

Published in final edited form as:

J Clin Epidemiol. 2009 February ; 62(2): . doi:10.1016/j.jclinepi.2008.08.007.

Methods for estimating confidence intervals in interrupted time series analyses of health interventions

Fang Zhang, PhD, Anita K. Wagner, PharmD, MPH, DrPH, Stephen B. Soumerai, ScD, and Dennis Ross-Degnan, ScD

Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA

Abstract

Objective—Interrupted time series is a strong quasi-experimental research design that is increasingly applied to estimate the effects of health services and policy interventions. We describe and illustrate two methods for estimating confidence intervals around absolute and relative changes in outcomes calculated from segmented regression parameter estimates.

Study Design and Setting—We used multivariate delta and bootstrapping methods to construct confidence intervals around relative changes in level and trend and around absolute changes in outcome based on segmented linear regression analyses of time series data corrected for auto-correlated errors.

Results—Using previously published time series data, we estimated confidence intervals around the effect of prescription alerts for interacting medications with warfarin on the rate of prescriptions per 10,000 warfarin users per month. Both the multivariate delta method and the bootstrapping method produced similar results.

Conclusion—The bootstrapping method is preferred for calculating confidence intervals of relative changes in outcomes of time series studies since it does not require large sample sizes when parameter estimates are obtained correctly from the model. Caution is needed when sample size is small.

Keywords

Interrupted time series; Segmented regression; Confidence interval; Relative change; Bootstrapping; Delta method

Introduction

Increasingly, the interrupted time series (ITS) design is used to evaluate the effects of health care interventions. After its introduction to the health services research literature by Gillings [1], interrupted time series analysis has been widely used in assessing the effects of health services and policy interventions, such as implementation of changes in institutional

© 2008 Elsevier Inc. All rights reserved.

Corresponding author: Fang Zhang, PhD, Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care, 133 Brookline Avenue, 6th Floor, Boston, MA 02215 USA, Phone : 617-509-9962, Fax : 617-859-8112, fang_zhang@hms.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

formularies [2], payment restrictions for medications in state programs [3], or educational and administrative interventions to improve prescribing [4]. Interrupted time series designs, especially when they involve analysis of comparison series, are the strongest observational designs [1,5] to evaluate changes due to interventions because they can account for the pre-intervention level and trend of the outcome measures.

Time series models allow for description, explanation, prediction, and control [6]. Depending on data structures, alternative statistical modeling methods are available. The first method is generalized estimating equations (GEE) with a time series correlation structure [7], which uses individual level data to estimate aggregate intervention effects. This method can be problematic for several reasons: parameter estimates of intervention effects take the form of odds ratios, which are more difficult for a general audience to interpret; assumptions are violated when covariates are missing not at random, which is a frequent occurrence [8]; and computation is daunting when population size is large. Another method is the autoregressive integrated moving average model (ARIMA) [9,10], which uses aggregate outcome estimates at each time to model intervention effects. However, ARIMA models assume a complex error correlation structure and require sample sizes of at least 50 consecutive time points [9], which is often impossible in health services studies.

This paper focuses on segmented linear regression autoregressive error models [1,11], which are more commonly used in health services research since they can be applied to aggregate outcome estimates, tolerate fewer time points than ARIMA models, and are amenable to intuitive graphical presentation of the data and effects, which is appealing to a general audience.

Example: Warfarin Alert Study

In a health maintenance organization with 15 primary care clinics and 450,000 members, electronic medical record (EMR) alerts were implemented to reduce the co-prescribing of interacting medications to warfarin users to guard against potential adverse drug interactions [4]. At the beginning of the study period in January 2000, there were 35 months prior to the intervention (baseline), 4 for implementation of the alerts, and 17 after the intervention.

Total co-prescribing of interacting medications was the primary outcome measured; the authors also measured co-prescribing of several individual medications, including acetaminophen. The effects of the alerts on these outcomes were demonstrated by reporting level and trend changes in rates of co-prescribing, controlling for pre-intervention level and trend (see figure 1). For example, there was a significant level change of -311.4 (95% CI: -480.3, -142.4) prescriptions of interacting medications per 10,000 warfarin users, which corresponds to a -9.5% relative change (95% CI: -15.6%, -3.6%) after the alert intervention. The absolute change describes a clinically meaningful absolute reduction in co-prescribing; the relative change represents this change as a proportion of the baseline rate, permitting comparison with the effects of similar interventions regardless of the background prescribing rate.

