7<br>(2.3-3.2)    | 3.2<br>(2.6-3.7)    | 3.8<br>(3.3-4.5)    | 2.9<br>(2.4-3.5)    | 3.3<br>(2.7-3.9)    | 3.3<br>(2.7-4.0)    |
| 1 - 7 days                       | 20.7<br>(19.5-21.9) | 23.0<br>(21.7-24.2) | 22.1<br>(20.8-23.4) | 23.1<br>(21.8-24.4) | 23.4<br>(22.0-24.8) | 24.5<br>(23.1-26.0) |
| Status of CABG Surgery           |                     |                     |                     |                     |                     |                     |
| Urgent                           | 62.0<br>(60.1-63.4) | 65.8<br>(64.3-37.2) | 66.6<br>(65.1-68.0) | 61.3<br>(59.8-62.8) | 59.6<br>(58.0-61.2) | 62.3<br>(60.7-64.0) |
| Emergent/Salvage                 | 3.7<br>(3.2-4.3)    | 2.9<br>(2.4-3.4)    | 3.1<br>(2.6-3.7)    | 2.5<br>(2.0-3.0)    | 2.6<br>(2.1-3.1)    | 3.3<br>(2.7-3.9)    |
| Pre-op Intra-Aortic Balloon Pump | 9.3<br>(8.5-10.2)   | 11.7<br>(10.7-12.6) | 11.6<br>(10.6-12.6) | 10.9<br>(9.9-11.9)  | 11.2<br>(10.2-12.3) | 11.3<br>(10.2-12.4) |

Years 2002-2005 were collected in the calendar year, and 2006-2007 were collected in the fiscal year from October of the previous year to September of the reported year. All data are percents (with 95% confidence intervals) unless otherwise specified as means (with standard deviations).

clinically significant. In this study, we chose an odds ratio threshold that would be clearly clinically significant.

There are a number of limitations to this work. It should be noted that the sensitivity and specificity are optimistic in that implementation of this system in a prospective setting would eliminate the time costly manual data adjudication and outcome chart review steps that are done by Mass-DAC, at the expense of the

accuracy of the data analyzed. Another limitation in the use of this system is the timing of data submission by the sites. Currently, data are submitted quarterly, rather than monthly (as used in this analysis), by MA cardiac institutions. The RA-SPRT method is intended for sequential case level analysis and performing the analysis in larger time segments results in over-correction of the repeated measurement adjustment. Real-time data surveillance would be ideal, but requires a regular and



relatively clean data stream submission from the source institution. In order to evaluate this potential and evaluate the impact of lack of data adjudication, we are currently conducting percutaneous coronary angiography prospective surveillance in a subset of institutions in the state with the necessary infrastructure for near real-time submission.

### Conclusions

The RA-SPRT method implemented within an automated surveillance system was able to detect institutional outliers in a statewide clinical registry. While either a significant electronic health record infrastructure or a state reporting mechanism is required to realize the full utility of this system for outcome profiling,

it could result in significant time savings in providing early warnings to local institutions and physicians.

### Acknowledgements

This study was funded in part by grants R01-LM-008142 from the National Library of Medicine, FDA-SOL-08-00837A from the Food and Drug Administration, and VA HSR&D CDP-09-387 from the Veteran's Administration Health Services Research and Development Service. We would like to acknowledge Richard Cope and Coping Systems for their technical expertise and programming work.

### Author details

<sup>1</sup>GRECC and Center for Health Services Research, Tennessee Valley Healthcare System, Veterans Administration, Nashville, TN, USA. <sup>2</sup>Division of General Internal Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>3</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>4</sup>Department of Health Care Policy, Harvard Medical School and Department of Biostatistics, Harvard

School of Public Health, Boston, MA, USA. <sup>5</sup>Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, MD, USA. <sup>6</sup>Division of Cardiology, Brigham & Women's Hospital, Boston, MA, USA.

#### Authors' contributions

MEM participated in the conception and design of the study, analysis and interpretation of the data, drafting of the manuscript, and critical revision of the manuscript. STN participated in the conception and design of the study, acquisition of the data, analysis and interpretation of the data, and critical revision of the manuscript. TPG, DMD, NLB, and SD participated in the conception and design of the study and critical revision of the manuscript. VDV participated in analysis and interpretation of the data and critical revision of the manuscript. FSR participated in the conception and design of the study, analysis and interpretation of the data, and critical revision of the manuscript. All authors read and approved the final manuscript.

