Monitoring radiation use in cardiac fluoroscopy imaging procedures

Nathaniel T. Stevens and Stefan H. Steiner
Department of Statistics and Actuarial Sciences, Business and Industrial Statistics Research Group,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Ian R. Smith[a)]
St. Andrew's Medical Institute, St. Andrew's War Memorial Hospital, Brisbane, Queensland 4000, Australia

R. Jock MacKay
Department of Statistics and Actuarial Sciences, Business and Industrial Statistics Research Group,
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Purpose: Timely identification of systematic changes in radiation delivery of an imaging system can lead to a reduction in risk for the patients involved. However, existing quality assurance programs involving the routine testing of equipment performance using phantoms are limited in their ability to effectively carry out this task. To address this issue, the authors propose the implementation of an ongoing monitoring process that utilizes procedural data to identify unexpected large or small radiation exposures for individual patients, as well as to detect persistent changes in the radiation output of imaging platforms.

Methods: Data used in this study were obtained from records routinely collected during procedures performed in the cardiac catheterization imaging facility at St. Andrew's War Memorial Hospital, Brisbane, Australia, over the period January 2008–March 2010. A two stage monitoring process employing individual and exponentially weighted moving average (EWMA) control charts was developed and used to identify unexpectedly high or low radiation exposure levels for individual patients, as well as detect persistent changes in the radiation output delivered by the imaging systems. To increase sensitivity of the charts, we account for variation in dose area product (DAP) values due to other measured factors (patient weight, fluoroscopy time, and digital acquisition frame count) using multiple linear regression. Control charts are then constructed using the residual values from this linear regression. The proposed monitoring process was evaluated using simulation to model the performance of the process under known conditions.

Results: Retrospective application of this technique to actual clinical data identified a number of cases in which the DAP result could be considered unexpected. Most of these, upon review, were attributed to data entry errors. The charts monitoring the overall system radiation output trends demonstrated changes in equipment performance associated with relocation of the equipment to a new department. When tested under simulated conditions, the EWMA chart was capable of detecting a sustained 15% increase in average radiation output within 60 cases (<1 month of operation), while a 33% increase would be signaled within 20 cases.

Conclusions: This technique offers a valuable enhancement to existing quality assurance programs in radiology that rely upon the testing of equipment radiation output at discrete time frames to ensure performance security. © *2011 American Association of Physicists in Medicine.*
[DOI: 10.1118/1.3524224]

Key words: dose area product, linear regression, residual, individuals control chart, exponentially weighted moving average (EWMA)

I. INTRODUCTION

I.A. Motivation

Recent reports concerning the accidental overexposure of stroke patients undergoing brain perfusion CT studies have highlighted the need for increased vigilance in the conduct of practices involving the use of ionizing radiation.[1] This issue, however, is not limited to CT scans alone but is of general concern, especially for modalities and procedures involving significant individual patient radiation risks, such as fluoroscopy procedures and nuclear medicine imaging exams.[2]

Radiation exposures arising from diagnostic and interventional cardiac catheterization procedures are among the highest routinely delivered in any ionizing radiation-based imaging procedure with the effective dose (E) for diagnostic coronary angiography procedures being reported to be of the order of 5–10 mSv, while for complex coronary angioplasty procedures, E can exceed 35 mSv.[3–5] Due to the magnitude and frequency of these types of studies, efforts aimed at maintaining exposures at the minimum required to achieve an effective outcome have the potential to benefit both individual patients and the population as a whole.

To ensure optimal performance of the imaging equipment, minimum performance standards are mandated.[6–8] These standards cover factors related to image quality and radiation dose delivery and stipulate a testing frequency for the verification of defined performance measures. For equipment such as that used in cardiac catheterization, the testing interval is commonly set at 12 months.

A flaw in this testing process concerns the potential for equipment performance to unexpectedly drift or shift discreetly at some point within the test interval either through the development of a fault in the imaging system or as an unintended change in the mode of operation of the equipment (default selection of a high or low dose mode of operation). Under certain circumstances, this change could result in patients and clinical personnel receiving radiation exposures higher than expected. Depending on the cause and magnitude of the variation in exposures being delivered, no marked effect on imaging performance may be detected, and therefore no corrective action to resolve the problem might be taken until the next scheduled preventive maintenance or mandatory test. In the example of the incident involving patients undergoing brain perfusion imaging, the problem does not appear to have been due to a malfunction of the imaging equipment but instead was linked to a process failure in the use of the CT scanner. In total, the scanner was used in more than 200 cases over some 18 months before the error was detected.[1]

Quality issues concerning dose delivery exist in the radiation therapy domain where a subtle drift in the performance of the equipment may result in the under- or overdelivery of radiation to a tumor. This change in radiation delivery can have significant consequences for the patient being treated. To manage this problem, a system of control charts using individual case data has been evaluated and deemed effective in detecting unacceptable drift in the delivery of treatment.[9,10] In general, control charts are more effective in detecting process changes than is a simple review of the data because they provide clear rules concerning when to look for process changes and also when there is no reason to undertake this search. In this paper, we extend the works of Pawlicki *et al.*[9] and Gerard *et al.*[10] in a number of ways.

I.B. Goals

The goal of this project was to develop and describe the use of a radiation exposure monitoring scheme in the cardiac catheterization facilities at St. Andrew's War Memorial Hospital (SAWMH) in Brisbane, Australia.

The proposed ongoing performance monitoring method is capable of

- Identifying unusually high or low exposure levels for individual patients and
- Detecting persistent changes in the average radiation dose applied by the imaging system.

We achieve these two goals by monitoring the process using two control charts: A chart for individuals and an exponentially weighted moving average (EWMA) chart.[12] EWMA control charts are known to be more sensitive than

Shewhart type charts to small to medium persistent process changes or slow drifts. We found the simultaneous use of the two control charts particularly useful in this context because often large individual outliers flagged by the individuals chart were found, upon further investigation, to be due to human error, such as a transcription error. By filtering these erroneous values, we are better able to detect real changes in the imaging equipment.

An obvious and overly simple way to monitor the radiation exposure arising from the imaging process is to chart the raw dose area product (DAP) values associated with each imaging procedure. However, using such a plot can result in a loss of sensitivity to changes in the functionality of the imaging systems because there is a substantial variation in the DAP values due to known factors. These include equipment related factors (e.g., the type of imaging equipment and the dose rate it is configured to deliver in the various modes of operation), as well as factors related to how the equipment is used (e.g., how much imaging is used, number of high dose frames acquired to document the case, number and selection of views employed, beam collimation, and field size), and finally various patient characteristics (e.g., patient size, disease state).