Since p-values fail to convey the precision [12] of effect estimates, it is recommended and sometimes required [13] to report confidence intervals (CI) for both absolute and relative change estimates from time series analyses. However, in 433 articles referenced in PubMed with the term “time series” in the title from 2005 to 2007, only an early version of the authors' multivariate delta methods [14] reported confidence intervals for both absolute and relative changes. Using two accepted statistical methods, in this paper we developed practical algorithms to calculate confidence intervals for time series estimates from segmented regression models.

Methods

Time Series Analysis Models

The linear segmented regression autoregressive error model is appropriate when the data suggest a linear relationship of time and the outcome under study, and when errors follow an autoregressive pattern. These models can adjust both parameter estimates and variance for autocorrelation in the error terms, which is indicated when sequential error terms are either positively or negatively correlated at a level greater than chance. The segmented linear autoregressive error time series model for the warfarin co-prescribing study example allows us to estimate the level (Coef__intervention) and trend (Coef__monthafterintervention) changes attributable to the EMR alerts as follows with autocorrelation error terms [15]:

$$\begin{aligned} Outcome(t) = & \text{Intercept} + \text{Coef_month} * month(t) \\ & + \text{Coef_intervention} * intervention(t) \\ & + \text{Coef_monthafterintervention} * month_after_intervention(t) + error(t) \end{aligned} \quad (1)$$

Here, *outcome*(t) is the rate of co-prescriptions per 10,000 warfarin users in month t; *intervention* is the indicator for whether month t occurs before (intervention=0) or after (intervention=1) the implementation of alerts; *month* and *month after intervention* are integer variables naturally ordered by month.

All models to analyze time series can produce maximum likelihood estimators (MLEs) of the parameters if they exist. Since the maximum likelihood method is considered one of the most appropriate approaches to use for small samples with autocorrelated errors [16], we present confidence interval estimates based on parameters estimated using maximum likelihood methods.

When the modeling process is complete, results from the model can be presented in the form of parameter estimates (and their confidence intervals) that express absolute and relative level and trend changes after the interventions compared to before, or in the form of linear combinations of the parameter estimates that express absolute and relative level and trend changes at a given point in time after the intervention.

The regression parameters from the model can be used to derive estimates for baseline level Intercept, baseline trend Coef__month, post-intervention level (Intercept + Coef__intervention), and post-intervention trend (Coef__month + Coef__monthafterintervention); The estimates of the absolute changes in level and trend after the intervention are EST(Coef__intervention) and EST (Coef__monthafterintervention) , respectively, while the relative changes in these parameters can be expressed as EST(Coef__intervention)/EST(Intercept) and EST(Coef__month)/EST(Coef__monthafterintervention).

Easier to interpret, however, is the overall effect of an intervention, combining level and trend changes into a single estimate. Denote EST(outcome_with) and EST(outcome_without) as estimates of the outcome with and without the intervention, respectively. Absolute change is the difference between EST(outcome_with) and EST(outcome_without) where relative change is {EST(outcome_with) – EST(outcome_without)}/EST(outcome_without). The outcomes EST(outcome_with) and EST(outcome_without) can be calculated as linear combinations of the parameter estimates from the segmented time series regression model. For example, if we want to consider the intervention effect x months after the policy, the absolute change would usually be represented as {EST(Coef__intervention) + x* EST(Coef__monthafterintervention)}. Relative

change due to the intervention would then equal the absolute change over $\{EST(Intercept) + (x + \text{number_of_baselinemonths}) * EST(Coef_month)\}$. Other more complicated functions (e.g., exponential) are also possible and needed when outcome variables have been transformed in the analysis (e.g., converted to logarithms). The method for calculating confidence intervals for each of these estimates of change follows below.

Confidence Intervals for Absolute and Relative Policy Effects

To estimate the 95% confidence interval around an absolute policy effect $\{EST(outcome_with) - EST(outcome_without)\}$, we can use $\{EST(outcome_with) - EST(outcome_without)\} \pm 1.96 * \text{Standard Error of } \{EST(outcome_with) - EST(outcome_without)\}$, where $\{EST(outcome_with) - EST(outcome_without)\}$ can be the absolute level change, the absolute trend change, or the absolute overall policy impact at a specific point in time.

