

Research Article

Segmented Linear Regression Models for Assessing Change in Retrospective Studies in Healthcare

Epaminondas Markos Valsamis ¹, David Ricketts,¹ Henry Husband,² and Benedict Aristotle Rogers¹

¹Brighton and Sussex University Hospitals NHS Trust, Trauma and Orthopaedic Department, Brighton BN2 5BE, UK

²Faculty of Mathematics, University of Cambridge, Cambridge CB3 0WA, UK

Correspondence should be addressed to Epaminondas Markos Valsamis; markosvalsamis@gmail.com

Received 17 June 2018; Revised 6 December 2018; Accepted 31 December 2018; Published 22 January 2019

Academic Editor: Luca Faes

Copyright © 2019 Epaminondas Markos Valsamis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. In retrospective studies, the effect of a given intervention is usually evaluated by using statistical tests to compare data from before and after the intervention. A problem with this approach is that the presence of underlying trends can lead to incorrect conclusions. This study aimed to develop a rigorous mathematical method to analyse temporal variation and overcome these limitations. *Methods.* We evaluated hip fracture outcomes (time to surgery, length of stay, and mortality) from a total of 2777 patients between April 2011 and September 2016, before and after the introduction of a dedicated hip fracture unit (HFU). We developed a novel modelling method that fits progressively more complex linear sections to the time series using least squares regression. The method was used to model the periods before implementation, after implementation, and of the whole study period, comparing goodness of fit using *F*-tests. *Results.* The proposed method offered reliable descriptions of the temporal evolution of the time series and augmented conclusions that were reached by mere group comparisons. Reductions in time to surgery, length of stay, and mortality rates that group comparisons would have credited to the hip fracture unit appeared to be due to unrelated underlying trends. *Conclusion.* Temporal analysis using segmented linear regression models can reveal secular trends and is a valuable tool to evaluate interventions in retrospective studies.

1. Introduction

The National Health Service has a strong culture of quality improvement [1]. A good evidence base is needed to drive this process, and these data are often derived from retrospective studies and audits [2].

An important part of the audit cycle is comparison of data before and after a given intervention [3]. Typically, mean values or ranks of outcome measures before and after the intervention are compared, and differences are tested statistically for significance. This method assumes stationarity of data which is often not the case. Nonstationarity of data is difficult to distinguish from causal change, and group comparisons are unable to distinguish underlying (secular) trends from intervention-induced change. This can lead to erroneous conclusions [4–6] and is illustrated by a simulated

example shown in Figure 1. This shows a hypothetical intervention that made no contribution to a change. Group comparison would attribute this to the intervention ($p = 0.0002$). While this is obvious in this example, more subtle trends obscured by highly variable data are more challenging to identify. Researchers attempt to limit the influence of trends by selecting data closer to the point of the intervention, but this does not annul the effect of a secular trend when one is present [6]. It also sacrifices much of the data, although it should be noted that data spanning long time intervals are more susceptible to confounding parameters that are external to the purpose of the study.

Alternative approaches that accommodate trends use interrupted linear or higher-order regression, but these suffer from bias because the interruption is chosen at the point of the intervention [7, 8]. Randomised controlled trials

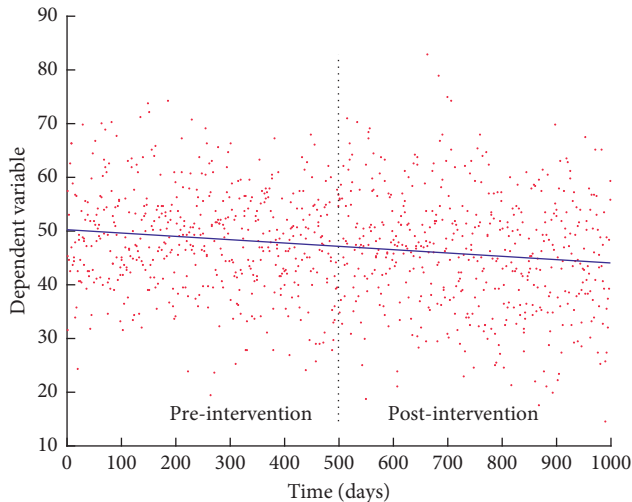


FIGURE 1: Gaussian noise superimposed on a mildly decreasing trend highlights a situation where the mean of pre-intervention measure is significantly different from that of the post-intervention measure ($p = 0.0002$) when the change is entirely due to the trend and not the intervention. The dashed vertical line separates the two groups. The blue line represents the line of regression of the mean value of the simulated variable.

avoid many of these issues but are often expensive and time consuming. Consequently, any method that can improve the reliability of retrospective studies has wide potential application to the wealth of data that such studies make available at relatively low cost.

The aim of this study was to describe a novel method of modelling the temporal variation of a time series of data to help identify secular trends in retrospective studies. Such a method should reliably reveal underlying trends spanning the entire study period, without bias to the intervention. Importantly, there should be a mathematically robust technique to determine whether a trend in the data is significant or due to random variation. We developed and applied a novel method to analyse the temporal variation in hip fracture outcomes in a retrospective study that aimed to evaluate the implementation of a dedicated hip fracture unit (HFU).

2. Methods

2.1. Hip Fracture Unit. In July 2015, a level 1 Major Trauma Centre (MTC) in the United Kingdom established a dedicated HFU to free up beds within the MTC. The HFU was located in a nearby district general hospital. Services were reconfigured to include ambulance triage, daily consultant-led dedicated theatre lists for proximal femoral fractures, and coordination of the necessary multidisciplinary team of staff at the district hospital. These changes were all instituted simultaneously on July 1st, 2015.

