

Adaptive Regression for Nonlinear Interrupted Time Series Analyses with Application to Birth Defects in Children of Vietnam War Veterans

George J. Knafl

School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Email: gknafl@unc.edu

How to cite this paper: Knafl, G.J. (2022) Adaptive Regression for Nonlinear Interrupted Time Series Analyses with Application to Birth Defects in Children of Vietnam War Veterans. *Open Journal of Statistics*, 12, 789-809.

<https://doi.org/10.4236/ojs.2022.126045>

Received: November 17, 2022

Accepted: December 27, 2022

Published: December 30, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The purpose of this article is to provide an overview of adaptive regression modeling and demonstrate its use in conducting nonlinear analyses of interrupted time series (ITS) data. Adaptive regression modeling is based on heuristic search over alternative models for data controlled by likelihood-cross validation (LCV) scores with larger scores indicating better models. Extended linear mixed models are used for correlated data like ITS data. Power transforms of predictor variables are used to account for nonlinearity. The use of adaptive regression modeling for assessing ITS effects is demonstrated using data on annual proportions of major birth defects in children fathered by male Air Force veterans of the Vietnam War over a 59-year period. The interruption for this ITS is conception after versus before the start of a participant's first tour in the Vietnam War. Whether the ITS effect is related to dioxin exposure is also addressed. Dioxin is a highly toxic contaminant of the herbicide Agent Orange used in the Vietnam War. The core findings of the reported analyses are that a substantial adverse ITS interruption effect is identified and that this adverse effect can reasonably be attributed to participants having a high dioxin exposure level. Moreover, these results indicate that adaptive regression modeling can identify nonlinear ITS effects in general situations that can lead to consequential insights into nonlinear relationships over time, possibly varying with other available predictors.

Keywords

Adaptive Regression, Air Force Health Study, Birth Defects, Dioxin, Interrupted Time Series

1. Introduction

The interrupted time series (ITS) design is commonly used in public health re-

search [1] [2] as well as in a variety of other application areas. This design is used to assess whether an interruption affects the underlying pattern for a time series. For example, the data used in reported analyses consist of annual proportions of major birth defects in children of male Air Force veterans of the Vietnam War. The interruption corresponds to the start of the first Vietnam tour for each veteran. The research question is whether conception after the start of the first tour affects the time-varying probability of a child having a major birth defect and whether this effect can be related to herbicide exposure for the father.

Standard straight-line models are typically used to represent the effect of the interruption. When the outcome is continuous and reasonably treated as normally distributed, the mean of this outcome is assumed to change linearly over time with possibly different intercepts and/or slopes before versus after the interruption. For categorical outcomes, generalized linear models are used with appropriately transformed means (using what is called the link function) modeled in the same way as linear in time. An intercept change, called a level change, corresponds to a shift up or down after the interruption in the linear pattern for the possibly link function transformed means. A slope change corresponds to a rotation of that linear pattern after the interruption. However, the assumption of a linear pattern whether shifted/rotated or not may not be appropriate.

ITS outcome measurements are often treated as independent. For example, Turner et al. [1] found that 31% of the studies they reviewed used linear regression methods to analyze their ITS data. However, it is more appropriate to account for autocorrelation, that is, correlation between outcome measurements at different times. When the outcome is continuous and reasonably treated as normally distributed, linear mixed modeling is one possibility to use to account for autocorrelation. For categorical outcomes, generalized estimation equations (GEE) can be used to account for autocorrelation. See Tab. 6 of [1] for a detailed list of methods used in the public health literature for ITS modeling.

When the outcome is continuous and reasonably treated as normally distributed, it is typically treated as having constant variances over time. For categorical outcomes, variances are treated as an appropriate function of the means possibly combined with constant dispersions over time. For example, count outcomes treated as Poisson distributed have variances equal to the means. These variances are then extended by multiplying the means by a constant dispersion parameter. However, nonconstant variances/dispersions can provide for distinct improvements. Most statistical software tools do not support modeling of nonconstant variances/dispersions.

Methods are needed for modeling nonlinear patterns for means over time with a possible discontinuity in that pattern due to an ITS interruption (a generalized level change) as well as a possible change in the nonlinear form of that pattern due to the interruption (generalizing a rotation). Methods are also needed to allow for nonconstant variances/dispersions. These methods need to account as well for autocorrelation.

The reported analyses use adaptive regression methods for these purposes [3] [4]. Briefly, adaptive regression methods are based on a heuristic search through alternative regression models for possibly correlated outcomes based on power transforms of predictor variables to identify an effective choice for a given data set. Models are evaluated using likelihood cross-validation (LCV) scores with larger values indicating better models. The goal of the search is to identify a model that is based on a parsimonious set of power transforms with as large as possible (or suboptimal) LCV score.

2. Methods

2.1. The Air Force Health Study

The Air Force Health Study (AFHS) was a longitudinal, prospective epidemiological study that investigated the impact of herbicide exposure on health, survival, and reproductive outcomes for male Air Force veterans of the Vietnam War [5] [6] [7] [8]. It is also called the Ranch Hand Study because study participants included personnel responsible for herbicide handling and spraying as part of what the Air Force called Operation Ranch Hand. The study started in 1979. Data were collected from 1982-2006 when the study ended. Available reproductive outcome data were released to the Vietnam Veterans of America in 2000. These consist of 8437 pregnancies with 6923 resulting in live births for 2614 participants, 94.8% of the 2758 participants of the full study. Dioxin levels in parts per trillion (ppt) are available for 2038 participants, 78.0% of the participants with available data and 73.9% of all study participants. Observed detectable dioxin levels range from 1.25 to 314.2 ppt. Dioxin is a highly toxic contaminant contained in the herbicide Agent Orange used in the Vietnam War. Reported analyses use reproductive outcome data for only the participants with measured dioxin levels in order to assess the impact of dioxin exposure.

Live births for AFHS participants were conceived over a 59-year period including the period of the Vietnam War (11/55-4/75) as well as years before and after the war. Birth defects for these live births were classified as major, minor, or unspecified. A major birth defect was defined as one that “could potentially affect survival, require substantial medical care, result in marked physical or psychological handicaps, or interfere with a child's prospects for a productive and fulfilling life” [5]. Knaf1 [9] reported on analyses of the AFHS birth defect data. One of these analyses used adaptive regression methods to identify a cutoff for a low versus high level of dioxin exposure for predicting the occurrence of major birth defects in children conceived after the start of the tour and fathered by participants with available dioxin levels. This analysis identified the cutoff 6.3 ppt with a significant effect (OR = 1.97, 95% CI 1.36 - 2.85, $p < 0.001$) on a child having a major birth defect for a high compared to a low dioxin exposure level for the father. This result is used in reported analyses to categorize dioxin values into low and high levels.

For reported analyses, live births are categorized into conceived before versus

after the start of the first Vietnam War tour in order to assess an interruption effect. Dioxin levels are categorized as indicated above into low levels ≤ 6.3 ppt (including values below the detectable limit) versus high levels > 6.3 ppt. Counts of numbers of live births and of live births with major birth defects are aggregated over each of the 59 years of available data within the four categories based on conceived before/after the start of the tour and low/high dioxin levels. Outcome measurements are missing for years with no observed live births within each of these categories. There is a total of 129 observations including 33 annual measurements over years 1 - 38 for conceived before the start of the tour with low dioxin, 32 annual measurements over years 4 - 38 for conceived before the start of the tour with high dioxin, 33 annual measurements over years 26 - 59 for conceived after the start of the tour with low dioxin, and 31 annual measurements over years 28 - 59 for conceived after the start of the tour with high dioxin. Note that, in the intermediate period of years 26 - 38, measurements are available both before and after the start of the tour.