To increase the sensitivity of the charts, we adopt two strategies. First, we stratify the charting by imaging system and procedure type. In our example, we built separate charts for the two imaging systems and used only the data from the most commonly occurring procedure, an exploration of the coronary arteries. Second, we account for variation in DAP values due to other measured factors using multiple linear regression.[11] We then construct control charts using the residual values (actual DAP − predicted DAP), which have far less variability and hence are more sensitive to process changes than the raw DAP measurements.

This paper describes the implementation, the interpretation, and the advantages of this technique. Further discussion about the differences and advantages of our approach compared to previously suggested approaches is given in Appendix A.

II. MATERIALS AND METHODS

II.A. Data

The data used in this study were obtained from records routinely collected during procedures performed in the cardiac catheterization imaging facility at St. Andrew's War Memorial Hospital, Brisbane, Australia, over the period January 2008–March 2010. The service deals with a comprehensive range of cardiac and peripheral vascular conditions. Procedures can be broadly divided into four categories: Diagnostic and interventional cardiac procedures (cardiovascular disease), electrophysiology and device implant procedures (cardiac rhythm related issues), miscellaneous cardiac procedures (valvuloplasties, defect closures, and pacemaker lead screening), and noncardiac procedures (embolizations, peripheral angiography, and angioplasty). A breakdown of the procedure types along with volumes is provided in Table I. This analysis demonstrates that as a subgroup, diagnostic

TABLE I. Summary of procedure types performed by the SAWMH Cardiac Imaging Service from January 2, 2008 to March 29, 2010.

Procedure classification	Volume	% of total
Cardiovascular disease related procedures		
• Diagnostic coronary angiogram (only) procedures	2656	42
• Coronary angioplasty procedures	1004	16
• Coronary artery bypass graft studies	456	7
Rhythm related procedures		
• Electrophysiology procedures (diagnostic and ablation)	803	13
• Device implants (pacemakers and defibrillators)	1246	20
Miscellaneous cardiac procedures		
• other cardiac procedures (valvuloplaties, defect closures, etc.)	69	1
Noncardiac imaging procedures		
• Diagnostic and interventional peripheral procedures	83	1
Total case load	6317	

coronary angiography (CA) is the most common single study type, accounting for approximately 42% of the service's total workload.

The service operates three imaging systems and we provide a demonstration of the monitoring for two of them, a Philips Allura Xper FD10 system (Philips Medical Systems, Best, The Netherlands) and a Toshiba Infinix CCi–FPD system (Toshiba Medical Systems Corporation, Tokyo, Japan). These two systems account for approximately 70% of all the records. The third imaging system, a Toshiba KXO-80 system (Toshiba Medical Systems Corporation, Tokyo, Japan), is not used for CA procedures and therefore was not included in subsequent analysis. As one of our purposes is to detect persistent changes in the imaging process, we monitor each imaging system separately.

Control charting for process monitoring has two phases.[11] In phase I, we retrospectively examine a set of data to build the charts and derive the control limits. In phase II, we use the charts to monitor the process in real time and take action if the charts signal that the process has changed. To demonstrate in our example, we have divided the existing data into phase I and phase II for each system and have used the relocation of each of the imaging systems to a new facility as the delineation between the two phases.

For the Philips system, the period January 2008 until August 25, 2009 was selected as phase I. The start date was chosen to coincide with a revision to the doses as prescribed by Philips. The end date corresponds to the movement of the imaging system to a new location. There are 1098 case records in phase I. Phase II started on September 28, 2009 when the imaging system began operation at its new location and continued until March 25, 2010. There are 422 case records in phase II. During relocation of this equipment, all cardiologists were required to use the Toshiba system for nonrhythm related procedures.

For the Toshiba system, the period February 2008 until September 25, 2009 was selected as phase I. The start date for this imaging system coincided with a major revision of the unit's dose algorithm. The end date corresponds to the day when the Toshiba system was moved to a new location. Phase I includes 2248 case records. Phase II started on October 20, 2009 when the imaging system began operation at its new location and continued to March 29, 2010. There are 584 case records in phase II. During relocation of this equipment, all cardiologists were required to use the Philips system for nonrhythm related procedures.

For each procedure, many factors were routinely recorded including patient characteristics such as height, weight, body mass index, and resting heart rate as well as procedural factors such as cardiologist, imaging system, fluoroscopy time, digital acquisition (DA) frame count, and total case DAP. Of particular interest were patient weight, number of fluoroscopy frames, and number of digital acquisition frames, as these factors were expected to be the main determinants of DAP, an expectation that arose from a physical understanding of the procedure.

II.B. Regression model building

In developing the monitoring process, we decided to focus on CA procedures only, the most frequently occurring procedure that accounts for approximately 60% of the nonrhythm related cardiac procedures on the two imaging systems. Including other procedures would have increased the volume of available data but would also have greatly increased the variability in DAP values. When we tried to include data from many procedure types, we were unable to build a regression model that could sufficiently compensate for this additional variability. By focusing only on a single (common) procedure, we were able to more quickly detect systematic changes in the imaging systems than if we included the additional DAP variation generated by uncommon procedures. Note that this choice also excludes all procedures other than CA in phase II.

Radiation delivery in CA procedures occurs due to two components; the first involves imaging while the cardiologist manipulates a catheter into position in the patient's heart prior to injecting a radio-opaque dye into the coronary arteries (to enable visualization of vessel disease). This component typically involves a low radiation exposure rate mode of operation (fluoroscopy). The second component involves documentation of the injection of contrast agent down the arteries. This commonly involves acquiring low noise, high resolution images at a higher radiation exposure. This stage is quantified by the DA frame count. The total study DAP is then a measure of the total radiation exposure resulting from the fluoroscopy and DA modes of operation.

With this in mind, we tried to build a good regression model using the phase I CA procedure data for each of the two imaging systems. This involved considerable trial and error, not all of which we describe here in detail. We considered transformations of the DAP score, the inclusion of various factors, and possible interactions among the selected fac-

tors. In Appendix B, we provide a brief description of the regression model building and some helpful references.