2.2. Study Group. We studied 2777 patients who sustained proximal femoral fragility fractures. Of these, 2117 patients (2176 fractures) sustained their fractures prior to commencement of the HFU (study period April 2011–June

2015), and 660 patients (671 fractures) were treated after the introduction of the HFU (study period July 2015–September 2016).

2.3. Data. From a retrospective review of patient notes, the following data were obtained:

- (1) Time to surgical intervention (hours from time of admission)
- (2) Median length of hospital stay in days (LOS)
- (3) Mortality rate at 30 days, 120 days, and 365 days

Patient demographic data were also collected. Both hospitals were part of the same trust, and as such the same sources of data collection were used: “e-Oasis” and “Symphony” databases.

2.4. Statistical Analysis. Data before and after the intervention (introduction of the HFU) were compared. The data were analysed using both conventional statistical tests and temporal analysis using a novel model-fitting method. MATLAB (Natick, MA, USA) was used to apply segmented least squares regression to determine parameters of the nested models that are fitted to the time series [9, 10]. The models were compared for goodness of fit using F -tests to determine the best-fitting model. The large number of data points ($n > 2800$) helped overcome limitations arising from the non-normality of data to validate the use of F -tests to compare nested models [11].

For conventional statistical tests, continuous variables (time to surgery and length of stay) were compared using Mann–Whitney tests while categorical data (30-day, 120-day, and 365-day mortality rates) were compared using Fisher’s exact tests. Demographic data were compared using Mann–Whitney tests (age) and chi-squared tests (gender, pathological fracture, ASA grade, and re-operation). These tests were applied using the MedCalc software suite (Ostend, Belgium) [12]. Statistical significance was set at $p < 0.05$.

2.5. Segmented Linear Regression Model Fitting. In this study, we used an adaptation of a segmented linear regression technique previously published by one of the authors (EMV) in the context of learning curve modelling [9, 13].

We used least squares regression to fit a set of four progressively more complex models to data of the time series. Our models ranged from a simple plateau to two adjoining straight lines (linear splines) with a single knot at the point in time that minimised the sum of the square of residuals. This range of models accommodates a respectable degree of complexity that can capture changing trends in the study period but also maintains simplicity and meaningfulness in the descriptions of trends.

The four models, in increasing order of complexity, are as follows:

- (i) *Plateau model.* This is the simplest model. It is a horizontal line (plateau) at the average value of the outcome measure over the study period.

$$y = \bar{y}, \quad (1)$$

where $\bar{y} = \sum_i y_i / n$.

- (ii) *Line model*. The next model is a straight line determined by least squares regression of the variable (y) on the time (t):

$$y = mt + c, \quad (2)$$

where $m = \sum_i t_i y_i / \sum_i t_i^2$ is the gradient and $c = (\sum_i y_i / n) - (m \sum_i t_i / n)$ is the y -axis intercept.

- (iii) *Line-plateau (or plateau-line) model*. This model is made up of two sections: a straight line adjoining a horizontal plateau at the k th value of t ($t = t_k$) or vice versa. The parameters of each section (gradient and intercept) are determined by least squares regression while the adjoining point ($t = t_k$) is determined as that which minimises the sum of squares of residuals (Appendix A).

The line-plateau model is given by $y_i = m_k t_i + c_k$ for $i = 1$ to k and $y_j = Y_k$ for $j = k + 1$ to n where m_k and c_k are the gradient and y -axis intercept of the straight line, respectively, and Y_k the value of the plateau.

The plateau-line model is given by $y_i = c_k$ for $i = 1$ to k and $y_j = m_k (t_j - t_k) + c_k$ for $j = k + 1$ to n where m_k is the gradient of the straight line and c_k the value of the plateau.

- (iv) *Line-line model*. This model is made up of two sections: a straight line adjoining a second straight line at the k th value of t ($t = t_k$). The two straight lines are determined by least squares regression of the variable (y) on the time (t) using a closed form. The adjoining point ($t = t_k$) is determined as that which minimises the sum of squares of residuals (Appendix B).

The resulting lines have equations $y_i = m1_k t_i + c_k$ ($i = 1$ to k) and $y_j = m2_k (t_j - t_k) + y_k$ ($j = k + 1$ to n) where $m1_k$ and c_k are the gradient and y -axis intercept of the first line and $m2_k$ is the gradient of the last straight line.

After all four models have been fitted, we select the best model as the simplest one unless a more complex one fits the data significantly better. Significance is confirmed by F -tests that compare the sum of the squares of residuals between models. A simple tabular method listing the p values of F -tests conducted pairwise between all four models helps select the best model. This method is illustrated in Table 1. This technique (F -test as applied to nested models) accounts for and prevents, the risk of overfitting whereby higher-order models (e.g., line-line) will always demonstrate improved fit, but at the expense of additional parameters [14].

2.6. Application of the Method to the Study. Our proposed method for selecting the best model was applied over three distinct periods of time: (i) before the intervention (pre-HFU), (ii) after the intervention (post-HFU), and (iii) the entire time series (including both pre-HFU and post-HFU periods). The model describing the entire time series was

TABLE 1: A tabular illustration of the results of F -tests between models.