The outcome measurements for reported analyses are the 129 observed counts of the numbers of live births with a major birth defect. An offset variable based on the natural logs of the 129 counts of the numbers of live births is used to adjust the model for the mean annual counts into the model for the annual probabilities for a major birth defect within the four categories of before/after the start of the tour and low/high dioxin level.

2.2. Poisson Regression Model

Knaf1 and Meghani [10] provide a detailed description of the Poisson regression model used in reported analyses. While [10] addresses the special case of a single time series, the model readily generalizes to multiple time series including the composite four time series addressed in reported analyses (a detailed generalization covering multiple time series is provided in [4]). Two modeling approaches are considered in [10] for accounting for temporal correlation: extended GEE modeling and extended linear mixed modeling (ELMM). ELMM is demonstrated in [10] to be preferable, and so only that approach is used in reported analyses and is described next.

The means for count outcome measurements are determined by a generalized linear model, that is, the natural logs of the means are modeled using a general linear model based on power transforms of predictor variables. The variances are products of the means (as appropriate for Poisson regression) and the dispersions. The natural logs of the dispersions are modeled using a general linear model based on power transforms of predictor variables. Both the means and the dispersions are adjusted using an offset variable to convert the model for the observed counts into a model for the associated observed proportions. Standard deviations equal square roots of products of means and dispersions.

Outcome measurements are grouped into matched sets (usually called “subjects” in the statistical literature). Outcome measurements are treated as corre-

lated within each matched set and as independent between different matched sets. For reported analyses, there are four matched sets consisting of annual measurements within the categories based on before/after the start of the tour and low/high dioxin levels. Outcome measurements within each matched set are ordered by year index with possible values 1 - 59.

Formally, let y_{ct} denote counts of birth defects within category c at year index t with associated total counts T_{ct} of live births and means $\mu_{ct} = E y_{ct}$. The natural log link function is used to transform the associated probabilities μ_{ct}/T_{ct} satisfying

$$\ln(\mu_{ct}/T_{ct}) = \ln \mu_{ct} - \ln T_{ct}$$

where the offsets are $\ln T_{ct}$. The natural logs $\ln \mu_{ct}$ of the means satisfy a general linear model based on multiple power transforms x^p of predictors x using powers p . The parameters correspond to slopes for these power transforms (including a possible intercept). The natural logs of the dispersions φ_{ct} also satisfy a general linear model in multiple power transforms of possibly different predictors with parameters corresponding to associated slopes. The counts y_{ct} are treated as Poisson distributed so that their variances satisfy $\text{VAR}(y_{ct}) = \mu_{ct}$. Extended variances σ_{ct}^2 are generated by multiplying by the dispersions giving the standard deviations

$$\sigma_{ct} = (\varphi_{ct} \cdot \mu_{ct})^{1/2}.$$

Correlations for the outcome measurements y_{ct} are based on spatial autoregression of order 1. Specifically, the correlation between measurements in the same category c with year indexes t and s is $\rho^{|t-s|}$ where $|t-s|$ denotes the distance between the year indexes t and s and ρ denotes the autocorrelation. Implementations of GEE methods typically support only non-spatial correlations assuming that indexes within each matched set are consecutive integers without any gaps. This is an unrealistic assumption for general matched sets of year indexes, and especially so for arbitrary subsets of those matched sets. Modeling of such subsets is needed to compute LCV scores as defined next.

2.3. Likelihood Cross-Validation

The likelihood is the multivariate normal density evaluated at the means and variances for the count outcome measurements. Parameters are estimated using maximum likelihood estimation. When the outcome is continuous, this is usual linear mixed modeling. For categorical outcomes, the likelihood is only an extended likelihood and not a true likelihood since it does not integrate to 1. However, it is used as a likelihood to generate parameter estimates, and so this distinction is ignored in what follows for brevity.

First, outcome measurements are randomly partitioned into k distinct subsets, called folds. Then, the outcome measurements in each fold are scored using the likelihood for the fold observations computed with parameters estimated using the remaining observations. Finally, the fold likelihood values are combined into

a composite k -fold likelihood cross-validation (LCV) score. The same initial seed is used in the randomization for all models of the same data set so that their LCV scores are comparable. A detailed formulation is provided in [10].

A larger LCV score indicates a better model. However, the addition of a power transform can sometimes result in a larger LCV score while only providing a superfluous or negligible improvement to the model. A small percent decrease in the LCV score for the simpler model without the added transform compared to the more complex model with the added transform is an indication that the added transform provides only a superfluous improvement. An effective model is one that has a suboptimal LCV score without any superfluous predictors. Such models are identifiable using adaptive regression modeling.

What constitutes a “large” versus “small” percent decrease can be formalized using LCV ratio tests based on the χ^2 (DF) distribution with DF degrees of freedom (as for standard likelihood ratio tests). The cutoff for a substantial (distinct, significant) percent decrease is computed from the 95th percentile of the χ^2 (DF) distribution (the significance level can be changed, but 5% is the usual choice as for p values). A percent decrease that is larger than the cutoff indicates that the model with the larger LCV score provides a substantial (distinct, significant) improvement over the model with the smaller LCV score or, equivalently, that the simpler model is substantially inferior. Otherwise, the model with the smaller LCV score is a competitive alternative to the model with the larger LCV score. If the model with the smaller LCV score is less complex than the model with the larger LCV score, then it is a parsimonious, competitive alternative. The value of DF can be set using the difference in the number of parameters for the two models, but it is usually conservatively computed using DF = 1. A brief overview of adaptive regression modeling is provided next.

2.4. Adaptive Regression

Given a set of untransformed predictors for the means and a possibly different set of untransformed predictors for the dispersions, adaptive regression methods are used to identify an effective model based on a parsimonious set of power transforms of the predictors from both sets with a suboptimal LCV score. The search is controlled by LCV scores and tolerances indicating how much of a percent decrease in the LCV score can be tolerated at given stages of the process. The base model is usually the model with constant means and constant dispersions but this can be adjusted (as in Section 3.5). First, the base model is systematically expanded by adding in power transforms one-at-a-time to the means or to the dispersions whichever generates the best LCV score over all predictors under consideration. The expansion stops when the LCV score generated by the next addition of a power transform generates a percent decrease in the LCV score larger than the largest allowable percent decrease based on a tolerance parameter.

By default, the expansion only considers additive models for predictors for the means and for the dispersions. However, the expansion can optionally also con-

sider geometric combinations between predictors. These consist of products of power transforms of multiple predictors considered in the expansion, generalizing standard interactions. Consideration of geometric combinations provides for a nonlinear assessment of what is called moderation [11], which is commonly based on only untransformed predictors.