Preliminary regression using DAP as the response suggested a systematic "funnel-shaped" pattern in the residual values, where the variation in the residual values increased as the fitted values increase. A common remedy for this is to model the natural logarithm of DAP instead. However, modeling the data in this way results in a multiplicative relationship between DAP and the explanatory variables. This is perhaps less intuitive and more challenging to interpret physically than an additive model, but the model on the log scale provides a better fit to the data and was therefore retained.

One nonlinear relationship evaluated and subsequently rejected was based on a physical model of the imaging process. This is described as

$$\text{DAP} = \beta_0 + (\text{FL}_{\text{time}} * \text{FL}_{\text{rate}})e^{\gamma_1+\beta_1(\text{Wt})}$$
$$+ (\text{DA}_{\text{frames}})e^{\gamma_2+\beta_2(\text{Wt})} + R,$$

where DAP is the case dose area product, Wt is the patient weight, FL_{time} is the case fluoroscopy time, FL_{rate} is the fluoroscopy frame rate, and $\text{DA}_{\text{frames}}$ is the number of DA frames acquired. The β_i's and the γ_i's are regression coefficients, and we assume $R \sim N(0, \sigma_r^2)$. However, this more complicated model did not provide a better fit to the data, and so it was abandoned in favor of the simpler linear relationship for ln(DAP) given in Eq. (1).

We expected that the patient's weight may affect the radiation administered differently for the two stages of the procedure (due to the different tube voltages used in fluoroscopy and digital acquisition). Accordingly, we initially incorporated weight-by-fluoroscopy frame and weight-by-digital acquisition frame interaction terms. We found that neither played a significant role in predicting DAP. An additional concern was whether DAP depended on the cardiologist performing the procedure. To test this notion, indicator variables representing cardiologists were included in the model. However, the inclusion of cardiologists did not explain a significant portion of the variation in the DAP values for either imaging system and so this factor was excluded.

In the end, the model that provided the best fit to the data on both the Philips and the Toshiba systems was of the form

$$\ln(\text{DAP}) = \beta_0 + \beta_1(\text{Wt}) + \beta_2(\text{FL}_{\text{time}} * \text{FL}_{\text{rate}})$$
$$+ \beta_3(\text{DA}_{\text{frames}}) + R, \tag{1}$$

We give the final fitted models in Sec. III A.

II.C. Control chart methodology

We propose to monitor the imaging systems using the residuals from the DAP regression models. Specifically, we monitor the quantities $\ln(\text{DAP})_{\text{observed}} - \ln(\text{DAP})_{\text{predicted}}$ for each new CA case. We address the two objectives discussed in Sec. I A using

• An individuals control chart of the residuals and

• An EWMA control chart for the residuals where extreme values identified in the individuals chart are excluded.

We use the residuals from phase I to determine the control limits for the two charts. We illustrate this in Sec. II D. In phase II, for each case, we calculate the estimated residual, denoted by $\hat{r}_i$ for case i, using the fitted model and then add the residual to the individuals control chart; a plot of $\hat{r}_i$ vs i. If the observed residual falls outside the chart control limits, we flag the individual case as outlying and investigate possible reasons for the unexpectedly large or small radiation dose for case i.

If the observed residual falls inside the control limits on the individuals chart, we add the residual value to the exponentially weighted moving average control chart with the following updating formula:[15]

$$E_i = \lambda\hat{r}_i + (1-\lambda)E_{i-1} = \lambda\hat{r}_i + \lambda(1-\lambda)\hat{r}_{i-1} + \lambda(1-\lambda)^2\hat{r}_{i-2}$$
$$+ \lambda(1-\lambda)^3\hat{r}_{i-3} + \cdots = \lambda\sum_{j=0}^{i-1}(1-\lambda)^j\hat{r}_{i-j} + (1-\lambda)E_0,$$

where λ is a smoothing constant and $0 < \lambda \leq 1$. In our context, $\hat{r}_i$ is the model residual for case i and E_0 is the starting value that we set to zero. Note that the EWMA, as the name suggests, is a weighted average of the residuals for all past cases, where the weights decrease exponentially for cases further in the past. To monitor the process, the EWMA chart is a plot of E_i vs i. If the EWMA chart signals, i.e., falls outside some prespecified control limits, we suspect a systematic and persistent change in the imaging process (equipment output, selection of views, use of collimation, etc.). We recommend a smoothing constant $\lambda = 0.1$, a typical choice in EWMA implementation. We address this choice of λ further in Appendix A.

By removing extreme residual values before plotting the EWMA, we keep the roles of the two charts separate. The individuals chart is designed to detect individual outliers for which the observed DAP is markedly different to that expected for the given parameters, while the EWMA chart detects systematic and persistent shifts in the average DAP. For this reason, our reaction to any signals on either the individuals or EWMA charts should be very different. As it turns out in this application, signals on the individuals chart were often due to data recording errors, such as transposing height and weight measurements. By filtering the residuals that contribute to the EWMA, we introduce a data quality check that reduces the chance that the EWMA will signal systematic problems due to simple data errors.

To implement the proposed monitoring methodology, we must set control limits, and in the case of the EWMA, select a reasonable value for λ. For individuals control charts, the control limits are usually set at ± 3 standard deviations away from the average. We use the standard deviation [e.g., an estimate for σ_r in model (1)] and average of the residuals from the phase I data. Since the residuals come from a linear regression model, the average is zero.
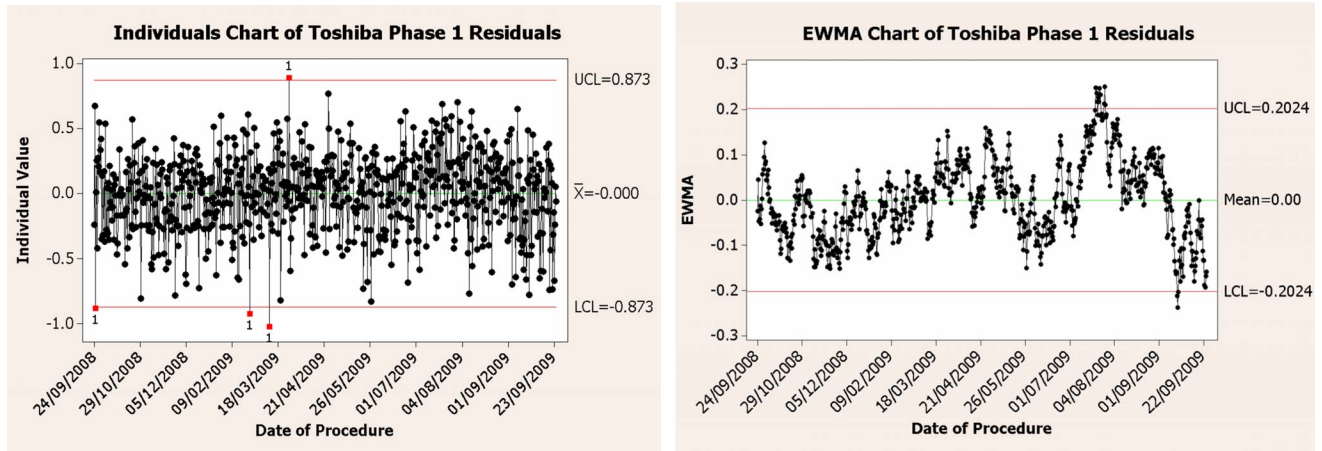
FIG. 1. Individuals and EWMA control charts from original Toshiba phase I data.

