Use of a Bayesian Changepoint Model to Estimate Effects of a Graduated Driver's Licensing Program

Michael R. Elliott[1] and Jean T. Shope[2]
[1]*University of Pennsylvania,* [2]*University of Michigan*

Abstract: In April 1997 the US state of Michigan implemented a graduated licensing program for novice drivers under the age of 18 that ensures that they gain experience and maturity under conditions of low risk before progressing to more risky driving situations. Since there is no reasonable control group of young Michigan drivers not exposed to graduated licensing during this period, the extent to which observed declines in crash rates can be attributed to graduated licensing versus other unobserved changes in crash reporting or driving behavior is important. We assemble a Bayesian changepoint model to assess the probability that changes in crash rate trends among 16-year Michigan drivers can be plausibly linked to the introduction of graduated licensing and, if it can, to make inference about graduated driver licensing effects that take into account the uncertainty in when these effects began and ended, and whether or not a "rebound" in crash rates occurs afterward. We show that, while there is a moderate degree of sensitivity to the choice of prior distributions for changepoints and rate slopes in determining the number of changepoints present in the crash trends, inference about whether GDL effects are present and their degree are relatively insensitive to prior choice. This analysis suggests that the decline in crash rates among 16-year old Michigan drivers observed between 1996 and 1998 can be reasonably attributed to graduated licensing for all crashes combined, but that observed changes in single-vehicle and especially nighttime crash rates might have been part of longer-term trends among this age group.

Key words: Linear splines, model averaging, model selection, prior sensitivity.

1. Introduction

On April 1, 1997 Public Act 387 went into effect in the US state of Michigan, requiring all unlicensed Michigan residents under the age of 18 who wished to obtain driver's licenses to do so through a graduated driver licensing (GDL) program. This program required a more structured introduction to driving than the previous law, and thus its effect on the crash rates of young Michigan drivers is a key question of interest. To answer this question, we construct a linear changepoint model that assumes a linear spline model for the mean crash rates conditional on the number and location of the changepoints or pulses, with the regression slopes and the number and location of the changepoints as parameters with specified prior distributions. We bring together several modeling ideas (Berry and Hartigan 1993, Carlin, Gelfand and Smith 1992, and Stephens 1994) into a unified framework that combines the flexibility of non-parametric regression models with the simplicity of interpretation of parametric regression models.

The effect of graduated licensing programs has been examined in New Zealand (Langley, Wagenaar, and Begg 1996), the Canadian provinces of Ontario (Boase and Tasca 1998) and Nova Scotia (Mayhew, Simpson, and desGroseilliers 1999), and the US state of Florida (Ulmer et al. 1999). The estimated effects have varied from a crash rate decline of 11% among 16 year-olds in the calendar years previous and subsequent to GDL introduction in Florida, where only time-of-day restrictions were put into place, to a crash rate decline of 31% between the calendar years preceding and following introduction of graduated licensing in Ontario, where a graduated licensing program involving supervision and other license restrictions for novice drivers of all ages was enacted.

The Canadian and US analyses considered before-GDL versus after-GDL crash rate differences via relative risk measures, where the "before" and "after" time points were chosen on an ad-hoc basis, typically the calendar year before and the one or two calendar years after the introduction of graduated licensing. Comparisons of these sorts may be confounded with underlying trends independent of GDL, and such confounding is difficult to sort out since typically there is no true control group of young drivers not subject to graduated licensing. In analysis of the New Zealand GDL program Lan-

44

gley et al. utilized ARIMA time series models (Box and Jenkins 1976) to estimate the effect of GDL on crash-related hospitalization rates. In the ARIMA modeling, it was assumed that the effect of GDL occurred in the form of a single "pulse" that occurred five months after the introduction of GDL; again, this choice of timepoint is reasonable but ad-hoc. Alternatively one might fit nonparametric regression models such as locally-weighted running line smoothers (Cleveland 1979) or cubic splines (Hastie and Tibshirani 1993) to the crash rate trend data and observe where and to what degree the rate appears to buckle, but the simple interpretability of the pulse or relative risk parameters is lost. Our analysis allows us to

1. Construct mean crash rate estimates that capture non-linear trends;

2. Assess the posterior probability that the changepoints (if any) occur within time frames that can be directly associated with GDL introduction;

3. Provide inference about the impact of graduated licensing that includes variability due to the uncertainty in estimating when the impact began and ended and whether the observed rate changes are plausibly related to GDL;