The expanded model often is too complex and needs contraction. Each power transform in the model is removed and the powers of the remaining transforms are adjusted to improve the LCV score. The power transform whose removal generates the adjusted reduced model with best LCV score is removed next. The contraction stops when the percent decrease in the LCV score with the removal of the next transform is larger than the largest allowable percent decrease in the LCV score based on another tolerance parameter. This tolerance parameter is set so that the largest allowable percent decrease for the contraction equals the cutoff for a substantial percent decrease in the LCV score using $DF = 1$. In this way, all power transforms in the adaptively generated model determined by the contraction provide substantial improvements in the sense that the removal of each transform generates a substantially inferior model. The contraction usually removes transforms from the expanded model, but in cases when the contracted model leaves the expanded model unchanged, its powers are adjusted to improve the LCV score. A detailed description of the formal adaptive modeling process is provided in [3].

Adaptive regression modeling is based on a complex algorithm, but one that generates effective models. The powers for transforms of these models can be changed by small amounts without much impact on generated results. Plotted results will not change by much if the models have competitive LCV scores. Stopping the contraction using LCV ratio tests guarantees the need for each of the generated transforms possibly with small adjustments to their powers.

Slopes for power transforms of adaptively generated models are usually all significantly nonzero at much smaller p values than the usual choice of 0.05 for a cutoff for a p value. Furthermore, power transforms in models with slopes significant at values close to $p = 0.05$ are typically removed from a model during the contraction. Consequently, even though LCV ratio tests are based on the usual 5% significance level, those tests are more conservative than standard tests for zero slopes conducted at 5% significance. For this reason, tests of zero slopes of adaptively generated models should not be used for inferential purposes and so are not reported in the analyses of Section 3. Those analyses are conducted using a SAS[®] macro available from the author.

3. Results

The 2038 participants with available dioxin levels fathered 5478 live births with 3290 (60.1%) conceived before the start of the tour and 2188 (39.9%) conceived after the start of the tour. Of the births conceived before the start of the tour, 1891 (57.5%) were fathered by participants with low dioxin levels and 1399

(42.5%) by participants with high dioxin levels. Of the births conceived after the start of the tour, 1333 (60.9%) were fathered by participants with low dioxin levels and 855 (39.1%) by participants with high dioxin levels. The aggregated proportions of live births with major birth defects over time within the four categories of before/after the start of the tour and low/high dioxin levels are plotted in **Figure 1**. Proportions are quite small early on and late in time, higher in between, and highest when conceived after the start of the tour for participants with high dioxin levels.

Adaptive regression modeling is used in subsequent sections to address a variety of ITS modeling issues. Unless otherwise specified, the base model for these analyses has constant means and constant dispersions based on two intercept parameters. LCV score are computed using 10 folds, containing from 7 to 19 outcome measurements each. The cutoff for a substantial percent decrease computed using $DF = 1$ for the AFHS major birth defect data rounds to 1.5%.

3.1. Assessing an Uninterrupted Time Effect

Adaptive regression can be used to identify the possibly nonlinear underlying pattern for the estimated probability of a major birth defect over time under the assumption of no effect to conception before/after the start of the tour and no effect to a low/high dioxin level. The model is expanded using power transforms of the variable t equal to the year index as the only predictor for both the means and the dispersions, and then the expanded model is contracted.

The adaptively generated model has constant means and constant dispersions not depending on t . Under this model, the estimated constant probability for all 59 years is 0.044 with estimated constant standard deviation 0.066 and estimated

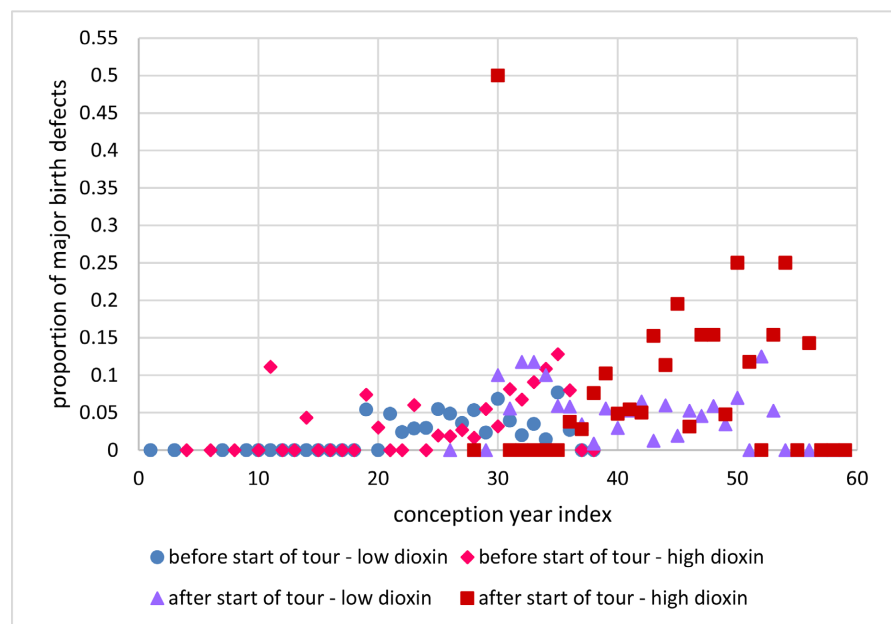


Figure 1. Observed proportions of major birth defects over time before and after the start of the Vietnam tour for participants with low and high dioxin levels.

correlation 0.22. The LCV score is 0.15262.

3.2. Assessing an Effect to Conception after the Start of the Tour

Adaptive regression modeling can be used to assess whether there is an interruption effect due to conception after the start of the tour assuming no effect to a low/high dioxin level. The expansion has means and dispersions both based on power transforms of the two predictor variables: the time variable t and the indicator variable 1_{after} with value 1 for conception after the start of the tour and 0 for conception before the start of the tour. The expansion also considers geometric combinations based on these two predictors. The addition of 1_{after} provides for a general level effect, that is, a discontinuity in the dependence over time for conception after the start of the tour. The addition of a power transform of the variable t provides for dependence over time independent of conception before/after the start of the tour. The addition of a power transform of the interaction $1_{\text{after}} \cdot t$ provides for a change in the dependence over time for conception after the start of the tour compared to before the start of the tour. Note that indicator variables like 1_{after} are unaffected by power transformation. The adaptively generated model has constant means and constant variances for conception before the start of the tour and nonlinear means after the start of the tour based on $1_{\text{after}} \cdot t^{-1.3}$ and nonlinear dispersions after the start of the tour based on $1_{\text{after}} \cdot t^{-1.17}$.