The construction of confidence intervals for relative change in this context requires statistical approximation, for which we propose two methods. The multivariate delta method [17] is very useful when exact calculation of confidence intervals is numerically impossible. It offers a solution based on Taylor's approximation for the variance of the relative effect using existing parameter estimates and the variance/covariance matrix from standard statistical software such as SAS [15] or R [18]. The generic formulas for the expectation and variance of the relative change produced by this method are in the appendix.

The delta method works well when the sample size is adequate to ensure reasonable normal approximation of the MLEs, a requirement that may not always be met. We propose to use the bootstrapping method [19] to construct confidence interval to compare with those produced by the delta method. Three steps are required in the bootstrapping process: 1) derive all estimates from the regression model, including mean parameter estimates, variances and autocorrelation estimates; 2) simulate data based on these estimates with normally distributed error a large number of times (in this case $B=10,000$); 3) obtain the bootstrap estimates for the parameters of interest from each simulation and use the 2.5 and 97.5 percentiles of these estimates to construct the 95% confidence interval. Bootstrapping methods can be applied to construct confidence intervals around both absolute and relative changes.

Since the delta method is a “first-order” approximation based on Taylor series expansion and the bootstrap can often have “second-order” accuracy [17], the bounds of the bootstrap confidence interval are usually closer to the true values while the delta method may sometimes produce confidence intervals that are too narrow [17]. However, when results are similar, the computation time of the Delta method is much shorter. We provide the SAS® code used to estimate confidence intervals around relative change estimates for our example using both methods at <http://www.dacp.org/zhangwebreferences.html>.

Application

We use the warfarin alerts study [4] to compare results from the multivariate delta method (MDM) and bootstrapping method (BM). Following standard practice, we first constructed a saturated model containing parameters for level and trend at baseline and in post-intervention segments, as detailed in equation 1 above. We then constructed the most parsimonious model by sequentially eliminating the least significant terms and re-estimating the model. At each iteration, we allowed correction for an autoregressive error structure with up to 12th order autocorrelations, using a back-step automatic autocorrelation selection procedure. The parsimonious model included the intercept (i.e., baseline level), level changes after alerts, and a trend change after alerts for the total co-prescription rate. Table 1

lists the alerts effect estimates and confidence intervals computed using the two methods, expressed in the form of absolute and relative level and trend changes in the co-prescription rate of all potential interacting drugs per 10,000 warfarin users per month comparing the periods before and after the alerts.

Compared to the baseline period, the alerts were associated with a level change of -311.4 (95%CI: -480.3,-142.4) per 10,000 warfarin users per month, which corresponds to a -9.5% relative change (95%CI: -14.7%,-4.4% using the MDM and -15.6%,-3.6% using the BM). There were no statistically significant trend changes comparing the period of the alerts to the baseline. A zero baseline trend, as estimated in this model, would result in infinitely large relative trend changes, so the post alert relative trend changes are not presented. As expected, we observed slightly wider confidence intervals of the BM compared to those produced by the MDM.

Table 1 also lists overall intervention effects as absolute and relative changes of the total co-prescription rate at a specific time (6 months) after the alerts were implemented, compared to the expected if no alert had been implemented. Six months after the alerts, we estimated a change of -439.1 (95%CI: -543.6,-334.6) per 10,000 warfarin users per month, which corresponds to a -13.5% (95%CI: -16.6%,-10.4% [MDM] and -17.8%,-8.6% [BM]) relative change compared to what would have been expected in the absence of the alerts.

The effects of alerts on co-prescribing of acetaminophen were more significant with an additional significant slope change of -22.2 (95%CI: -38.1, -6.3) per 10,000 warfarin users per month, which corresponds to a large relative change of -267.5% (95% CI: -475.7%, -59.3% [MDM] and -650.3%, -80% [BM]).

Discussion

Interrupted time series studies are a strong quasi-experimental method for evaluating intervention effects when randomized controlled trials are not feasible. Absolute and relative intervention effects, which compare the overall changes in outcome attributable to an intervention with counterfactual estimates of what would have happened without the intervention, are the most intuitive and informative summary of the results of autoregressive error time series models.