4. Provide inference about any "rebounds" in crash rates; and

5. Provide summary model checks of functional and distributional forms via posterior predictive distributions.

1.1 Michigan's graduated driver licensing program

Prior to the introduction of GDL, young Michigan drivers could obtain a driver's license at age 16 with parental permission after passing both a driver's education course (with on-the-road driving) and a written test. Under graduated licensing, residents 14 years and 9 months and older who have completed the first segment of driver's education can obtain a "Level I" license approximately equivalent to the pre-GDL learner's permit, which allowed the holder to drive under parental supervision. But 16-year-old drivers wishing to obtain a "Level II" license permitting independent driving

45

privileges are now required to: 1) complete a second segment of driver's education; 2) spend 50 hours of behind-the-wheel practice driving, including 10 hours of nighttime driving, with a licensed parent or guardian (or a designate if no parent or guardian is a licensed driver); 3) pass a road test conducted by a state-approved testing firm; 4) be free of traffic-related moving violations for the prior 3 months; and 5) have held a Level I license for 6 months. Level II licenses also are more restrictive than pre-GDL licenses in that no unsupervised driving is allowed between midnight and 5 am unless driving to or from places of employment. 17-year-old drivers can graduate to Level III licensing (no time-of-driving restrictions) after holding a Level II license for at least 6 months and going at least 12 months without a moving violation or violation of graduated licensing requirements. Drivers are "graduated" from graduated licensing at age 18, though they remain in the state's existing three-year probationary period for all new licensees.

Because many young drivers were in the process of obtaining their licenses under the pre-GDL system as of April 1, 1997, the introduction of the program was somewhat gradual among 16 year-olds. Those who had started a driver's education course on or before March 31, 1997 were not subject to the graduated licensing requirement, except that those who had not yet obtained a full license were required to pass the road test along with their written test, rather than substitute the driver's education road experience. Hence through mid-1999 the cohort of 16 year-old drivers included a mix of pre-GDL, Level I, and Level II drivers, although few pre-GDL drivers remained after mid-1998.

An analysis of the Michigan GDL law by Shope et al. (2001) found that the 1998 crash rate among Michigan 16 year-olds declined by 26%(95% CI 25%-28%) from the 1996 crash rate, with declines of 46%(39%-51%) for nighttime (midnight-5am) crashes, and 31%(28%-34%) for single-vehicle crashes (all $p < .0001$). Note that this analysis implicitly assumes that all changes between 1996 and 1998 were due to the introduction of GDL.

2. A Bayesian Linear Spline Changepoint Model

Our outcomes of interest are crash rates for three types of crashes involving 16-year-old drivers: all crashes, nighttime crashes (midnight-5 am), and

Table 1: Crash rates per 1,000 Michigan 16-year-olds

period	(a)	(b)	(c)	period	(a)	(b)	(c)
1	32.64	1.37	7.72	15	34.83	1.48	7.40
2	36.80	1.40	8.34	16	33.77	1.24	8.99
3	40.18	1.73	9.22	17	26.62	0.80	6.50
4	43.72	1.99	11.61	18	27.25	0.81	5.21
5	34.65	1.41	8.80	19	28.52	1.04	6.26
6	37.61	1.62	7.88	20	30.82	0.75	8.08
7	39.07	2.21	9.07	21	24.07	0.63	5.69
8	47.05	1.79	12.87	22	27.53	0.62	5.48
9	36.96	1.44	9.75	23	29.43	0.71	5.88
10	35.14	1.69	7.40	24	29.71	0.70	7.89
11	38.42	1.45	8.32	25	27.88	0.42	7.54
12	43.46	1.68	12.37	26	29.06	0.51	6.17
13	34.71	1.15	8.09	27	30.18	0.79	6.32
14	36.85	1.50	7.52	28	33.88	0.62	8.41

(a)= all crashes , (b)= nightime (midnight-5am) , and (c)= single-vehicle (SV) crashes. First period is January to March of 1994, and the last period is October to December of 2000.

single-vehicle crashes. These data are available from the Michigan State Police, and have been computed on a per-1,000-16 year-old basis at quarterly periods from January-March 1994 through October-December 2000 (see Table 1). The estimated total population of MI 16-year-olds during 1994 through 2000 (US Census 2001) has been used as the denominator to account for the fact that 16 year-olds driving under pre-GDL learner's permits were not counted as licensed drivers, although they were included in the count of police-reported crashes. We use a changepoint model since the crash rate trends must be continuous, and we want to capture the effect, if any, of the GDL pulse on the crash rates. By using a larger number of changepoints we can also capture more complex underlying trends in the data. The relatively limited number of data points – 28 – suggests that a linear trend between the points will be sufficient to estimate the mean rates