Figure 2 provides the plot of the estimated probability of a major birth defect over time while **Figure 3** provides the plot of the associated estimated standard deviation over time. For conception before the start of the tour, the estimated probability of a major birth defect is constant over years 1 - 38 at 0.026 while the

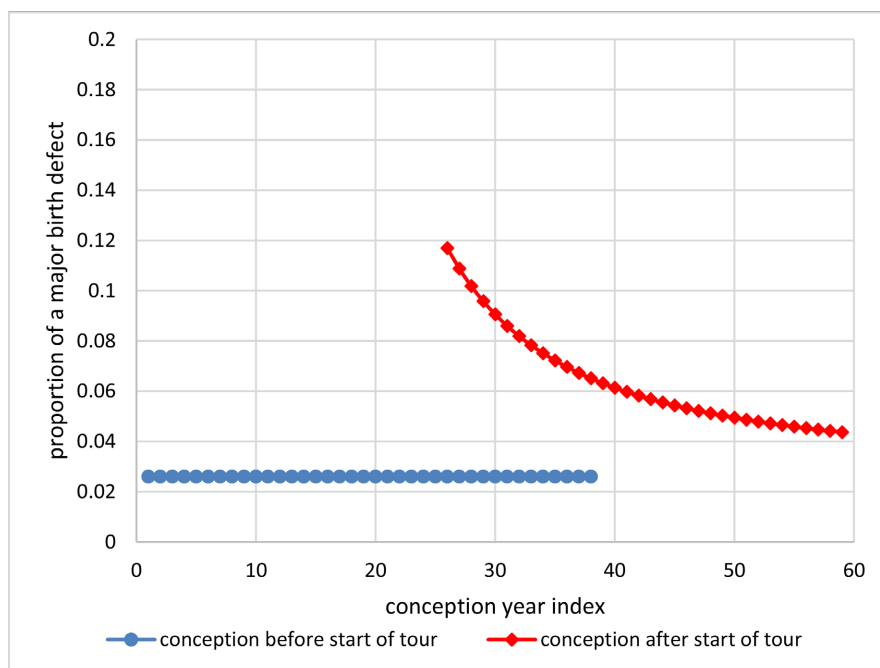


Figure 2. Estimated probability of a major birth defect over time before and after the start of the Vietnam tour for participants with measured dioxin values.

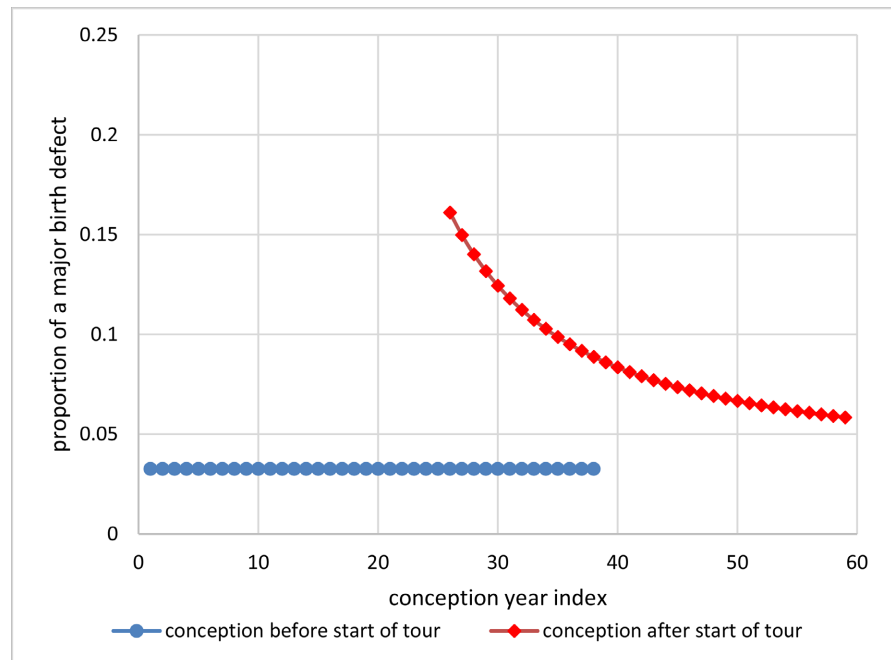


Figure 3. Estimated standard deviation for the proportion of major birth defects over time before and after the start of the Vietnam tour for participants with measured dioxin values.

associated estimated standard deviation is constant at 0.041. For conception after the start of the tour, the estimated probability of a major birth defect decreases nonlinearly from 0.117 at year 26 to 0.044 at year 59 while the associated estimated standard deviation has a similar nonlinear pattern decreasing from 0.161 to 0.058. The estimated autocorrelation is 0.28.

The LCV score for this model is 0.21321 while the LCV score is 0.15262 for the model in uninterrupted time of Section 3.1 with percent decrease 28.4%. This is a substantial percent decrease (*i.e.*, larger than the cutoff for the data of 1.5% for a substantial percent decrease) indicating that the model in uninterrupted time is not a competitive alternative, thereby supporting a substantial interruption effect to conception after the start of the tour with an increased probability of a major birth defect that decreases nonlinearly over time. Note that the percent decrease in the LCV score is used here to assess whether the interruption effect is substantial rather than using a p value.

3.3. Comparison to the Standard Straight-Line ITS Model

The standard straight-line model for an ITS interruption effect has means based on an intercept, the indicator 1_{after} to account for a level effect, the untransformed year index variable t , and the untransformed interaction $1_{\text{after}} \cdot t$ to account for a rotation effect. The dispersions are constant. Using standard Wald t tests, the slopes for 1_{after} , t , and $1_{\text{after}} \cdot t$ in this model are significantly nonzero with p values 0.002, <0.001, and <0.001, respectively, thereby supporting both a level effect and a rotation effect.

Figure 4 provides the plot of the estimated probability of a major birth defect over time. Although this is based on the straight-line model, the plotted values are somewhat curved due to using the natural log link function. For conception before the start of the tour, the estimated probability of a major birth defect increases from 0.004 at year 1 to 0.035 at year 38. For conception after the start of the tour, the estimated probability of a major birth defect decreases from 0.108 at year 26 to 0.049 at year 59. The plot for the standard deviations has a similar form, and so is not displayed. The estimated autocorrelation is 0.15.

The standard ITS model may not provide an accurate depiction of the underlying time varying pattern. This can be addressed using adaptive regression modeling as follows. The LCV score for the straight-line ITS model is 0.08054 with a very substantial percent decrease 62.2% compared to the LCV score 0.21321 of the adaptive ITS model of Section 3.2. This result indicates that the dependence on time is highly nonlinear. Furthermore, the percent decrease in the LCV score for the straight-line ITS model is 47.2% compared to the LCV score 0.15262 for the adaptively determined constant model in time of Section 3.1. This result indicates that the straight-line ITS model does not accurately represent these ITS data and suggests that in general using p values to assess ITS effects can be result in misleading conclusions.

3.4. Assessing Effects to Both Conception after the Start of the Tour and a High Dioxin Level

Adaptive regression modeling can be used to assess whether there is an interruption

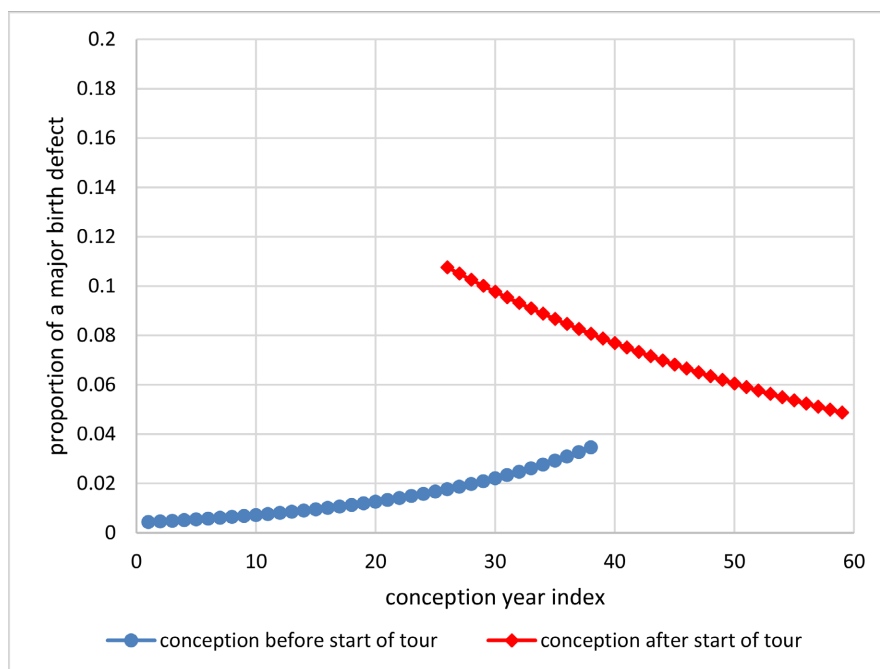


Figure 4. Estimated probability of a major birth defect over time before and after the start of the Vietnam tour for participants with measured dioxin values using the straight-line ITS model.