	Plateau	Line	Line \longleftrightarrow plateau
Line	$P_{\text{line/plateau}}$		
Line \longleftrightarrow plateau	$P_{\text{line-plateau/plateau}}$	$P_{\text{line-plateau/line}}$	
Line-line	$P_{\text{line-line/plateau}}$	$P_{\text{line-line/line}}$	$P_{\text{line-line/line-plateau}}$

The best model is the one that has a row of uninterrupted significant p values ($p < 0.05$) stretching furthest to the right. In the case of a tie, the lower-order model is preferred. If no significant p values feature in the table, the plateau is chosen as the best model.

subsequently compared to the models for the separate pre-HFU and post-HFU periods using F -tests to determine whether separate models offered a significantly better fit.

3. Results

3.1. Time to Surgery. Conventional tests: median time to surgical intervention decreased after introduction of the HFU. This change was not significant (21.51 hours pre-HFU to 20.75 hours post-HFU, Mann-Whitney test: $p = 0.150$).

Temporal analysis: the best-fitting model for the entire data set was the line-plateau (Figure 2), illustrating that time to surgery decreased during the initial period of the study, well before the introduction of the HFU. The line-line model was the best model for the period pre-HFU and showed a decreasing trend followed by an increasing trend. The best model post-HFU was the plateau, taking a value at about the average plateau of the line-plateau model fitted on the entire study period. Separate models for the periods before and after the intervention did not offer a better fit compared to using a single model for the entire time series ($p = 0.100$). As a result, a single model over the entire period offered the best temporal description, suggesting that a decrease in the time to surgery occurred early in the study period, well before the intervention.

3.2. Length of Stay. Conventional tests: median length of stay (LOS) decreased after introduction of the HFU, though not significantly (15 days pre-HFU to 14 days post-HFU, Mann-Whitney test: $p = 0.410$).

Temporal analysis: the best-fitting model for the entire time series was the line-line model (Figure 3). It demonstrated a slowly decreasing rate in LOS followed by a sharply decreasing rate in LOS that started about half a year after the introduction of the HFU. The onset of this sharp change in the rate explained the difference in median LOS pre- and post-HFU. Separate pre-HFU analysis and post-HFU analysis demonstrated the single line to be the best model in both. Separate models offered a better fit than the overall model ($p = 0.017$) confirming that the intervention (HFU) contributed to change by causing an acceleration in the decrease of LOS.

3.3. 30-Day Mortality. Conventional tests: 30-day mortality rate decreased significantly after introduction of the HFU (5.47% pre-HFU to 3.13% post-HFU, Fisher's Exact test: $p = 0.014$).

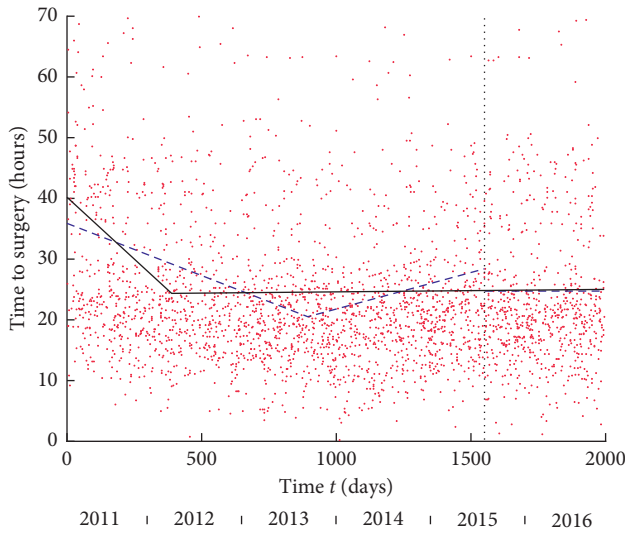


FIGURE 2: Time to surgery versus time t in days from the start of the study. The vertical dashed line marks the onset of the HFU. The black line shows line-plateau as best-fitting linear model for entire period: the line has equation $y = -0.0414t + 40.1868$; plateau at $y = 24.7033$ reached after 375 days. Dashed blue lines are best-fitting linear models for pre-HFU and post-HFU periods. For pre-HFU period, the best model in line-line model with first line $y = -0.0172t + 35.8863$ and second line $y = 0.0120(t - 894) + 20.5137$ with delimiting time 894 days. For post-HFU period, the best model is the plateau at $y = 24.7359$.

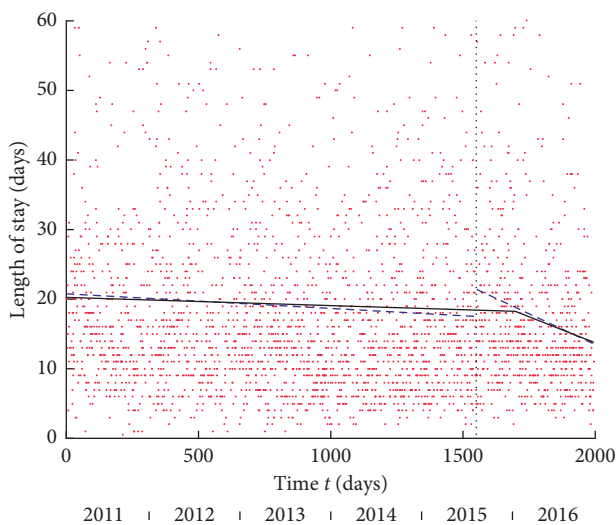


FIGURE 3: Length of stay versus time t in days from the start of the study. The vertical dashed line marks the onset of the HFU. The black line shows the line-line model as the best-fitting linear model for entire time period: first line $y = -0.0012t + 20.2556$ and second line $y = -0.0150(t - 1698) + 18.2532$ with delimiting time at 1698. The dashed blue lines are the best-fitting linear models for the pre-HFU and post-HFU periods. The best model for the pre-HFU period is the single line with equation $y = -0.0021t + 20.7696$, and best model for post-HFU is the single line with equation $y = -0.01769(t - 1550) + 48.8532$.