The control limits for the EWMA chart are slightly more complicated. One option is to use[13]

$$\pm k\hat{\sigma}_r \sqrt{\frac{\lambda}{2-\lambda}},$$

where k is a constant determined by the user, λ is the smoothing constant, and $\hat{\sigma}_r$ is the residual standard deviation from the regression model (1). The choice of k is driven by the desired in-control average run length (ARL), which is defined as the average number of cases before the EWMA signals (i.e., goes outside the control limits) (see Appendix A for more details).

II.D. Phase I analysis

We assume that phase I data come from a stable (i.e., in-control) process and we use these data to set up the control charts that will be used prospectively in phase II. In practice, using the phase I data to arrive at a final regression model and set up the control charts is an iterative process since we are concerned that the phase I data may come from an unstable process.[14] To achieve a final regression model that best predicts the process, an initial model must be created. Next, we

- Plot the residuals from this model on an individuals control chart with control limits set at $\pm c\hat{\sigma}_r$, where c is a constant (and is normally taken to be 3) and $\hat{\sigma}_r$ is the residual standard deviation.
- Plot the filtered residuals on the EWMA chart, as described in Sec. II C.
- Remove any cases that suggest the process was not stable if reasons can be found. This could include cases whose residuals lie outside the control limits on the individuals chart or series of cases that result in a signal on the EWMA chart.
- Refit the regression model to the remaining data.
- Repeat these three steps until (almost) all of the residuals fall within the control limits.

We illustrate this process with the Toshiba phase I data. Fitting the model (1) to the phase I data gives a residual

standard deviation $(\hat{\sigma}_r)$ of 0.291, so we set the control limits for the individuals chart at ± 0.873. Recall that the average of the residuals is zero, so the chart is centered on zero. Crowder[13] described how to determine λ and k, and hence how to set EWMA control limits that will minimize the out-of-control ARL for a specified shift in the process mean, based on a given in-control ARL. We selected a smoothing parameter of $\lambda = 0.1$ and an in-control ARL of 800 for the Toshiba system (roughly one year's worth of CA cases and equal to the time between mandated major tests of the equipment), which resulted in control limits of ± 0.2024. We give a more thorough explanation of how these limits are derived in Appendix A. Figure 1 shows both the individuals and the EWMA charts from the original phase I.

Next, we examine the control charts to see if there is any evidence of instability. The individuals chart (left panel) shows a small number of out-of-control points, but with $\pm 3\sigma$ limits, a normality assumption, and over 700 CA cases, this is to be expected. The EWMA chart (right panel) is more worrying. We see a signal above the upper control limit, followed by a signal below the lower control limit. Since the number of cases in phase I is reasonably large and because the in-control ARL is 800 cases, a signal is again not that unexpected (even for a stable process). However, further investigation is warranted. The period between August 4, 2009 and September 22, 2009 when the EWMA drops consistently corresponds to the time period when the Philips system was not in use, and hence all cardiologists were using the Toshiba system. It seems that when all of the cardiologists use one imaging system, the process does not behave as it usually does. The two groups of cardiologists who typically use different systems also have different habits when using the same system. This realization suggested changing the Toshiba phase I dates to exclude the time period when only the Toshiba system was in service. This change in phase I dates necessitated the construction of a new Toshiba regression model and the recalculation of control limits. With the new phase I data, the residual standard deviation increased slightly to 0.3033 and so the control limits for the individuals chart are now ± 0.910 and the control limits for the EWMA
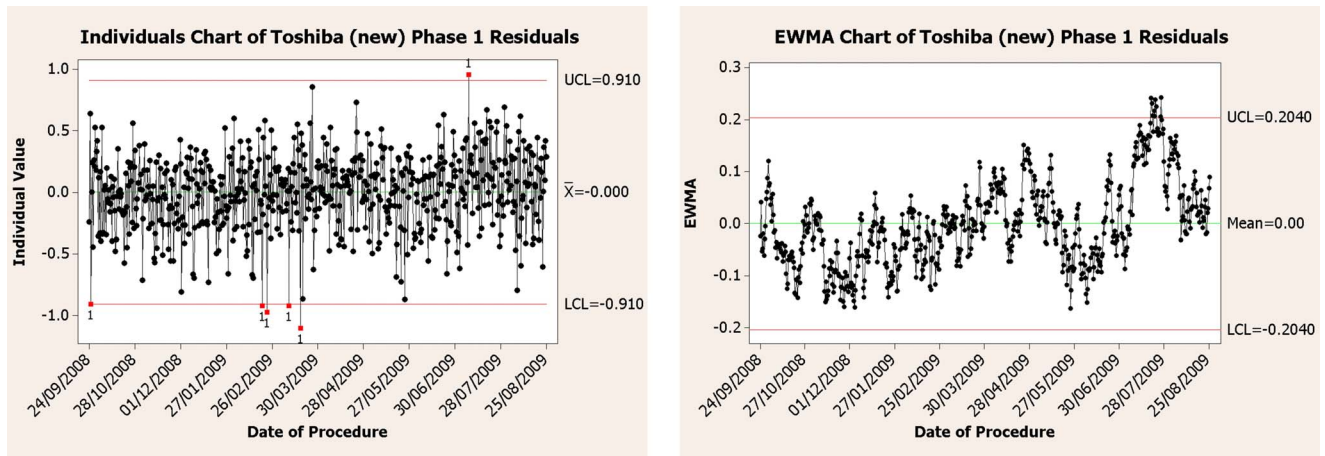
FIG. 2. Individuals and EWMA control charts from reduced Toshiba phase I data.

are now ± 0.2040. The individuals and EWMA charts for the new Toshiba phase I dates are given in Fig. 2. We now conclude that the remaining phase I data could reasonably have come from a stable process and we move on to phase II.