effect to conception after the start of the tour possibly moderated by a high dioxin level. The expansion has means and dispersions based on power transforms of the same three predictor variables: the time variable t , the indicator variable I_{after} , and the indicator I_{high} for participants having a high dioxin level. The expansion also considers geometric combinations based on products of two or three of these three predictors. The addition to the model of power transforms of the variables t and $I_{\text{after}} \cdot t$ serves the same purposes as before. The addition to the model of the indicator I_{high} provides for a discontinuity in the dependence over time for participants having a high compared to a low dioxin level. The addition of a power transform of the interaction $I_{\text{high}} \cdot t$ provides for a change in the dependence over time for a high compared to a low dioxin level while the addition of a power transform of the three-way interaction $I_{\text{after}} \cdot I_{\text{high}} \cdot t$ provides for a change in the dependence over time for conception after the start of the tour for participants with a high compare to a low dioxin level.

The adaptively generated model has constant means and constant variances for conception before the start of the tour for participants with low dioxin levels. For the other cases, the model has nonlinear means based on $I_{\text{after}} \cdot t^{-1.56}$ and $I_{\text{high}} \cdot t^{0.2}$ and nonlinear dispersions based on $I_{\text{after}} \cdot I_{\text{high}} \cdot t^{-1.17}$.

Figure 5 provides the plot of the estimated probability of a major birth defect over time while **Figure 6** provides the plot of the associated estimated standard deviation over time. For conception before the start of the tour for participants with low dioxin levels, the estimated probability of a major birth defect is constant over years 1 - 36 at 0.018 while the associated estimated standard deviation is constant at 0.037. For conception before the start of the tour for participants

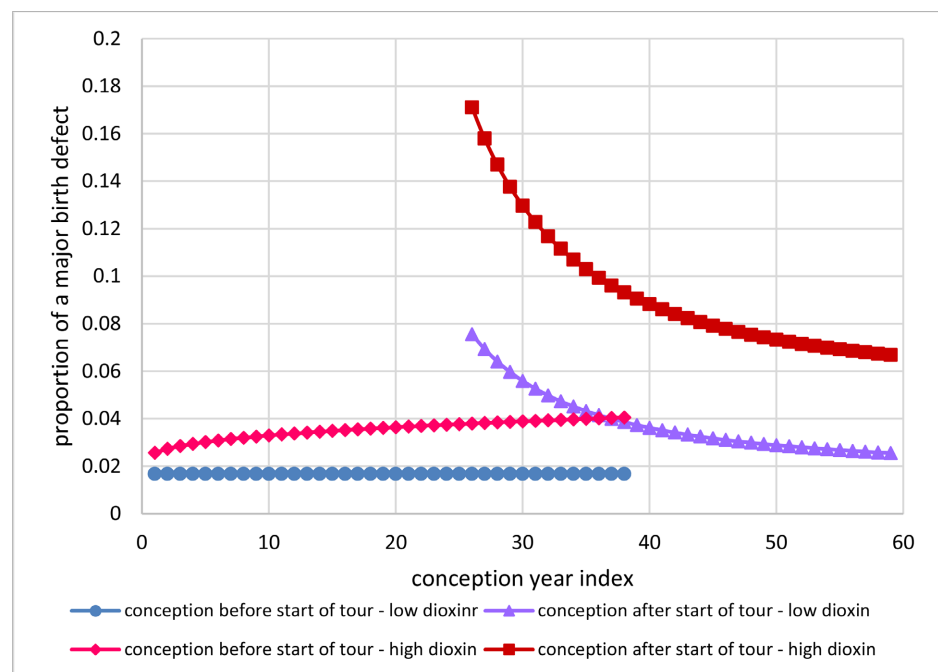


Figure 5. Estimated probability of a major birth defect over time before and after the start of the Vietnam tour for participants with low and high dioxin levels.

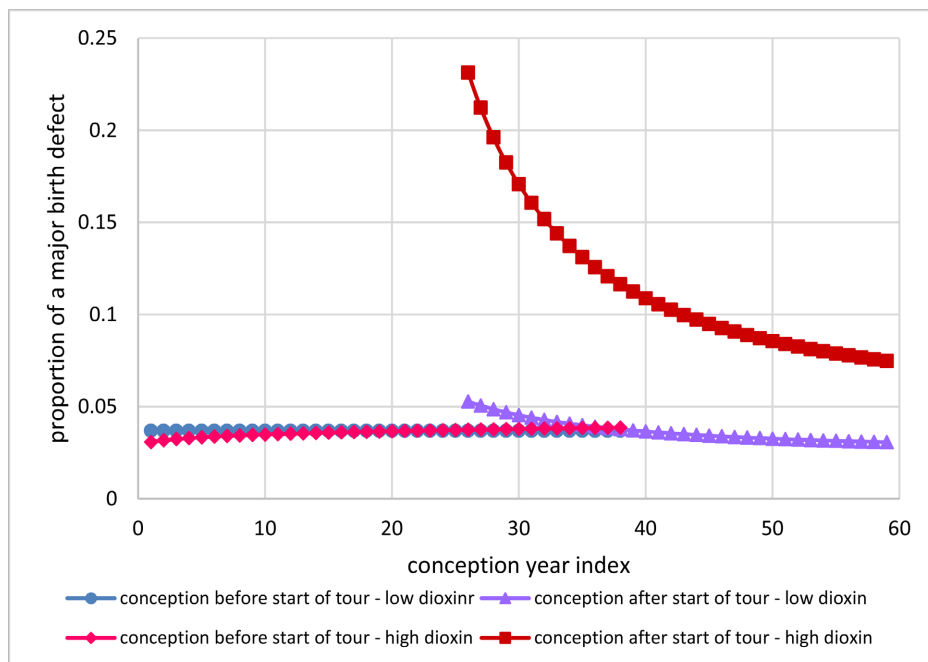


Figure 6. Estimated standard deviation for the proportion of major birth defects over time before and after the start of the Vietnam tour for participants with low and high dioxin levels.

with high dioxin levels, the estimated probability of a major birth defect increases somewhat from 0.026 at year 1 to 0.040 at year 36 while the associated estimated standard deviation increases only a little from 0.031 to 0.038. For conception after the start of the tour for participants with low dioxin levels, the estimated probability of a major birth defect decreases nonlinearly from 0.076 at year 26 to 0.028 at year 59 while the associated estimated standard deviation decreases somewhat from 0.053 to 0.031. For conception after the start of the tour for participants with high dioxin levels, the estimated probability of a major birth defect decreases nonlinearly from 0.171 at year 26 to 0.067 at year 59 while the associated estimated standard deviation decreases nonlinearly from 0.231 to 0.075. The estimated autocorrelation is 0.25.