Temporal Analysis: The single line model was the best model when analysing the entire time period, suggesting that 30-day mortality rate followed a declining trend throughout

the study period (Figure 4). Separate pre-HFU and post-HFU analysis demonstrated that the plateau fit best in both. Separate models did not offer a better fit compared to the overall model ($p = 1.000$). We concluded that 30-day mortality most probably followed a declining trend over the entire study period which appeared to be responsible for the significant reduction found using group comparison. This means that the significant reduction in mortality rates from pre-intervention to post-intervention cannot be simply credited to the intervention.

3.4. 120-Day Mortality. Conventional tests: 120-day mortality rate decreased, but not significantly, after introduction of the HFU (12.68% pre-HFU to 10.13% post-HFU, Fisher's Exact test: $p = 0.078$).

Temporal analysis: the single line was the best model for the entire time period (Figure 5). Separate pre-HFU and post-HFU analysis demonstrated the line model fit best for both. Separate models did not offer a better fit compared to the model for the entire period ($p = 0.187$). The 120-day mortality rate appeared to follow a declining trend that spanned the entire study period, and the HFU did not appear to have caused a reduction in this measure.

3.5. 365-Day Mortality. For this outcome measure, post-HFU data are limited to a smaller sample ($n = 316$ instead of $n = 677$) due to fewer patients having been followed-up.

Conventional Tests: There was a small and non-significant decrease in 365-day mortality (21.46% pre-HFU to 20.57% post-HFU, Fisher's Exact test: $p = 0.769$)

Temporal Analysis: The single line model fits the entire time series best (Figure 6). Separate pre-HFU and post-HFU analysis demonstrated the best fit models were the line and plateau, respectively. Separate models did not offer a better fit when compared to the overall model ($p = 0.380$). This suggests that a change in one-year mortality was not due to the intervention. Instead, temporal analysis suggested that it followed a decreasing trend that spanned the entire study period. Other causal factors may be implicated in the background changes in 30-day, 120-day, and 365-day mortality rates (Discussion).

3.6. Patient Demographics. Patient demographics were analysed before and after implementation of the HFU (Table 2). There were two statistically significant differences: post-HFU, the rate of pathological fracture was less and the distribution of ASA grade was different. These are likely confounding variables which may have also influenced secular trends, further emphasising the importance of temporal analysis in retrospective studies.

4. Discussion

In this study, temporal analysis provided important insight into the observed changes before and after the intervention. Importantly, it demonstrated that some of the changes in the

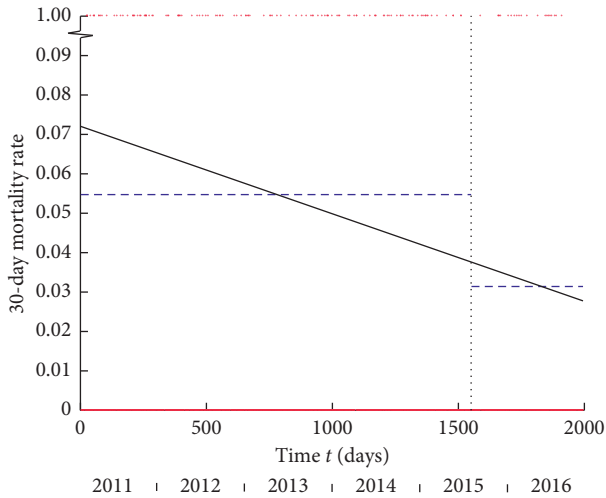


FIGURE 4: 30-day mortality versus time t in days from the start of the study. The vertical dashed line marks the onset of the HFU. The single line (black) is the best-fitting linear model for the entire period: equation $y = -0.00002t + 0.0720$. The dashed blue lines show plateaus are the best-fitting linear models for the pre-HFU ($y = 0.0547$) and post-HFU ($y = 0.0314$) periods. Data points 0 or 1 for dead or alive in red.

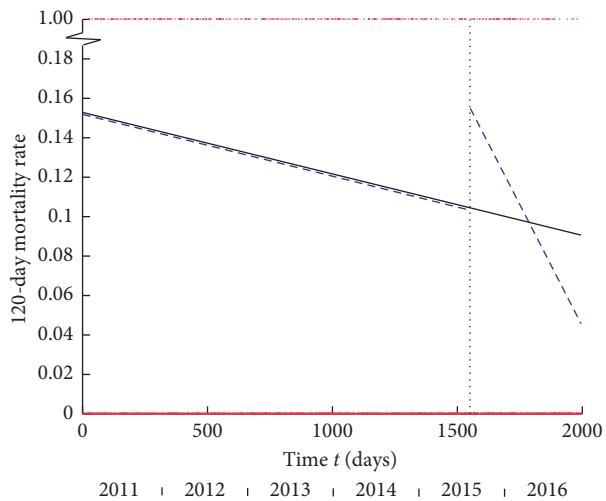


FIGURE 5: 120-day mortality versus time t in days from the start of the study. The vertical dashed line marks the onset of the HFU. The black line shows single line as the best-fitting linear model for the entire period: equation $y = -0.00003t + 0.1529$. The dashed blue lines are the best fitting linear models for the pre-HFU and post-HFU periods. The best model for the pre-HFU period is the single line with equation $y = -0.00003t + 0.1519$, and the best model for the post-HFU period is the single line with equation $y = -0.00025(t - 1550) + 0.1497$. Data points 0 or 1 for dead or alive in red.

outcome measures were likely due to trends that occurred over a longer time period, independent of the intervention.