We demonstrate an application of multivariate delta and bootstrapping methods for estimating confidence intervals around such changes in level and trend, and around overall changes in study outcome at a given point in time. These methods enable researchers to estimate confidence intervals when exact solutions are not available. Using the algorithms provided, investigators can estimate confidence intervals in a similar way for other time series applications. Complete parameter estimates from the linear segmented autoregressive error time series models are required to implement the bootstrapping method and the additional variance-covariance matrix of the main parameters is required to implement the multivariate delta method.

Several limitations and precautions should be noted. First, certain functions have boundaries, and the 95% CI should reflect those boundaries. For example, a decrease of a positive outcome such as number of prescriptions should not be represented as more than a 100% relative change, since this is the maximum effect size.

Second, the resulting confidence intervals should not be interpreted as exact confidence intervals when the number of time points in a time series segment is small. The delta method is based on large sample size theory and small sample sizes are likely to result in confidence intervals that are too narrow, exaggerating precision [17]. The bootstrapping method does

not suffer from the small sample size problem if the correct model estimates are used for all relevant parameters because bootstrapping repeated samples achieves unbiased confidence interval estimates [19]. However, problems can occur when estimating variance and autocorrelation parameters for time series with small sample sizes using typical estimation methods such as maximum likelihood estimation when these parameters are entered into the bootstrapping trial. Park et al. demonstrated underestimation of confidence intervals for small sample sizes which could lead to rejection of a correct null hypothesis up to 20% of the time in simulated examples with very large autocorrelations (0.9) and about 9% of the time in examples with moderate autocorrelations (0.4) for a time series with 50 time points [20], compared with the desired 5% level. We conducted simulations using estimates from previously published health services research data sets with sample sizes varying from 20 to 50 time points, assuming a single policy level change only with autocorrelations varying from 0.1 to 0.8, and found a bias in estimating confidence intervals around the level change that varied from 6% to 23%. Because bias depends on multiple factors, for time series with a small number of time points and large autocorrelation, we can conclude that even the bootstrapping method may produce artificially narrow confidence intervals. Thus, reporting a 95% confidence interval using the 99% critical value may be desirable practice [21].

Despite the continuing need for exact statistics for estimating confidence intervals in segmented linear autoregressive error time series analysis [22], the approximate confidence interval estimation methods described in this paper are useful in a wide range of evaluations of health interventions. The bootstrapping method is preferred over the delta method for calculating confidence intervals since it does not require large sample sizes and it produces conservative confidence intervals. Given the reality of small sample sizes in many studies, further research on sample size and power calculations is needed to clarify the number of data points required for reliable estimation of intervention effects.

Acknowledgments

This research was supported by grant R01 AG19808-01A1 from the National Institute on Aging (Dr Soumerai, principal investigator) and grant R01DA10371-01 from the National Institute on Drug Abuse (Dr Soumerai, principal investigator). Drs. Zhang, Ross-Degnan, Soumerai and Wagner are investigators in the HMO Research Network Center for Education and Research in Therapeutics, supported by the U.S. Agency for Healthcare Research and Quality (Grant # 2U18HS010391).

Additional contributions: We are grateful for the generosity of Adrienne C. Feldstein, MD, MS; David H. Smith, RPh, MHA, PhD and Xiuhai Yang, MS of the Center for Health Research, Kaiser Permanente, Portland, Ore for providing the data.

Appendix

Denote \hat{y}_{with} and $\hat{y}_{without}$ as estimates of the outcome with and without the intervention, respectively. The generic formulas for the expectation and variance of the relative change using the delta method:

$$\begin{aligned}
E[(\hat{Y}_{with} - \hat{Y}_{without})/\hat{Y}_{without}] &\approx \frac{E(\hat{Y}_{with})}{E(\hat{Y}_{without})} - 1 \text{VAR}[(\hat{Y}_{with} \\
&- \hat{Y}_{without})/\hat{Y}_{without}] = \text{VAR}(\hat{Y}_{with}/\hat{Y}_{without}) \approx \left(\frac{E(\hat{Y}_{with})}{E(\hat{Y}_{without})} \right)^2 \\
&\times \left(\frac{\text{VAR}(\hat{Y}_{with})}{E^2(\hat{Y}_{with})} \right. \\
&+ \frac{\text{VAR}(\hat{Y}_{without})}{E^2(\hat{Y}_{without})} \\
&\left. - 2 \frac{\text{COV}(\hat{Y}_{with}, \hat{Y}_{without})}{E(\hat{Y}_{with})E(\hat{Y}_{without})} \right).
\end{aligned}$$