The LCV score for this model is 0.26684 while the LCV score for the model in time and conception after the start of the tour of Section 3.2 is 0.21321 with percent decrease 20.1%. This is a substantial percent decrease indicating that the model in time and conception after the start of the tour unadjusted for dioxin level is not a competitive alternative, thereby supporting a substantial interruption effect to conception after the start of the tour that is distinctly moderated by having a high dioxin level. There is an increased probability of a major birth defect after the start of the tour for participants having a high dioxin level. While that probability decreases over time, it is higher for each of the years 26-59 than the associated probability for participants with a low dioxin level. This indicates that the interruption effect identified in Section 3.2 is reasonably considered to be related at least in part to participants having a high dioxin exposure level.

3.5. Assessing Constant Low/High Dioxin Effects for Conception before the Start of the Tour

The curves in **Figure 5** for conceptions after the start of the tour appear non-constant over time for each of the low and high dioxin levels. On the other hand, for conception before the start of the tour, the curve for a low dioxin level is exactly constant over time while the curve for a high dioxin level is only mildly increasing over time. This suggests the possibility that both the low and the high dioxin effects for conceptions before the tour might be constant over time.

Adaptive regression modeling can be used to assess this issue. Start with a base model having an intercept and a slope for the interaction $I_{\text{before}} \cdot I_{\text{high}}$ where $I_{\text{before}} = 1 - I_{\text{after}}$ is the indicator for conception before the start of the tour. The interaction $I_{\text{before}} \cdot I_{\text{high}}$ represents a constant effect over time to a high dioxin level for conceptions before the start of the tour. Base the expansion for the means on the predictors $I_{\text{after}} \cdot I_{\text{high}}$ and $I_{\text{after}} \cdot I_{\text{high}} \cdot t$ representing respectively a discontinuity and a nonlinear dependence on time for a high dioxin level for conception after the start of the tour. Note that these latter two interactions need to be computed prior to the adaptive search rather than automatically generated as geometric combinations so that the dependence of the means for conceptions before the start of the tour is not changed by the expansion. The dispersions are based on the same three predictor variables as before: the time variable t , the indicator variable I_{after} , and the indicator I_{high} along with geometric combinations based on two or three of these three predictors, but geometric combinations are automatically generated only for the dispersions. In this way, the expanded model for conceptions before the start of the tour depends on only an intercept and the indicator $I_{\text{before}} \cdot I_{\text{high}}$ and so with different constant effects for low/high dioxin levels. An adaptive relationship in time for means is generated for conception after the start of the tour for low/high dioxin levels and a general adaptive relationship is generated for the dispersions. The contraction allows the interaction $I_{\text{before}} \cdot I_{\text{high}}$ to be removed from the means to account for the possibility of no difference between the low/high dioxin effects for conception before the start of tour. The adaptively generated model has means based on an intercept and $I_{\text{after}} \cdot I_{\text{high}} \cdot t^{-0.7}$ and dispersions based on an intercept, $I_{\text{after}} \cdot t^{-1.26}$, $I_{\text{high}} \cdot t^{0.1}$, and $I_{\text{after}} \cdot I_{\text{high}} \cdot t^{-0.18906}$.

Figure 7 provides the plot of the estimated probability of a major birth defect over time while **Figure 8** provides the plot of the associated estimated standard deviation over time. For the three cases of conception before the start of the tour for participants with a low dioxin level and with a high dioxin level as well as conception after the start of the tour for participants with a low dioxin level, the estimated probability is constant over time at 0.026. For conception after the start of the tour for participants with high dioxin levels, the estimated probability decreases nonlinearly from 0.139 at year 26 to 0.067 at year 59. For conception before the start of the tour for participants with low dioxin levels, the estimated standard deviation is constant at 0.023. For conception before the start of the

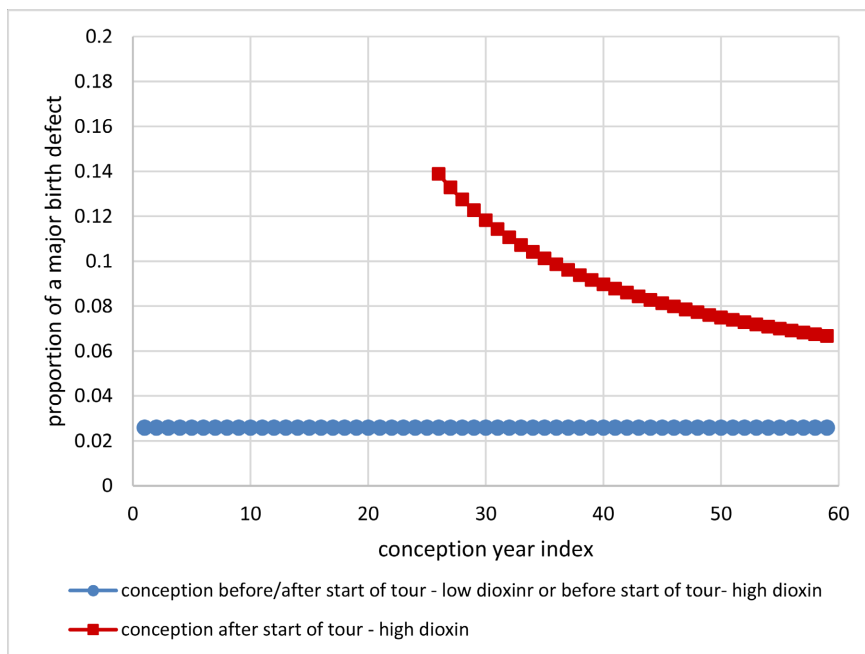


Figure 7. Estimated probability of a major birth defect over time before and after the start of the Vietnam tour for participants with low and high dioxin levels assuming constant probabilities for low and high dioxin levels before the start of the tour.