The phase I construction of control limits for the Philip's system required no iteration and we do not show the resulting phase I control charts here. The residual standard deviation from the regression model was 0.2417, resulting in individuals chart control limits of ± 0.725. We selected an in-control ARL of 400 for the Philips system (again this corresponds to 12 months of CA cases), which resulted in EWMA control limits of ± 0.1490.

III. RESULTS

III.A. Phase II analysis

We demonstrate the monitoring approach using the phase II data for both the Philips and the Toshiba systems. In actual application, we would build the phase II control charts in real time as patients undergo the imaging procedure. The regression models we used as a result of the model building described in Sec. II are as follows:

Toshiba:

$$\ln(DAP) = 0.666 + 0.0162 * (Wt)$$
$$+ 0.00584 * (FL_{time} * FL_{rate})$$
$$+ 0.00175 * (DA_{frames}),$$

$R^2 = 68.3\%, \quad \hat{\sigma}_r = 0.3033.$

Philips:

$$\ln(DAP) = 0.825 + 0.0191 * (Wt)$$
$$+ 0.00567 * (FL_{time} * FL_{rate})$$
$$+ 0.00122 * (DA_{frames}),$$

$R^2 = 83.3\%, \quad \hat{\sigma}_r = 0.2417.$

As mentioned in Sec. I A, the purpose of constructing regression models and using the residuals is to reduce the variation in the plotted data; hence, making the charts more sensitive to systematic process changes. In Table II, we see reductions of approximately 66% for the Philips system and 48% for the Toshiba system.

Next, we plot the control charts for the phase II data with the control limits as derived in phase I. The phase II control charts for the Toshiba system are given in Fig. 3. Because there were no out-of-control points identified by the individuals control chart, no cases were omitted in the construction of the EWMA chart. As is evidenced by the significant increasing trend in the EWMA chart in Fig. 3, the process seems to have shifted upward at the start of the phase II period. Since the EWMA remains (mostly) positive, it appears that the upward shift is sustained. In other words, we found that the DAP values in this period were larger than expected, even after adjusting for the factors in the regression model. In real time, we would have examined the imaging system carefully when the signal on the EWMA chart first occurred.

A possible explanation for the shift in the process lies in what happened to the imaging equipment between the phase I and phase II periods. The equipment was relocated to a new cardiac imaging facility; in so doing, it was disassembled into its major components and reinstalled in a new location. During the reinstallation process, the imaging equipment was fully recalibrated. Although a standardized process, the complexity of the system is such that it is likely there were minor inconsistencies in the performance of the systems between the two periods (as evidenced by the data).

For the Phillips system, we see a different pattern on the charts shown in Fig. 4. There are several unusually large residuals shown on the individuals chart. Further retrospec-

TABLE II. Standard deviations of raw data and residual values for both imaging systems.

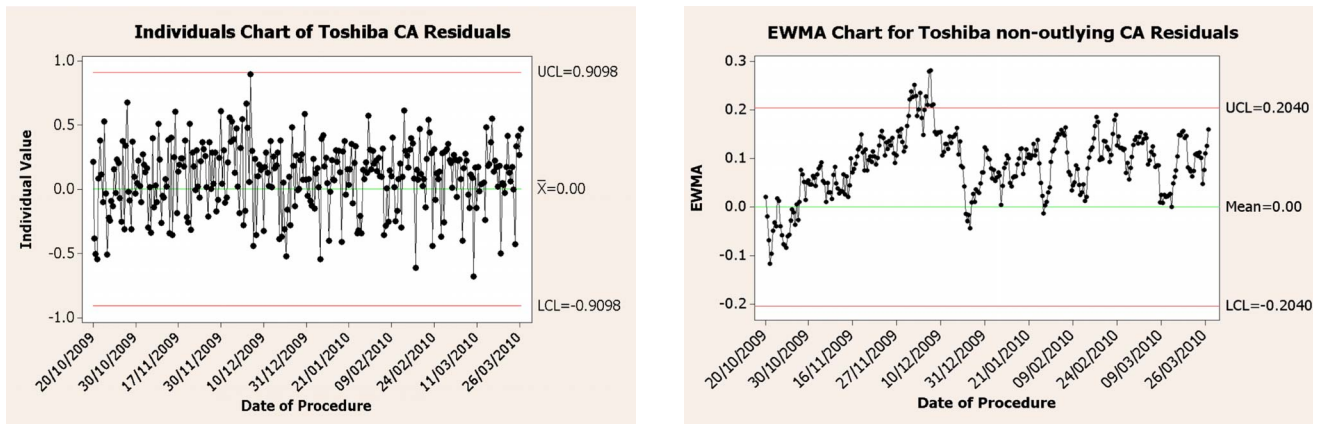| System	ln(DAP) St. deviation	Residual St. deviation
Philips FD10	0.7212	0.2417
Toshiba Cci–FPD	0.5799	0.3033

FIG. 3. Individuals and EWMA control charts of Toshiba phase II residuals.

tive investigation of these cases failed to reveal any clear reasons for the unusual residuals. However, had the charts been constructed in real time the investigation would have been more timely and reasons may have been found. The right panel of Fig. 4 gives the EWMA control chart for the same data with the four outliers from the individuals chart removed.

Looking at the EWMA chart in Fig. 4, we see that the process exceeds the upper control limit around October 13, 2010 and drops below the lower control limit around December 11, 2010. An explanation exists for the first signal on the EWMA. It was during this time period that only the Philips system was in use, and hence all cardiologists were performing CA procedures on the Toshiba system. As described in Sec. II D, when all of the cardiologists use one system, the process does not behave as it usually does. This suggests that while we found no substantial differences between cardiologists on a single system, there appears to be a substantial difference between cardiologists when grouped by their preferred imaging system. Generally, the cardiologists who regularly use the Toshiba system have an inherently lower average DAP than those who normally use the Philips system (taking into account the frame rate difference between the two systems). With regard to the second signal, to determine whether deterioration in the performance of the imaging system has occurred, or whether some other factor is responsible, an investigation to search for the cause of this drift is necessary.