Reference List

1. Gillings D, Makuc D, Siegel E. Analysis of Interrupted Time Series Mortality Trends: an Example to Evaluate Regionalized Perinatal Care. *American Journal of Public Health*. 1981; 71(1):38–46. [PubMed: 7258429]
2. Brufsky JW, Ross-Degnan D, Calabrese D, Gao X, Soumerai SB. Shifting Physician Prescribing to a Preferred Histamine-2-Receptor Antagonist. Effects of a Multifactorial Intervention in a Mixed-Model Health Maintenance Organization. *Medical Care*. 1998; 36(3):321–32. [PubMed: 9520957]
3. Soumerai SB, Avorn J, Ross-Degnan D, Gortmaker S. Payment Restrictions for Prescription Drugs Under Medicaid. Effects on Therapy, Cost, and Equity. *New England Journal of Medicine*. 1987; 317(9):550–6. [PubMed: 3302713]
4. Feldstein AC, Smith DH, Perrin N, Yang X, Simon SR, Krall M, Sittig DF, Ditmer D, Platt R, Soumerai SB. Reducing Warfarin Medication Interactions: an Interrupted Time Series Evaluation. *Arch Intern Med*. May 8; 2006 166(9):1009–15. [PubMed: 16682575]
5. Cook, TD.; Campbell, DT. *Quasi-experimentation: design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company; 1979.
6. Zeger SL, Irizarry R, Peng RD. On Time Series Analysis of Public Health and Biomedical Data. *Annu Rev Public Health*. 2006; 27:57–79. [PubMed: 16533109]
7. Zeger SL. A Regression-Model for Time-Series of Counts. *Biometrika*. 1988; 75(4):621–9.
8. Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, S. *Analysis of longitudinal data*. Second. New York NY: Oxford University Press; 2002.
9. Box, GEP.; Jenkins, GM. *Time Series Analysis: Forecasting and Control*. San Francisco CA: Holden-Day; 1976.
10. Perez A, Dennis RJ, Rodriguez B, Castro AY, Delgado V, Lozano JM, Castro MC. An Interrupted Time Series Analysis of Parenteral Antibiotic Use in Colombia. *J Clin Epidemiol*. 2003; 56(10): 1013–20. [PubMed: 14568634]
11. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research. *Journal of Clinical Pharmacy and Therapeutics*. 2002; 27(4):299–309. [PubMed: 12174032]
12. Atkins CD. A Clinician's View of Statistics. *Journal of General Internal Medicine*. 1997; 12(8): 500–4.
13. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals. *JAMA*. 1997; 277:927–34. [PubMed: 9062335]
14. Zhang, F.; Wagner, AK.; Soumerai, SB.; Ross-Degnan, D. Estimating confidence intervals around relative changes in outcomes in segmented regression analysis of time series data; Paper presented at: NESUG 15th Annual Conference; Buffalo, NY. 2002;
15. SAS Institute Inc. *SAS OnlineDoc®*, Version 9.1.3. Cary NC: 2005.

16. Spitzer JJ. Small-Sample Properties of Non-Linear Least-Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors. *Journal of the American Statistical Association*. 1979; 74(365):41–7.
17. Casella, G.; Berger, RL. *Statistical Inference*. Second. Pacific Grove, CA: Duxbury; 2002.
18. R: A Language and Environment for Statistical Computing. Development Core Team; Vienna, Austria: 2008.
19. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Boca Raton FL: Chapman & Hall/CRC; 1993.
20. Park RE, Mitchell BM. Estimating the Autocorrelated Error Model With Trended Data. *Journal of Econometrics*. 1980; 13:185–201.
21. Zinkgraf, SA.; Willson, VL. The use of the Box-Jenkins approach in causal modeling: an investigation of the cost of the misidentification of selected stationary models. In: Anderson, OD.; Perryman, MR., editors. *Time Series Analysis*. New York NY: New Holland Publishing Company; 1981.
22. Agresti, A. *Categorical Data Analysis*. Hoboken NJ: John Wiley & Sons; 2002.

What is new?