In modelling the temporal effect of patient mortality, we used simple linear regression instead of logistic regression. We did this because the latter is better suited when dependence on a number of covariates is sought. As we were evaluating the time dependence of mortality rate, simple

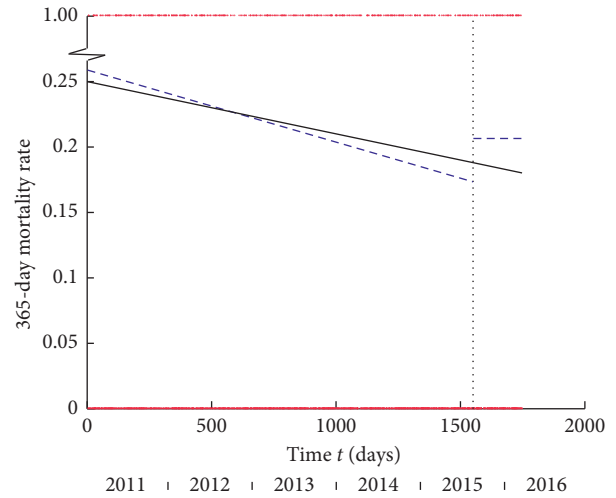


FIGURE 6: 365-day mortality versus time t in days from the start of the study. The vertical dashed line marks the onset of the HFU. The black line shows the single line as the best-fitting linear model for the entire period: equation $y = -0.00004t + 0.2497$. The dashed blue lines are the best-fitting linear models for the pre-HFU and post-HFU periods. The best model for the pre-HFU period is the single line with equation $y = -0.00006t + 0.2586$, and the best model for the post-HFU period is the plateau at $y = 0.2063$. Data points 0 or 1 for dead or alive in red.

TABLE 2: Patient demographics before and after implementation of the HFU.

Factor	Pre-HFU ($n = 2176$)	Post-HFU ($n = 671$)	p value
Age	84	85	0.0635
Gender	F:1575 M:601	F:477 M:194	0.5463
Pathological fracture	59	6	0.0091
ASA grade (1, 2, 3, 4, 5)	91, 574, 1178, 302, 3	12, 201, 378, 76, 1	0.0094
Reoperation	31	15	0.2035

Significant p values are in bold.

linear regression offered the advantage of yielding results in a directly interpretable form.

We retrospectively investigated factors that could explain the observed trends as this would help us in evaluating the reliability of the modelling method. We found several factors which are listed below:

- (1) *Tariff*. The Best Practice Tariff (BPT) was implemented nationwide in April 2010. One criterion was time to surgery within 36 hours (decreased from 48 hours previously) [15]. This could explain the falling trend in time to surgery observed in the early phase of the study in 2011 and 2012.
- (2) *Orthogeriatrician*. The first orthogeriatrician was appointed to the trust in 2006. Three more have since been appointed and are supported by junior staff.
- (3) *Implant use*. We discontinued the use of both the dynamic condylar screw (in 2009) and the Austin Moore hemiarthroplasty implant (in 2011). Modular

hemiarthroplasty was preferred in our trust after 2010 and was shown to decrease length of stay in a separate study [16].

- (4) *Clerking proforma*. A dedicated proximal femoral fracture clerking proforma was introduced in 2011 [17].
- (5) *Theatre prioritisation*. Proximal femoral fractures became prioritised cases on theatre lists in 2011. Prior to this, other cases often took precedence over fractured neck of femur surgery.
- (6) *Anaesthetic use*. Anaesthetic practice gradually improved, including introduction of a new anaesthetic proforma in 2012 (lower anaesthetic doses and increased use of regional anaesthesia).

The decline in time to surgery that occurred during the initial phase of the study period (long before the HFU) was responsible for the detected drop in median time to surgery from pre-HFU to post-HFU. Temporally, this coincided with the introduction of the Best Practice Tariff in 2010 and introduction of a neck of femur clerking proforma and theatre prioritisation in 2011, which may have contributed to the observed trend. Conventional statistical tests would have misattributed this to the HFU.

Temporal analysis demonstrated that the HFU significantly accelerated the rate of decrease in length of stay, yet this was missed in group comparisons using rank tests. The change in use of hip implants at our trust between 2009 and 2011 has been shown to decrease length of stay [16], possibly explaining the underlying improvement in this variable before the introduction of the HFU. Improvements in anaesthetic use may have also contributed to this.

The underlying decreasing trend in 30-day mortality represented by the single line model over the entire study period and the fact that separate models did not offer a significantly better fit meant that the decrease in 30-day mortality was probably due to a secular trend. Using group comparison tests alone would have misattributed this to the HFU. Temporal analysis of 120-day and 365-day mortality rates similarly demonstrated that an underlying decrease throughout the entire study period was the reason for any observed reduction when using group comparison tests. Interestingly, the post-HFU model for 120-day mortality showed a rapid decline possibly because of a small initial rise subsequently regressing to its normal trend.

The secular trends demonstrating improvement in mortality rate may have been a result of some or all of the aforementioned improvements, although other changes not identified here may have contributed as well.