IV. DISCUSSION

Driven largely by improvements in technology and clinical understanding, the practice of medical imaging is dealing with increasingly complex procedures.[14] While it can be argued that this change has resulted in substantial improvements in healthcare, recent events in the United States also reveal a corresponding increase in the opportunity for failures of the overall imaging process resulting in significant and unnecessary risk increases for individual patients and the population as a whole.[2] As was the case with the brain perfusion CT cases, these failures are not constrained to malfunctions of the imaging equipment but can arise due to unintended and unacknowledged changes in the imaging process.[1]

At present, the recommendation of most professional bodies is for ionizing radiation based imaging equipment to un-
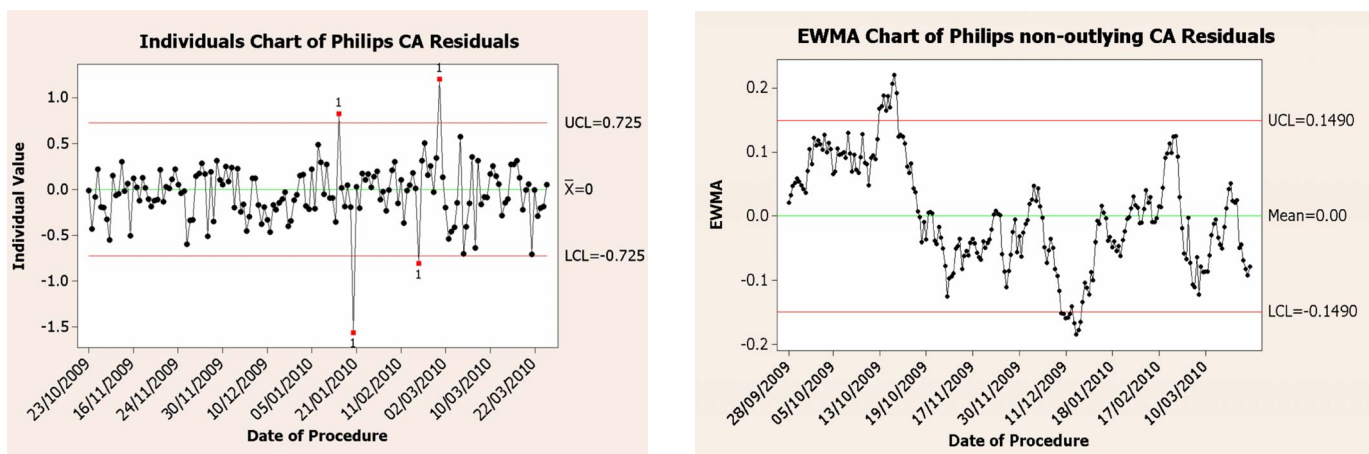

FIG. 4. Individuals and EWMA control charts of Philips phase II residuals.

dergo regular performance assessments. For the type of equipment used in cardiac angiography (as noted in this paper), this testing is normally performed every 12 months using standardized protocols.[6] However, most testing procedures are designed to be independent of the patient, assessing discrete aspects of the performance of the equipment with emphasis placed on image quality (resolution and contrast) and radiation output. The limitation of this process is that it does not take into consideration the potential for noncatastrophic failures to occur (that may result in radiation output increases) at a point in time between the major performance assessments, nor does it account for unintended variations (specifically increases) in radiation use arising from changes in clinical practice or system protocols. For angiography systems, this might include frame rate changes, changes in collimation practice, and use of different (less efficient dose vs clinical yield) imaging projections.

Although it is possible that certain discrete and significant changes (equipment failures and protocol changes, for example) would be detected promptly by imaging staff, the sophistication of newer digital imaging systems (e.g., CT scanners and cardiac and vascular angiography systems) is such that modest changes in radiation output may be overlooked due to the imaging system's inherent ability to compensate for patient load and hence detector input variation. The process proposed here addresses these concerns by supplementing the conventional testing program with a secondary process that makes use of clinically derived data acquired routinely by most imaging services as part of their regulatory obligations.

The success of the proposed monitoring scheme depends critically on the quality of the data and the fitted regression model. In this example application, the routine capture of the factors of interest such as patient weight and frame rate was not introduced until September 2008. As such, these data were not always recorded, or were occasionally recorded improperly. Accordingly, in phase I, approximately 40% of the cases could not be used to estimate the parameters for the final regression models. Later, in phase II, far fewer cases were missing factor values, but there were still occasionally errors that became clear upon closer inspection of particular data records. To utilize this monitoring methodology at its full potential, it is important to keep complete and accurate documentation of all cases.

To maintain the effectiveness of the proposed methodology, the models and control limits should occasionally be updated to allow for accepted process changes such as technological improvements or changes in clinical practice. We recommend refitting the model and updating the control limits whenever a substantial process change has taken place. A regression model update may also be required if the process signals frequently, or for prolonged periods of time.

When an update of the regression model and control charts is deemed warranted, we are faced with the difficult questions of how much data we need. From a statistical perspective, more data are better as this will allow us to better model the relationship between the response and the factors and estimate the unexplained variability. However, waiting until a large amount of data is available delays the application of the monitoring. We propose a compromise. To get the monitoring started quickly, we initially use the first 20 cases, only to estimate a new model intercept term (β_0), which is estimated as the original β_0 estimate minus the average residual (from the original model) for the 20 cases. In other words, until we have more data, we assume that the effect of the factors remains the same and only the average DAP value has changed. Once we have collected additional new cases, say at least 100, we can investigate to see if a new regression model is necessary.

In our example, we decided to base the monitoring on only cases with CA procedures. We investigated including other procedures in the model building but were unable to construct a single model that was appropriate for more than CA procedures. An alternative idea that we did not pursue here is to build a separate model for each procedure and then construct the two control charts using estimated residuals from each model, standardized to have the same variability. This approach assumes that if the imaging system drifts, then the effect is the same across all procedures included in the monitoring.

The methodology that we propose here can be applied to any measured output, such as cumulative air kerma, from a high volume imaging process. If the volume is low, then the monitoring procedures are likely to be ineffective since it will take a long time relative to performance standards to detect any process shifts.

The chief value of the proposed monitoring scheme is that it operates in real time and will detect underlying shifts in average DAP levels as quickly as possible. There is substantial administrative cost due to the requirement to maintain the records for each case and to review the charts looking for signals. The charting can be easily automated using standard software such as EXCEL.