#### Competing interests

A patent application by Frederic Resnic, Michael Matheny, and Richard Cope for several key analytic components of DELTA is currently under review at the U.S. Patent Office. As current and former employees of Brigham & Women's Hospital, the intellectual property rights of Frederic Resnic and Michael Matheny related to the development of DELTA are assigned to the institution. Neither Frederic Resnic nor Michael Matheny have previously received nor anticipate any financial compensation, stock options, income, or ownership in DELTA, Coping Systems, or any other related commercial entity. Michael Matheny has no consulting income to report, and Frederic Resnic reports having modest consulting income from Boston Scientific Inc., Abbott Vascular Inc., Cordis Corp., and St. Jude Medical Inc., as well as research grants from The Medicines Company. Dr. Normand receives minor consulting income from the Medicines Company and Analytica International.

Received: 10 December 2010 Accepted: 14 December 2011  
Published: 14 December 2011

#### References

1. Shahian DM, Normand SL, Torchiana DF, Lewis SM, Pastore JO, Kuntz RE, Dreyer PI: **Cardiac surgery report cards: comprehensive review and statistical critique.** *Ann Thorac Surg* 2001, **72**(6):2155-2168.
2. Shahian DM, Torchiana DF, Normand SL: **Implementation of a cardiac surgery report card: lessons from the Massachusetts experience.** *Ann Thorac Surg* 2005, **80**(3):1146-1150.
3. **Massachusetts Data Analysis Center.** [http://www.massdac.org/].
4. Shahian DM, Normand SL: **Comparison of "risk-adjusted" hospital outcomes.** *Circulation* 2008, **117**(15):1955-1963.
5. **Publicly Released Reports for the Commonwealth of Massachusetts Cardiac Surgery Cohort.** [http://www.massdac.org/reports/surgery.html].
6. Ettinger WH, Hylka SM, Phillips RA, Harrison LH Jr, Cyr JA, Sussman AJ: **When things go wrong: the impact of being a statistical outlier in publicly reported coronary artery bypass graft surgery mortality data.** *Am J Med Qual* 2008, **23**(2):90-95.
7. **Time-line summary for the reporting of cardiac surgery project of the Massachusetts Data analysis Center (MASS-DAC).** [http://massdac.org/sites/default/files/CSTimelines.pdf].
8. Matheny ME, Ohno-Machado L, Resnic FS: **Monitoring device safety in interventional cardiology.** *J Am Med Inform Assoc* 2006, **13**(2):180-187.
9. 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**(4):399-407.
10. Matheny ME, Arora N, Ohno-Machado L, Resnic FS: **Rare Adverse Event Monitoring of Medical Devices with the Use of an Automated Surveillance Tool.** *AMIA Annu Symp Proc* 2007, 518-522.
11. Matheny ME, Morrow DA, Ohno-Machado L, Cannon CP, Sabatine MS, Resnic FS: **Validation of an automated safety surveillance system with prospective, randomized trial data.** *Med Decis Making* 2009, **29**(2):247-256.
12. Spiegelhalter D, Grigg O, Kinsman R, Treasure TOM: **Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery.** *Int J Qual Health Care* 2003, **15**(1):7-13.
13. Rogers CA, Reeves BC, Caputo M, Ganesh JS, Bonser RS, Angelini GD: **Control chart methods for monitoring cardiac surgical performance and their interpretation.** *J Thorac Cardiovasc Surg* 2004, **128**(6):811-819.
14. Matheny ME, Ohno-Machado L, Resnic FS: **Risk-adjusted sequential probability ratio test control chart methods for monitoring operator and institutional mortality rates in interventional cardiology.** *Am Heart J* 2008, **155**(1):114-120.
15. Shahian DM, Blackstone EH, Edwards FH, Grover FL, Grunkemeier GL, Nafel DC, Nashef SAM, Nugent WC, Peterson ED: **Cardiac Surgery Risk Models: A Position Article.** *Ann Thorac Surg* 2004, **78**(5):1868-1877.
16. Kass RE, Carlin BP, Gelman A, Neal RM: **Markov Chain Monte Carlo in Practice: A Roundtable Discussion.** *The American Statistician* 1998, **52**(2):93-100.
17. Steiner SH, Cook RJ, Farewell VT, Treasure T: **Monitoring surgical performance using risk-adjusted cumulative sum charts.** *Biostatistics* 2000, **1**(4):441-452.
18. Matheny ME, Ohno-Machado L, Resnic FS: **Monitoring device safety in interventional cardiology.** *J Am Med Inform Assoc* 2006, **13**(2):180-187, Epub 2005 Dec 2015.
19. **Standards for Privacy of Individually Identifiable Health Information: Final Rule.** *Federal Register* 2002, **67**(157):53182-53273.

#### Pre-publication history

The pre-publication history for this paper can be accessed here:  
http://www.biomedcentral.com/1472-6947/11/75/prepub

doi:10.1186/1472-6947-11-75

**Cite this article as:** Matheny et al.: Evaluation of an automated safety surveillance system using risk adjusted sequential probability ratio testing. *BMC Medical Informatics and Decision Making* 2011 **11**:75.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