- In 433 articles referenced in PubMed with the term "time series" in the title from 2005 to 2007, only one reported confidence intervals for both absolute and relative changes
- The study describes two methods for obtaining confidence intervals for absolute and relative changes in parameter estimates from studies using segmented regression interrupted time series designs.
- Multivariate delta and bootstrapping methods both enable researchers to estimate confidence intervals around absolute and relative changes.
- The bootstrapping method is preferred for calculating confidence intervals of relative changes when parameter estimates are obtained

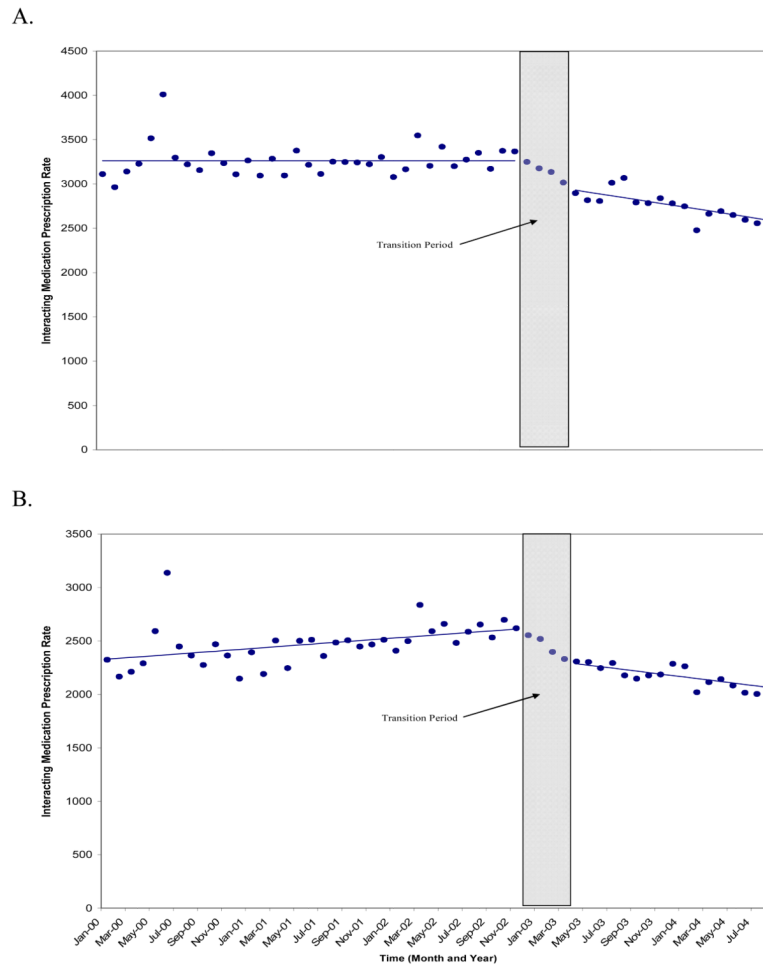


Figure 1.

Time series of warfarin-interacting medication prescription rates by month (warfarin-interacting medication prescriptions per 10,000 warfarin users per month). A, Interacting medications (combined); B, warfarin-acetaminophen. Fitted trend lines show predicted values from the segmented regression analysis. Transition period is the intervention implementation period.

Table 1
Absolute and relative changes in level and trend of the monthly co-prescription rate (per 10,000 warfarin users) following alert changes, and associated 95% confidence intervals around these changes estimated with the multivariate delta and bootstrapping methods

	Absolute Change [Exact]	Relative Change [Multivariate Delta Method]	Relative Change [Bootstrapping Method]
Total co-prescription			
EMR alerts level to baseline level	-311.4 (-480.3,-142.4)	-9.5% (-14.7%,-4.4%)	-9.5% (-15.6%,-3.6%)
Six months after the EMR alerts	-439.1 (-543.6, -334.6)	-13.5% (-16.6%, -10.4%)	-13.5% (-17.8%, -8.6%)
Acetaminophen co-prescription			
EMR alerts level to baseline level	-311.0 (-495.0,-127.0)	-13.4% (-21.6%,-5.3%)	-13.4% (-21.7%,-5.2%)
EMR alerts trend to baseline trend	-22.2 (-38.1,-6.3)	-267.5% (-475.7%,-59.3%)	-266.9% (-650.3%, -80.0%)
Six months after the EMR alerts	-444.2 (-598.4, -290.0)	-16.7% (-21.8%, -11.5%)	-16.7% (-21.7%, -11.3%)