V. CONCLUSIONS

The process outlined in this paper uses two control charts, informed by clinically derived data, to monitor the use of radiation by system. It is offered as a supplement to existing equipment performance assessment programs. The benefit of this process is twofold; first, as the data are captured and analyzed on a case by case basis, data errors can be identified and addressed promptly and, second, small sustained shifts in the output of the imaging system (arising from equipment performance drift or changes in imaging practice) can be identified. Although performance of the example charts provided in this paper may appear overly sensitive, the minor inconvenience caused by investigating the signals arising is a small price to pay when compared to the benefits for patients (and associated personnel).

APPENDIX A: EWMA CHARTS

Assuming a Guassian model for the residuals, we can derive the properties of the combined individuals and EWMA control charts using a Markov chain approximation.[15] The approximation suggests that for an in-
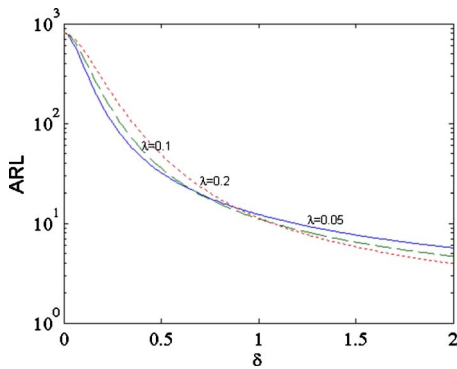
FIG. 5. Average run length as a function of mean shift for different values of λ.

control average run length of 400 and $\hat{\sigma}=0.24$, the EWMA control limits should be set at ±0.1490, and for an in-control average run length of 800 and $\hat{\sigma}_r=0.30$, the control limits should be set at ±0.2040. This coincides very closely with what Crowder *et al.*[13] suggests when 1 $\hat{\sigma}_r$ above or below the mean is the magnitude of the shift in the process that must be detected quickly.

We can also use the Markov chain approximation to illustrate how quickly we expect the EWMA to detect a persistent shift in the mean DAP. Denoting the mean μ and standard deviation σ, we let $\mu_\delta=\mu\pm\delta\sigma$. Thus, when we speak of a 1 $\hat{\sigma}$ shift in the mean (i.e., $\delta=1$), we mean that the average of the residuals has increased or decreased by 1 standard deviation and is hence given by $\mu_1=\mu\pm\sigma$. Recall that these residuals are from a regression model where ln(DAP) is the response, and so a shift in the mean refers to a shift on the log scale.

Figure 5 plots the ARL of the Toshiba system vs δ (the number of $\hat{\sigma}$-shifts in the mean) when we assume the residuals are normally distributed. Three lines are depicted. Each represents the change in ARL for common values of λ (EWMA smoothing constant). We used $\lambda=0.1$, and so any discussion below is in reference to the dashed line.

The results shown in Fig. 5 give the change in ARL using the proposed monitoring scheme, where outlying points detected on the individuals control chart are not included in the EWMA. Assuming the control limits on the individuals chart are set at $\pm3\hat{\sigma}$, we found that there is very little difference between the ARL of the proposed monitoring scheme and the EWMA where all points are included.

We can use the results in Fig. 5 to quantify how quickly we expect to detect systematic changes in the average ln(DAP) for the Toshiba system. Table III gives the ARL for specific values of δ and gives the resultant actual shift in the average DAP for the Toshiba system. Note that for a clearer understanding of this concept, the information in this chart has been transformed from the log scale to the actual DAP scale. For instance, Table III suggests we can expect that a 1 $\hat{\sigma}$ shift in the average ln(DAP) will, on the average, signal as out-of-control in approximately 18 cases, and that a shift of this magnitude translates into a shift in the average DAP of 6.8 Gy cm^2, an increase of approximately 35%.

TABLE III. ARL for various shifts in the mean DAP value for the Toshiba CA data with $\lambda=0.1$.

δ	μ_{start}	μ_{end}	% increase	ARL
0	19.3	19.3	0	800
0.5	19.3	22.5	16.4	59.1
1	19.3	26.1	35.4	18.4
1.5	19.3	30.4	57.6	10.8

Figure 6 explores the best choice of the smoothing constant λ as a function of the size of shift in the mean ln(DAP) values. We define best as the value of λ that yields the smallest ARL when the ln(DAP) mean has shifted while still having fixed in-control (i.e., no mean shift) ARL. Figure 6 suggests $\lambda=0.1$ is optimal for a mean shift of $\delta=0.84$.

It is of interest to compare the performance of our proposed EWMA chart based on filtered model residuals with the previously suggested methods such as using only an individuals control chart based on the raw data.[10] It is well known that an individuals chart is good at detecting large shifts in the process, while an EWMA chart is better suited to detecting small to medium sustained shifts.[11] However, in our context to fairly compare the approaches, we must also consider the influence of monitoring the raw ln(DAP) values rather than model residuals. Assuming the model is reasonable, using residuals has the advantage of reducing the unexplained variation, thereby making it easier to detect process changes. However, in our context, this advantage is partially offset because we could only successfully build a model for the CA procedure and thus lose some data in our resulting monitoring scheme.

For illustration, we consider monitoring the Toshiba system. The regression model R^2 is 0.68; thus, the standard deviation of the model residuals is $0.566(\approx\sqrt{1-R^2})$ times the residuals of the raw ln(DAP) values. Hence, assuming process mean shifts effect all procedures equally, a 1 standard deviation shift in mean ln(DAP) corresponds to a 1.77 (1/0.566) standard deviation shift in the mean of the residuals. This captures the effect of the modeling (assuming there is no estimation error in the model). We use the Markov chain approximation to determine the ARL of the two methods. To account for the fact that the CA procedure rep-
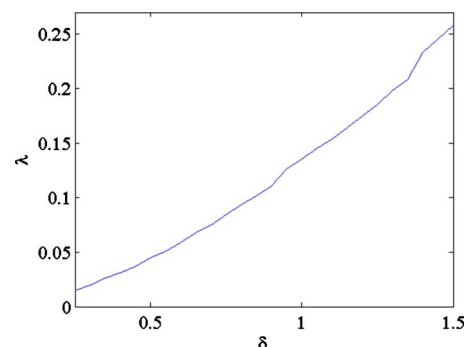

FIG. 6. Optimal smoothing constant λ to detect different sized mean ln(DAP) shifts.