47

if enough changepoints are used. We rescale the crash rates y_t for a given type of crash to have sample mean 0 and sample variance 1, in order to easily "tune" the prior distributions as we will discuss below. We assume

$$y_t \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, K = k \sim N(\eta_t, \sigma^2)$$

$$\eta_t \equiv \eta_t(\boldsymbol{\theta}, k) = \beta_0 + \sum_{j=0}^{k} \beta_{j+1}(t - \theta_j)_+ + \boldsymbol{\gamma}' \boldsymbol{x}_t \qquad (2.1)$$

where $t = 1, \ldots, T = 28$, $\theta_0 = 0$, $(z)_+$ equals z if $z > 0$ and 0 otherwise and $\boldsymbol{x}_t$ consists of dummy variables for seasonal effects (January-March, April-June, July-September, October-December) centered at 0 (that is, $\boldsymbol{x}_t$ has dummy variables for winter, spring, and summer, with all three set to -1 for winter; this allows de-trending of seasonal effects without changing the overall location of the mean). Thus we assume that, after adjusting for seasonal effects (assumed constant over the crash period), the mean crash rate $\mu_t \equiv \mu_t(\boldsymbol{\theta}, k) = \beta_0 + \sum_{j=0}^{k} \beta_{j+1}(t - \theta_j)_+$ follows linear trends parameterized by $\boldsymbol{\beta}$ between changepoints, where both the location of the changepoints $\theta_1, \ldots, \theta_k$ and the number of changepoints k is unknown. The prior specifications are then given by

$$\boldsymbol{\beta} \mid \boldsymbol{\theta}, K = k \sim N_{k+2}(0, dI) \qquad (2.2)$$

$$\boldsymbol{\gamma} \mid \boldsymbol{\theta}, K = k \sim N_3(0, dI)$$

$$\sigma^2 \mid \boldsymbol{\theta}, K = k \sim \text{IG}(.001, .001)$$

$$P(\boldsymbol{\theta} \mid K = k) = \prod_{j=1}^{k} P(\theta_j \mid \theta_0, \ldots, \theta_{j-1}, K = k) = \prod_{j=1}^{k} P(\theta_j \mid \theta_{j-1}, K = k)$$

where

$$\begin{matrix} P(\theta_j \leq t + 1 \mid \theta_{j-1} = t, K = k) = 0 \\ P(\theta_j > t + 1 \mid \theta_{j-1} = t, K = k) \propto 1 \end{matrix} \quad \text{for } j = 1, \ldots, k.$$

$$K \sim p(k)$$

where $N_p(\mu, \Sigma)$ is the p-variate normal distribution and $\text{IG}(a, b)$ is the inverse-gamma distribution given by $f(x; a, b) = \Gamma(a)^{-1} b^{-a} x^{-(a+1)} e^{-1/bx}$.

Thus we have a conjugate prior on the linear spline trend and seasonal effects, an essentially flat prior on residual variance; and, conditional on the number of changepoints, a uniform prior on the location of these changepoints in time, subject to the constraint that the changepoints must be separated by at least 6 months. Note that any discrete distribution could be used as a prior for the changepoints. We use a constrained discrete uniform prior that forces the linear spline to have at least three points estimating the linear trend. We choose a uniform distribution because we do not want to favor one point over another *a priori*: we want the data to tell us where the crash rates change, not let our preconceptions or even our observations the time series plots (which can be prone to optical illusions) make the decision.

Elements of these models are discussed in Berry and Hartigan (1993), Carlin Gelfand and Smith (1992), and Stephens (1994); they can also be viewed as a special case of "Bayesian Treed models" discussed in Chipman et al. (2001).

2.1 Posterior distributions

Integrating out $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and σ^2 from the likelihood yields (e.g., Halpern 1973)

$$p(\boldsymbol{y} \mid \theta_1 = t_1, \ldots, \theta_k = t_k, k) = \pi(\boldsymbol{y}, \boldsymbol{\theta}, k) \propto$$
$$d^{-(k+5)/2} |d^{-1} I_{k+5} + X'_{\theta,k} X_{\theta,k}|^{-1/2}$$
$$\times \left(\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}' X_{\theta,k} (d^{-1} I_{k+5} + X'_{\theta,k} X_{\theta,k})^{-1} X'_{\theta,k} \boldsymbol{y} \right)^{-T/2} \quad (2.3)$$