4.1. Evaluation of the Proposed Method. The chosen set of models balance simplicity with the complexity that is required if a single straight line cannot capture underlying trends. The decision to enforce joined segments in models (iii) (line-plateau) and (iv) (line-line) accounted for the expectation that any change within the fitted period is expected to be gradual.

Discontinuity in the variable, such as when an intervention results in a sudden change, can be accommodated

by fitting separate models in the phases before and after the intervention. By using linear segments (as opposed to higher-order curves), our models yield meaningful parameters such as plateau values and rates of change. Furthermore, the models are nested which allows statistical comparison with a rigorous tabular method for selection of the best model. The best model offers the most reliable description of the temporal variation of the data.

Our method examined whether the intervention was significant by considering the periods before and after an intervention as separate but also as part of the whole study period. If separate models fit the data significantly better, we concluded that the intervention was significant in bringing change.

4.2. Comparison with Other Methods. The simplest form of temporal analysis is a visual inspection of the data of the time series, though this is of little practical value when change is small and obscured by data featuring high variability. Regression analysis is more reliable in revealing the presence of trends and has been used extensively in time series analysis [18].

An adaptation to linear regression is the interrupted time series (ITS) analysis where separate regressions are attempted for the periods before and after the intervention, allowing for discontinuities that could be due to the intervention [19]. This is a well-established method to test the hypothesis that an intervention causes a significant change in the outcome measure over time. However, fitting separate models before and after the intervention biases the method toward finding change at the intervention, and this may not offer the most effective description of how the time series evolved and what secular trends existed (or how these changed) throughout time. Most applications of the ITS method do not compare separate pre- and post-intervention lines with a single model for the entire time period, nor do they envisage using more than one linear segment for each section. This can miss more complex trends [7, 18]. For example, a change in the outcome measure that occurs well before or after the intervention, can be missed by ITS, which would credit the change to the intervention.

To contrast our proposed method with the ITS, we applied ITS analysis to one of our outcome variables: the time to surgery (Figure 7). The resulting intercept change and slope change of the ITS at the intervention marginally miss significance (change in slope ($p = 0.0532$) and change in intercept ($p = 0.0812$)). Nonetheless, as ITS focuses its attention at the intervention, the analysis would have deemed the HFU ineffective while missing the important reduction in time to surgery early in the study. F -tests demonstrated significantly improved goodness of fit when using our proposed method, compared to the ITS ($p < 0.0001$). This is not surprising given that our method enabled model change at different time points, yielding a lower sum of the square of residuals. In this scenario (time to surgery), our method concurred with the conclusion when using ITS analysis, but the two methods are fundamentally different, and their conclusions may not agree in all cases.

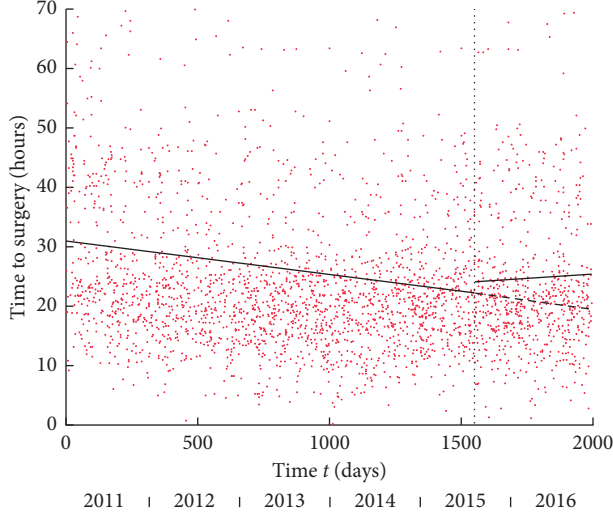


FIGURE 7: Interrupted time series analysis for time to surgery versus t days. The line of best fit pre-HFU (solid black line) was $y = -0.0056t + 30.97$. Extrapolation of the pre-HFU trend is shown as the dashed line. The line of best fit post-HFU (solid black line) was $y = 0.0029(t - 1550) + 19.59$. Slope change = $+0.0085$. Intercept change = $+1.862$. Change in slope: $p = 0.0532$; change in intercept: $p = 0.0812$.

Indeed, our method did not solely evaluate the effect of an intervention (as the ITS does) but described the overall trends within the study period, as well as providing a method to test the significance of change at the intervention.

4.3. Limitations. The first limitation of the proposed method is that it cannot capture change that is more complex than envisaged by two adjoining segments, (for example, multiple segmental trends or exponential trends) [18, 20]. In principle, the method can be extended to incorporate three or more segments, but in embracing higher complexity, one risks yielding unnecessary and meaningless information. Expanding to more complex models should be undertaken with caution.

Second, the method requires dedicated computer programming, as most statistical and spreadsheet software do not envisage the fitting of segmented linear regression with variable adjoining points between the segments. Consequently, a basic level of computer programming is required, which limits its use to more advanced healthcare analysts.

Another important limitation is in using F -tests to compare the goodness of fit of the various models when residuals do not meet the assumption of being normally distributed [11]. While this may have an effect on the reliability of p values that are obtained from the F -distribution, particularly when applying the method to mortality rates, it affects all models equally and should not decrease the effectiveness of the method in selecting the best model. Indeed, setting significance at 5% is imposing an arbitrary threshold which affects model selection, and different significance levels will undoubtedly yield different best models in marginal situations.