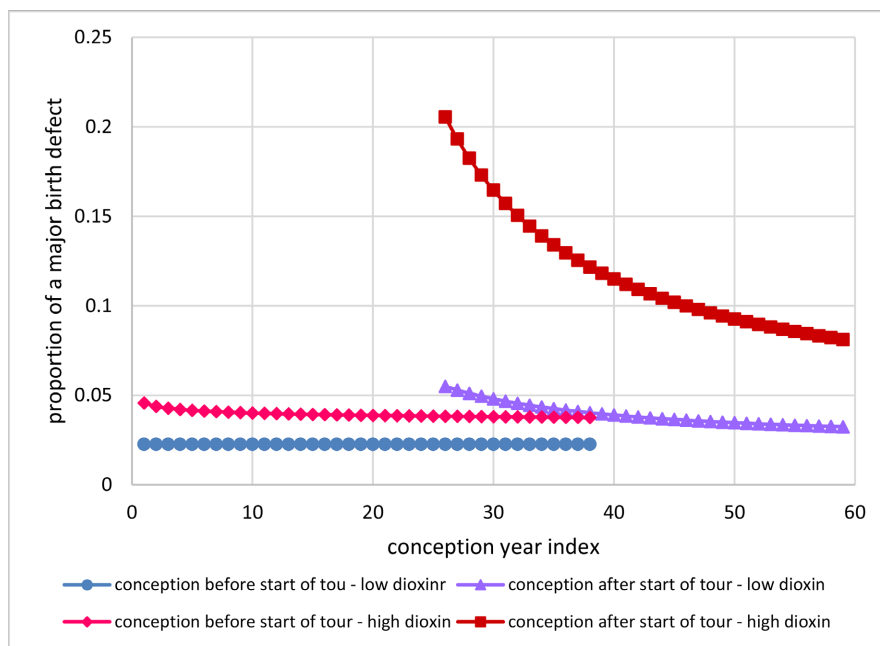


Figure 8. Estimated standard deviation for the proportion of major birth defects over time before and after the start of the Vietnam tour for participants with low and high dioxin levels assuming constant probabilities for low and high dioxin levels before the start of the tour.

tour for participants with high dioxin levels, the estimated standard deviation decreases somewhat from 0.046 to 0.038. For conception after the start of the tour for participants with low dioxin levels, the estimated standard deviation de-

creases somewhat from 0.055 to 0.032. For conception after the start of the tour for participants with high dioxin levels, the estimated standard deviation decreases nonlinearly from 0.206 to 0.081. The estimated autocorrelation is 0.32.

The LCV score for this model is 0.26789 while the LCV score is 0.26684 for the model in time, conception after the start of the tour, and a high dioxin level of Section 3.4 with small percent decrease 0.4%. This indicates that the two models are competitive alternatives. The model of Section 3.4 is based on six parameters including an intercept and two slopes for the means, an intercept and a slope for the dispersions, and an autocorrelation. The model of this section is more complex, based on seven parameters including an intercept and one slope for the means, an intercept and three slopes for the dispersions, and an autocorrelation. Although the model of this section is more complex, it is a competitive alternative with a larger LCV score, and so it can reasonably be used to draw conclusions about how the probability of a major birth defect changes over time. Both models support the conclusion that the probability of a major birth defect for live births conceived after the start of the tour for participants with a high dioxin level decreases over time but always at a higher level at each time than for the other three conception cases. However, the model of this section indicates that the variability in the probabilities for the other three cases identified in the Section 3.4 model can be reasonably considered to result from variability around a constant probability over time. Consequently, the model of this section provides support for the conclusion that the probability of a major birth defect is constant except for an increase for conception after the start of the tour for participants with a high dioxin level that decreases in time after the start of the tour. In other words, the interruption effect on the probability of a major birth defect identified in **Figure 2** can reasonably be fully attributed to a high dioxin exposure level.

3.6. Assessing the Impact of Autoregressive Correlations

The estimated autocorrelation for the model of Section 3.5 is 0.32. This is a relatively small value with limited impact on temporal correlations. Specifically, rounded estimated correlations for outcome measurements 2 - 4 years apart are 0.10, 0.03, and 0.01, respectively, and are less than 0.01 for 5 or more years apart. This suggests that treating outcome measurements as independent might provide a competitive model. LCV scores can be used to make this assessment serving as an alternative to a hypothesis test of zero autocorrelation. For the model of Section 3.5 with the autocorrelation set to zero, the LCV score is 0.25126 with percent decrease 6.2% compared to the model allowing for a nonzero autocorrelation. This percent decrease is substantial and so large enough to support the importance in this case for accounting for autocorrelation.

4. Discussion

4.1. Overview of Analysis Results

Adaptive regression methods were used to analyze the interrupted time series

(ITS) consisting of annual proportions of major birth defects in children fathered by male Air Force veterans of the Vietnam War conceived after versus before the start of a participant's first tour in the Vietnam War (Figure 1). A substantial adverse ITS interruption effect was identified. For births conceived before the start of the tour, the estimated probability of a major birth defect (Figure 2) was constant at 0.026 over time. In contrast, for births conceived after the start of the tour, the estimated probability decreased nonlinearly over time from 0.116 to 0.044 always at a larger value than for births conceived before the start of the tour. This result supports the conclusion that conception after the start of the tour had an adverse effect on the occurrence of major birth defects for children fathered by Vietnam veterans.

Adaptive regression was also used to assess whether the identified adverse effect to conception after the start of the tour could be related in some way to the father's exposure to dioxin. That adverse effect can be fully attributed to participants having a high dioxin exposure level, that is, a dioxin value greater than 6.3 ppt as identified in a previous analysis of these data. For births conceived after the start of the tour for participants with a high dioxin exposure level, the estimated probability over time of a major birth defect (Figure 7) decreased over time from 0.139 to 0.067 while the estimated probability over time was constant at the lower value of 0.026 for the other three cases of births conceived before the start of the tour for participants with either a low or a high dioxin exposure level and births conceived after the start of the tour for participants with a low dioxin exposure level. One possible explanation for the decrease in the adverse high dioxin effect over time could be decay in the dioxin exposure over time. Another possible explanation is that participants who fathered a child with a major birth defect were less likely to father more children.

Observed detectable dioxin values ranged from 1.25 to 314.2 ppt, and so 6.3 ppt seems like a small value. However, 1202 (59.0%) of the participants with available dioxin values had a low dioxin level. Also, dioxin values were determined from blood draws taken from 9.8 to 30.4 years after the end of a participant's last Vietnam tour, and so exposures in Vietnam were likely quite a bit larger.

The CDC conducted a study of major birth defects in metropolitan Atlanta, GA over the period 1978-2005, the beginning of which overlaps with the end of the 59-year period for the AFHS data. They concluded that the prevalence of a major birth defect was relatively stable over that period ranging from 2.8 per 100 births in 1978 to 3.0 per 100 births in 2005 [12]. It is not clear whether the CDC and the AFHS used the same definition of a major birth defect, but the estimated baseline probability identified using the AFHS data of 0.026, corresponding to a prevalence of 2.6 per 100, is very consistent with the CDC results.

Adaptive regression methods were used to assess whether the standard straight-line ITS model based on untransformed time supported the conclusion of a substantial interruption effect. Adverse ITS level and rotation effects were identified using tests for significant slopes of this standard model. However, the

adaptive model constant in time and independent of conception before/after the start of the tour of Section 3.1 provided a substantial improvement over the standard straight-line ITS model as measured with LCV scores. Consequently, the use of p values for the standard straight-line ITS model to assess level and rotation effects can sometimes result in misleading conclusions. Adaptive models allowing for nonlinear patterns over time are needed to support or refute such conclusions. They might also be needed in general to identify ITS effects when the standard straight-line ITS model does not.

Adaptive regression methods were used to assess whether there was a benefit to accounting for autocorrelation. Autocorrelation was important to account for in reported analyses as measured by LCV scores.