where $X_{\theta,k}$ is the design matrix for (2.1) where $\boldsymbol{\theta}$ and K are fixed at $\theta_1 = t_1, \ldots \theta_k = t_k$ and k. Consequently we can factor the joint posterior distribution into our key posterior distributions of interest:

$$p(k \mid y) = \frac{[p(k)/c_k] \sum_\theta \pi(\boldsymbol{y}, \boldsymbol{\theta}, k)}{\sum_k \{[p(k)/c_k] \sum_\theta \pi(\boldsymbol{y}, \boldsymbol{\theta}, k)\}} \quad (2.4)$$

$$p(\theta_1 = t_1, \ldots, \theta_k = t_k \mid k, y) = \frac{\pi(\boldsymbol{y}, \boldsymbol{\theta}, k)}{\sum_\theta \pi(\boldsymbol{y}, \boldsymbol{\theta}, k)} \quad (2.5)$$

49

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2 \mid y) = \sum_{\theta,k} \pi(\boldsymbol{y}, \boldsymbol{\theta}, k) N(\hat{\boldsymbol{\beta}}_{\theta,k}, V_{\beta_{\theta,k}}) / \sum_{\theta,k} \pi(\boldsymbol{y}, \boldsymbol{\theta}, k) \qquad (2.6)$$

where

$$\hat{\boldsymbol{\beta}}_{\theta,k} = (X'_{\theta,k} X_{\theta,k} + I_{k+5}(\sigma^2/d))^{-1} X'_{\theta,k} \boldsymbol{y}$$
$$V_{\beta_{\theta,k}} = \sigma^2 (X'_{\theta,k} X_{\theta,k} + I_{k+5}(\sigma^2/d))^{-1}.$$

Equation (2.4) follows from the fact that the joint prior $P(\boldsymbol{\theta}, K = k) \propto p(k)/c_k$, where c_k is the number of elements of $\{t_1, t_2, \ldots, t_k : P(\theta_1 = t_1, \theta_1 = t_2, \ldots \theta_k = t_k \mid K = k) > 0\}$. The identity $p(\boldsymbol{\theta} \mid k, y) = \frac{\pi(\boldsymbol{y}, \boldsymbol{\theta}, k) p(\boldsymbol{\theta})}{p(k|y)}$ yields (2.5) after cancelling the $p(k)/c_k$ terms in the numerator and denominator. Equation (2.6) follows immediately from (2.5) and the fact that $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2 \mid \boldsymbol{\theta}, k, y) \sim N(\hat{\boldsymbol{\beta}}_{\theta,k}, V_{\beta_{\theta,k}})$ (see, e.g., Carlin and Louis 2000, p. 25). The X matrix is $k + 5 \times T$, since it includes the $k + 2$ changepoints and the 3 seasonal effect dummy variables. Thus the marginal posterior of the number of changepoints and the location of these changepoints, (marginally and conditional on the number of changepoints) is available analytically; the marginal distribution of μ_t is obtained by integrating σ^2 out of the weighted mixture of normals given by (2.6), and is easily obtained by simulation by simulating a draw from $\boldsymbol{\theta}, k \mid y$ from (2.4) and (2.5) and then running a short Markov Chain Mote Carlo to obtain a draw of $\boldsymbol{\beta}_k, \boldsymbol{\gamma}, \sigma^2 \mid \boldsymbol{\theta}, k, y$.

The choice of d and of $p(k)$ is critical: for a fixed d, the increase in the likelihood associated with higher-dimension model will tend to lead to larger values of K being selected, while for a fixed $p(k)$, increasingly flat priors (large d) lead to smaller values of K being selected, as inspection of (2.3) reveals. This is a problem inherent in the choice of non-informative priors in model selection or model averaging, since as the prior becomes more dispersed it puts less and less prior weight on reasonable values of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ relative to the whole of $\Re^{k+5}$. Thus we determine $d = d(k)$ from (2.2) such that $P(-c < \beta_k < c) = .95$, which algebra shows is given by $d = \left[c/\Phi^{-1}\left(.5(1 + .95^{1/(k+5)})\right)\right]^2$. Because y_t has been standardized to have mean 0 and variance 1, the β_k can in some sense be viewed relative to a standard normal distribution; we consider $c = 3$ or $c = 10$, with the former being a reasonable "maximum" value for most of the maximum likelihood estimates and the latter being less informative, but not to the point of

50

placing a great deal of the prior distribution at extreme values. We consider two priors for K: $K \sim POI(2)$,and a BIC-type penalty (Schwartz 1978) against larger models given by $p(k) \propto T^{-k/2}$. We restrict our analysis to consider models of up to 5 changepoints, so we truncate these priors so that $P(K > 5) = 0$. (We back-transform to the original scale when reporting our results.)