5. Conclusion

We proposed a novel, systematic method to model the temporal variation of a time series based on segmented linear regression. Its application to a real healthcare intervention demonstrated the method's ability to identify and describe trends over the study period, without bias to the intervention. We described a mathematically rigorous technique to determine whether trends are significant. The method offers a reliable tool in evaluating interventions and in detecting change, improving the information that can be drawn from retrospective studies.

Appendix

A. Derivation of the Least Squares Parameters for the Line-Plateau and Plateau-Line Models

The line-plateau model is a straight line $y_i = m_k t_i + c_k$ for $i = 1$ to k adjoining to a plateau $y_j = Y_k$ for $j = k + 1$ to n . m_k and c_k are the gradient and y -axis intercept of the straight line, respectively, and Y_k the value of the plateau. For each k from $k = 1$ to n , parameters m_k and c_k are found so as to minimise R , the sum of the squares of residuals:

$$R = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^k (y_i - m \cdot t_i - c)^2 + \sum_{i=k+1}^n (y_i - m \cdot t_k - c)^2, \quad (\text{A.1})$$

where y_i are the ordinates of the points of the data series (t_i, y_i) and \hat{y}_i are the predicted values of the model $y_i = m \cdot t_i + c$, $i = 1$ to k and $y_i = Y$, $i = k + 1$ to n .

Least squares fitting requires that $\partial R / \partial m = 0$ and $\partial R / \partial c = 0$ which together give the following matrix equation:

$$\begin{pmatrix} A & B \\ B & n \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} E \\ F \end{pmatrix}, \quad (\text{A.2})$$

where $A = \sum_{i=1}^k t_i^2 + (n - k)x_k^2$, $B = \sum_{i=1}^k t_i + (n - k)xt_k$, $E = \sum_{i=1}^k t_i y_i + xt_k \cdot \sum_{i=k+1}^n y_i$, and $F = \sum_{i=1}^n y_i$.

Equation (A.2) can be solved by linear algebra to yield values for m and c while Y (the plateau) is a constant value obtained from $Y = m \cdot t_k + c$.

The best line-plateau model is that which yields a minimum value of R for different k ($k = 1$ to n):

$[m_k, c_k, k] = \min [\sum_i (y_i - \hat{y}_i)^2]$; the derivation of the plateau-line model^k is very similar. Here a plateau $y_i = c$, $i = 1$ to k adjoins a straight line. $y_i = m \cdot (t_i - t_k) + c$, $i = k + 1$ to n .

Then, equation (A.2) becomes

$$R = \sum_{i=1}^k (y_i - c)^2 + \sum_{i=k+1}^n [y_i - m \cdot (t_i - t_k) - c]^2. \quad (\text{A.3})$$

Least squares fitting requires that $\partial R / \partial m = 0$ and $\partial R / \partial c = 0$ which yields the following matrix equation:

$$\begin{pmatrix} G & H \\ H & n \end{pmatrix} \begin{pmatrix} m \\ c \end{pmatrix} = \begin{pmatrix} L \\ F \end{pmatrix}, \quad (\text{A.4})$$

where $G = \sum_{i=k+1}^n t_i^2 - 2t_k \cdot \sum_{i=k+1}^n t_i + (n-k)t_k^2$, $H = \sum_{i=k+1}^n t_i - (n-k)t_k$, $L = \sum_{i=k+1}^n t_i y_i - t_k \cdot \sum_{i=k+1}^n y_i$, and $F = \sum_{i=1}^n y_i$.

Equation (A.4) can be solved by linear algebra to yield values for the plateau c and the slope of the adjoining straight line m that joins the plateau at the k th data point. Once all values of k have been tried, the best plateau-line model is the one which yields a minimum value of R for different values of k ($k=1$ to n):

$$[m_k, c_k, k] = \min_k \left[\sum_i (y_i - Y_k)^2 + \sum_j (y_j - \hat{y}_j)^2 \right]. \quad (\text{A.5})$$

B. Derivation of the Least Squares Parameters for the Line-Line Model

In the line-line model, a first straight line with equation $y_i = m_1 t_i + c_k$ ($i=1$ to k) adjoins a second straight line with equation $y_j = m_2 (t_j - t_k) + y_k$ ($j=k+1$ to n).

For each value of k , parameters m_1 , c , and m_2 are found that minimise R :

$$R = \sum_{i=1}^k (y_i - m_1 \cdot t_i - c)^2 + \sum_{i=k+1}^n (y_i - m_2 \cdot (t_i - t_k) - m_1 \cdot t_k - c)^2. \quad (\text{B.1})$$

Least squares fitting requires that $\partial R / \partial m_1 = 0$, $\partial R / \partial m_2 = 0$ and $\partial R / \partial c = 0$ which results in the following linear equation:

$$\begin{pmatrix} B & Q & n \\ S & T & Q \\ U & S & B \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ c \end{pmatrix} = \begin{pmatrix} F \\ Z \\ E \end{pmatrix}, \quad (\text{B.2})$$

where $B = \sum_{i=1}^k t_i + (n-k)t_k$, $Q = \sum_{i=k+1}^n t_i - (n-k)t_k$, $S = t_k \sum_{i=k+1}^n t_i - (n-k)t_k^2$, $T = \sum_{i=k+1}^n t_i^2 + (n-k)t_k^2 - 2t_k \sum_{i=k+1}^n t_i$, $U = \sum_{i=k+1}^n t_i^2 + (n-k)t_k^2$, $F = \sum_{i=1}^n y_i$, $Z = \sum_{i=1}^k t_i y_i - t_k \cdot \sum_{i=k+1}^n y_i$, and $E = \sum_{i=1}^k t_i y_i + t_k \cdot \sum_{i=k+1}^n y_i$.