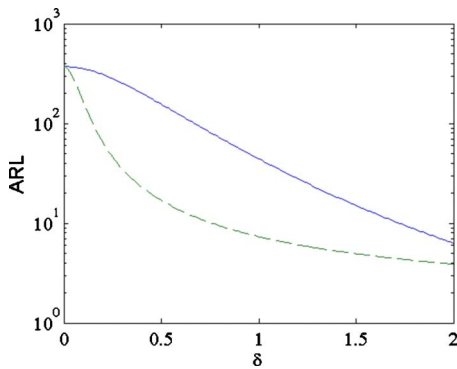FIG. 7. Comparison of the average run length for two monitoring approaches. Solid line—Individuals chart on raw ln(DAP). Dashed line—Combined individuals and EWMA chart on residuals.

resents only about 60% of all the nonrhythm related cardiac procedures, and thus that the filtered EWMA approach loses some data, we divide the ARL values for the EWMA by 0.6 to make them comparable to the individuals chart.

In Fig. 7, we compare the performance of the two approaches. For the individuals charts, we use the usual ±3 standard deviation unit control limits. For the approach of Gerard *et al.*,[10] this results in an in-control (i.e., when there is no mean shift) ARL of 370. We then determine the control limit for the filtered EWMA so that, with the 0.6 factor adjustment, the in-control ARL is also 370. Figure 7 suggests that the proposed filtered EWMA method, despite not using all the data, is superior to an individuals chart based on the raw data at detecting ln(DAP) mean shifts of less than about 2.5 standard deviation units. Table IV quantifies the results numerically. For instance, for a mean shift in ln(DAP) of 1 standard deviation unit, we expect the filtered EWMA chart to signal, on the average, after only 7.3 cases, while the individuals chart is expected to signal after about 44 cases.

APPENDIX B: REGRESSION MODELING

For each case, a regression model describes the relationship between a response, such as log(DAP), and the measured factors or explanatory variates such as patient weight and fluoroscopy time. The model has the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + r_i,$$

where, for the *i*th case, y_i is some function of DAP and $x_{i1}.,\ldots,x_{ip}$ are functions of the explanatory variates. The residuals r_i account for the variation in y_i not explained by

TABLE IV. Comparison of ARLs by mean shift size.

Methodology	Mean shift			
	$\delta=0$	$\delta=0.5$	$\delta=1$	$\delta=1.5$
Individuals chart	370	155.2	43.9	15
Filtered EWMA chart	370	16.8	7.3	4.9

the explanatory variates. We use the available phase I data to build the model, i.e., choose y_i and $x_{i1}.,\ldots,x_{ip}$, estimate the coefficients $\beta_0,\ldots,\beta_p$ and the residuals r_i, determine if there are any outliers, and so on.[12,16] Building the model is highly iterative in that we try different possibilities and use the tools given in the above references to select among candidate models.

If we can successfully build such a model, then the variation in the estimated residuals will be substantially smaller than the variation in the response. In phase II, if the process drifts, then β_0 will change, resulting in a shift in the mean of the response and the estimated residuals. Because the variation in the estimated residuals is smaller, a control chart based on these residuals will detect the change more quickly than will a chart based solely on the response.

a)Electronic mail: ismith@sawmh.com.au

[1]M. Wintermark and M. H. Lev, "FDA investigates the safety of brain perfusion CT," AJNR Am. J. Neuroradiol. **31**(1), 2–3 (2010).

[2]U.S. Food and Drug Administration, Initiative to Reduce Unnecessary Radiation Exposure from Medical Imaging, 2010 [updated 2010; cited 18 June 2010], available from http://www.fda.gov/Radiation-EmittingProducts/RadiationSafety/RadiationDoseReduction/ucm199994.htm.

[3]I. R. Smith and J. T. Rivers, "Measures of radiation exposure in cardiac imaging and the impact of case complexity," Heart Lung Circ. **17**(3), 224–231 (2008).

[4]A. J. Einstein, K. W. Moser, R. C. Thompson, M. D. Cerqueira, and M. J. Henzlova, "Radiation dose to patients from cardiac diagnostic imaging," Circulation **116**(11), 1290–1305 (2007).

[5]T. Topaltzikis, C. Rountas, R. Moisidou, I. Fezoulidis, C. Kappas, and K. Theodorou, "Radiation dose to patients and staff during angiography of the lower limbs. Derivation of local dose reference levels," Phys. Med. **25**(1), 25–30 (2009).

[6]D. A. Causer *et al.*, "Recommendations for a technical quality control program for diagnostic x-ray equipment," Australas. Phys. Eng. Sci. Med. **28**(2), 69–75 (2005).

[7]American College of Radiology, "ACR standard for diagnostic medical physics performance monitoring of radiographic and fluoroscopic equipment," *Standards* (American College of Radiology, Reston, VA, 1999), pp. 167–169.

[8]Diagnostic X-Ray Imaging Committee Task Group 12, The American Association of Physicists in Medicine (AAPM), "Quality control in diagnostic radiology," Report No. 74, 2002.

[9]T. Pawlicki, M. Whitaker, and A. L. Boyer, "Statistical process control for radiotherapy quality assurance," Med. Phys. **32**(9), 2777–2786 (2005).

[10]K. Gerard, J. P. Grandhaye, V. Marchesi, P. Aletti, F. Husson, A. Noel, and H. Kafrouni, "Feasibility study of using statistical process control to optimize quality assurance in radiotherapy," J. Qual. Maint. Eng. **15**(4), 331–343 (2009).

[11]D. C. Montgomery, *Introduction to Statistical Quality Control* (Wiley, Hoboken, NJ, 2009).

[12]D. E. Matthews and V. T. Farewell, *Using and Understanding Medical Statistics*, 4th ed. (S. Karger AG, Basel, 2007).

[13]S. V. Crowder, "Design of exponentially weighted moving average schemes," J. Quality Technol. **21**(3), 155–162 (1989).

[14]E. S. Amis, Jr. *et al.*, "American College of Radiology white paper on radiation dose in medicine," J. Am. Coll. Radiol. **4**(5), 272–284 (2007).

[15]J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: Properties and enhancements," Technometrics **32**(1), 1–12 (1990).

[16]D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 4th ed. (Wiley, Hoboken, NJ, 2006).