Adaptive regression methods were used to account for possible nonconstant dispersions. Nonconstant dispersions were important to account for. The model allowing for nonlinearity in probabilities for conception both before and after the start of the tour depicted in **Figure 5** provided support for an effect to participants having a high dioxin exposure level on conception after the start of the tour, but raised the question of how could there be such a dioxin effect on conception before the start of the tour when presumably the dioxin exposure had not yet occurred. The competitive model of **Figure 7** indicates that the probability of a major birth defect for conception before the start of the tour can be reasonably considered to be unaffected by a high dioxin exposure level as would be expected. Furthermore, it indicates that this conclusion holds as well for conception after the start of the tour for participants with a low dioxin exposure level. The differences in probabilities of **Figure 5** for these three baseline cases is accounted for by variability for these three cases depicted in **Figure 8**, indicating that the identification of a constant baseline probability for these three cases was a consequence of allowing for more complex dispersions.

4.2. Limitations

Shortcomings of the AFHS may have influenced reported results. For example, veterans with children having major birth defects may have been more likely to participate. Veterans with measured dioxin may have been more likely to have children with major birth defects. Measuring dioxin exposures long after the presumed exposure in Vietnam may have had unknown effects on results. Participants may have had other toxic exposures. Maternal exposures to dioxin or to other toxic substances have not been accounted for. Reported results are only generally applicable if all study shortcomings are inconsequential, which cannot be verified. However, the conclusions are appropriate for the substantial numbers of 5478 births for 2038 male Air Force veterans of the Vietnam War indicating that a high dioxin exposure level had an adverse impact resulting in an increased level of major birth defects for children of Vietnam veterans represented by this sample.

Adaptive regression modeling is not directly supported in standard statistical

software tools. However, adaptive analyses of general data sets can be conducted using a SAS[®] macro available from the author.

4.3. Conclusions

In summary, adaptive regression modeling can identify nonlinear ITS effects in general situations that can lead to consequential insights into relationships over time, possibly varying with other available predictors. Specifically, adaptive regression modeling provided substantial support for an adverse nonlinear ITS effect on the probability of a major birth defect over time after the start of the first tour in Vietnam for children of male Vietnam veterans. Moreover, it provided substantial support for the conclusion that this ITS effect could be fully attributed to male veterans' dioxin exposure after the start of their first Vietnam tour.

The AFHS collected a variety of other reproductive outcomes including a child having multiple birth defects, only minor birth defects, a developmental disability, a low birth weight, and defects within a variety of ICD-9 codes, among others. Future research is needed to address nonlinear ITS effects on these outcomes related to conception after the start of the tour and whether such effects can be attributed in some way to dioxin exposure. Comparisons are also needed of results on ITS effects to conception after the start of the tour for participants with missing dioxin values to participants with measured dioxin values to assess whether having a measured dioxin level influenced conclusions about ITS effects. Analyses are needed to address differences in ITS effects for a variety of characteristics, for example, child sex, parental/maternal age, and race, among others. Analyses are also needed to assess whether categorizing dioxin exposures into low versus high levels has somehow affected results in comparison to using observed continuous dioxin values.

Reported analyses accounted appropriately for autocorrelation over time in annual proportions of major birth defects within each of the conception before and after groups, but treated annual proportions of conceptions before the start of the tour as independent of annual proportions for conceptions after the start of the tour. This shortcoming may not have had much of an impact for this analysis given the relatively small autocorrelation estimates. It is only an issue for ITS effects like the one considered here with overlap in time between pre-interruption and post-interruption outcome measurements. Future research is needed to account in general not only for autocorrelation over time in outcome measurements within each of the before and after ITS interruption groups but also for correlations between outcome measurements in the before ITS group with outcome measurements in the after ITS intervention group. Future research is also needed into adaptive regression modeling using more complex ARIMA correlation structures.

Acknowledgements

The author acknowledges support to conduct statistical analyses of AFHS re-

productive outcome data as part of project at the Yale School of Nursing funded by a grant in 2000 from the Vietnam Veterans of America, Linda S. Schwartz Principal Investigator. Development of the SAS macro used to conduct reported analyses was supported in part by grants R01 AI57043 from the National Institute of Allergy and Infectious Diseases and R03 MH086132 from the National Institute of Mental Health.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Turner, S.L., Karahalios, A., Forbes, A.B., Taljaard, M., Grimshaw, J.M., Cheng, A.C., Bero, L. and McKenzie, J.E. (2020) Design Characteristics and Statistical Methods Used in Interrupted Time Series Studies Evaluating Public Health Interventions: A Review Study. *Journal of Clinical Epidemiology*, **122**, 1-11. <https://doi.org/10.1016/j.jclinepi.2020.02.006>
- [2] Tetali, S., Jammy, G., Asirvatham, E., Kumar, B. and Choudhury, L. (2021) An Interrupted Time Series Analysis of COVID-19 Positivity before, during and after Lockdown in Four States of India. *Open Journal of Epidemiology*, **11**, 47-55. <https://doi.org/10.4236/ojepi.2021.111005>
- [3] Knafl, G.J. and Ding, K. (2016) Adaptive Regression for Modeling Nonlinear Relationships. Springer International Publishing, Switzerland. https://doi.org/10.1007/978-3-319-33946-7_20
- [4] Knafl, G.J. Modeling Correlated Outcomes Using Extensions of Generalized Estimating Equations and Linear Mixed Modeling. Springer Nature, Berlin. (In Press)
- [5] Wolfe, W.H., Michalek, J.E., Miner, J.C., Rahe, A.J., Silva, J., Thomas, W.F., Grubbs, W., Lustig, M.B., Karrison, T.G., Roegner, R.H. and Williams, D.E. (1990) Health Status of Air Force Veterans Occupationally Exposed to Herbicides in Vietnam: I. Physical Health. *Journal of the American Medical Association*, **264**, 1824-1831. <https://doi.org/10.1001/jama.1990.03450140046032>
- [6] Michalek, J.E., Wolfe, W.H. and Miner, J.C. (1990) Health Status of Air Force Veterans Occupationally Exposed to Herbicides in Vietnam: II. Mortality. *Journal of the American Medical Association*, **264**, 1832-1836. <https://doi.org/10.1001/jama.1990.03450140054033>
- [7] Wolfe, W.H., Michalek, J.E., Miner, J.C., Rahe, A.J., Moore, C.A., Needham, L.L. and Patterson Jr., D.G. (1995) Paternal Serum Dioxin and Reproductive Outcomes among Veterans of Operation Ranch Hand. *Epidemiology*, **6**, 17-22. <https://doi.org/10.1097/00001648-199501000-00005>
- [8] The Air Force Health Study Assets Research Program. <https://www.ncbi.nlm.nih.gov/books/NBK286036/>
- [9] Knafl, G.J. (2018) A Reassessment of Birth Defects for Children of Participants of the Air Force Health Study. *Open Journal of Epidemiology*, **8**, 187-200. <https://doi.org/10.4236/ojepi.2018.84015>
- [10] Knafl, G.J. and Meghani, S.H. (2021) Regression Modeling of Individual-Patient Correlated Discrete Outcomes with Applications to Cancer Pain Ratings. *Open Journal of Statistics*, **12**, 456-485. <https://doi.org/10.4236/ojs.2022.124029>

- [11] Baron, R.M. and Kenny, D.A. (1986) The Moderator-Mediator Variable Distinction in Social Psychology Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality & Social Psychology*, **51**, 1173-1182.
<https://doi.org/10.1037/0022-3514.51.6.1173>
- [12] Update on Overall Prevalence of Major Birth Defects, Atlanta, Georgia, 1978-2005.
<https://cdc.gov/>