Equation (B.2) can be solved using linear algebra to yield values for m_1 , m_2 , and c and the process is repeated for all k to find the set of values $[m_{1k}, c_k, k, m_{2k}]$ that yields the s minimum value of R for $k=1$ to n .

$$[m_{1k}, c_k, k, m_{2k}] = \min_k \left[\sum_{i=1}^k (y_i - \hat{y}_i)^2 \right]. \quad (\text{B.3})$$

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We gratefully acknowledge the support of the SPRINT charity who funded the publication cost of the article.

References

- [1] AHRQ, *National Healthcare Quality and Disparities Report and 5th Anniversary Update On the National Quality Strategy*, vol. 40, Rockville, MD, USA, 2015.
- [2] F. G. R. Fowkes, "Medical audit cycle," *Medical Education*, vol. 16, no. 4, pp. 227–238, 1982.
- [3] M. C. Portela, P. J. Pronovost, T. Woodcock, P. Carter, and M. Dixon-Woods, "How to study improvement interventions: a brief overview of possible study types: table 1," *BMJ Quality and Safety*, vol. 24, no. 5, pp. 325–336, 2015.
- [4] A. D. Harris, J. C. McGregor, E. N. Perencevich et al., "The use and interpretation of quasi-experimental studies in medical informatics," *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 16–23, 2006.
- [5] W. R. Shadish, T. D. Cook, and D. T. Campbell, "Experimental and quasi-experimental for generalized designs causal inference," *Handbook of Industrial and Organizational Psychology*, vol. 223, p. 623, 2002.
- [6] N. Khan, H. Abboudi, M. S. Khan, P. Dasgupta, and K. Ahmed, "Measuring the surgical "learning curve": methods, variables and competency," *BJU International*, vol. 113, no. 3, pp. 504–508, 2013.
- [7] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, 2016.
- [8] M. Taljaard, J. E. McKenzie, C. R. Ramsay, and J. M. Grimshaw, "The use of segmented regression in analysing interrupted time series studies: an example in pre-hospital ambulance care," *Implementation Science*, vol. 9, no. 1, p. 77, 2014.
- [9] E. M. Valsamis, R. Golubic, T. E. Glover, H. Husband, A. Hussain, and A. R. Jenabzadeh, "Modeling learning in surgical practice," *Journal of Surgical Education*, vol. 75, no. 1, pp. 78–87, 2018.
- [10] The Mathworks Inc, "MATLAB-MathWorks," 2016, <http://www.mathworks.com/products/matlab>.
- [11] T. Lumley, P. Diehr, S. Emerson, and L. Chen, "The importance of the normality assumption in large public health data sets," *Annual Review of Public Health*, vol. 23, no. 1, pp. 151–169, 2002.
- [12] F. Schoonjans, A. Zalata, C. E. Depuydt et al., "MedCalc: a new computer program for medical statistics," *Computer Methods and Programs in Biomedicine*, vol. 48, no. 3, pp. 257–262, 1995.
- [13] E. M. Valsamis, T. Chouari, C. O'Dowd-Booth, B. Rogers, and D. Ricketts, "Learning curves in surgery: variables, analysis and applications," *Postgraduate Medical Journal*, vol. 94, no. 1115, pp. 525–530, 2018.
- [14] R. G. Lomax and D. L. Hahs-Vaughn, *Statistical Concepts: A Second Course*, 4th edition, 2013.
- [15] "NHS Best Practice Tariff (BPT) for Fragility Hip Fracture Care User Guide," 2010, [http://www.nhfd.co.uk/20/hipfractureR.nsf/0/9b0c5ea2e986ff56802577af0046b1df/\\$FILE/Best_Practice_Tariff_User_Guide.pdf](http://www.nhfd.co.uk/20/hipfractureR.nsf/0/9b0c5ea2e986ff56802577af0046b1df/$FILE/Best_Practice_Tariff_User_Guide.pdf).

- [16] E. Lindisfarne, E. Dawe, and P. Stott, "Unipolar modular exeter hemiarthroplasty reduces length of stay and allows faster rehabilitation after hip fracture: a retrospective study of 117 patients," *European Surgical Research*, vol. 50, pp. 205-206, 2013.
- [17] E. A. O. Lindisfarne, E. J. C. Dawe, C. Willis, S. White, and P.M. Stott, "Improving the clinical assessment of patients with fractured neck of femur: audit and loop closure of clerking practice after introduction of a clerking proforma," *Injury Extra*, vol. 43, no. 10, p. 117, 2012.
- [18] A. Fretheim and O. Tomic, "Statistical process control and interrupted time series: a golden opportunity for impact evaluation in quality improvement," *BMJ Quality and Safety*, vol. 24, no. 12, pp. 748-752, 2015.
- [19] J. Lopez Bernal, S. Cummins, and A. Gasparrini, "Interrupted time series regression for the evaluation of public health interventions: a tutorial," *International Journal of Epidemiology*, vol. 46, no. 1, pp. 348-355, 2016.
- [20] D. Weiss, S. I. Dunn, A. E. Sprague et al., "Effect of a population-level performance dashboard intervention on maternal-newborn outcomes: an interrupted time series study," *BMJ Quality and Safety*, vol. 27, no. 6, pp. 425-436, 2017.