3. Results

3.1 Number of changepoints

Table 2 indicates that the posterior probability of the number of change-points is sensitive to the choice of prior. Considering models of up to 5 changepoints, the BIC-type prior on changepoints suggests the 2-changepoint or perhaps the 3-changepoint model is most appropriate for all crashes combined, whereas the Poisson prior on changepoints gives credence to 2-,3-, 4- and 5-changepoint models. For nighttime crashes, the BIC-type prior gives the 1-changepoint model the greatest posterior probability, while the Poisson prior weights more toward the 2-changepoint model. For single-vehicle crashes, a 2-changepoint model has substantial posterior probability under all the priors considered, although the BIC prior with a less informative prior regression model places a majority of weight on a 0-changepoint model, while the Poisson priors suggest a 3-changepoint model might be plausible as well.

Because of the sensitivity to the prior, we will conduct all analyses under two choices that tend to reflect the extremes considered: (2.1) $K \sim POI(2)$ and $P(-3 < \beta_k < 3) = .95$, and (2.2) $p(K) \propto T^{-k/2}$ and $P(-10 < \beta_k < 10) = .95$.

We proceed with our analysis by considering first the posterior mean and 95% posterior predictive interval of the mean crash rate at time t $\mu_t \mid y$ and the posterior probability of the changepoint location of the jth changepoint

$$P(\theta_j = t \mid y) = \sum_{k=j}^{5} P(\theta_j = t \mid K = k, y)P(K = k \mid y)$$

for $j = 1, \ldots, 5$. (Since the location of the jth changepoint is predicated on

51

Table 2: Posterior probability of K-changepoint model

K	All Crashes				Night Crashes				SV Crashes			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
0	.00	.00	.00	.01	.00	.00	.00	.03	.00	.02	.14	.56
1	.00	.00	.00	.00	.10	.24	.46	.66	.01	.01	.05	.04
2	.17	.36	.55	.75	.50	.54	.46	.28	.36	.54	.60	.34
3	.37	.34	.33	.20	.25	.17	.07	.03	.36	.28	.17	.05
4	.30	.15	.10	.03	.11	.03	.01	.00	.19	.08	.04	.01
5	.16	.15	.03	.02	.04	.02	.00	.00	.08	.07	.01	.00

(a) $K \sim POI(2)$ and $P(-3 < \beta_k < 3) = .95$, (b) $K \sim POI(2)$ and $P(-10 < \beta_k < 10) = .95$, (c) $p(K) \propto T^{-k/2}$ and $P(-3 < \beta_k < 3) = .95$, and (d) $p(K) \propto T^{-k/2}$ and $P(-10 < \beta_k < 10) = .95$.

the jth changepoint existing, the area under the posterior distribution of the jth changepoint is downweighted to be equal to the posterior probability that $K \geq j$.) Then, since the timing of GDL suggests that GDL effects include one changepoint inside an "effect window" within three months of the introduction of GDL and the subsequent changepoint inside a second "effect window" within 6-18 months of GDL introduction, the posterior probability of the GDL effect is given by

$$P(12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19 \mid y)$$
$$= \sum_{k=2}^{5} \sum_{j=1}^{k-1} P(12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19 \mid K = k, y) P(K = k \mid y).$$

Similarly, conditional of the GDL effect being plausibly present, we can compute the effect of GDL as the posterior mean difference in crash rates at the two GDL-relevant changepoints given by

$$E(\mu_{\theta_{j+1}} - \mu^*_{\theta_{j+1}} \mid 12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19, y) = \frac{A}{B}$$

52

where

$$A = \sum_{k=2}^{5} \sum_{j=1}^{k-1} E(\mu_{\theta_{j+1}} - \mu^*_{\theta_{j+1}} \mid \theta, k, y) \times$$

$$P(12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19 \mid K = k, y) P(K = k \mid y)$$

$$B = \sum_{k=2}^{5} \sum_{j=1}^{k-1} P(12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19 \mid K = k, y) P(K = k \mid y)$$

where $\mu^*_{\theta_{j+1}}$ is computed by assuming $\beta_{j+1} = 0$, that is, by extrapolating from μ_{θ_j} assuming no change in trend due to GDL. Finally, we consider potential rebound effects by fixing $K = 2$ and defining the "rebound effect" as $\beta_3^* = \sum_{j=1}^{3} \beta_j$, the slope of the crash rates after the second changepoint, again conditional on the changepoints being plausibly related to GDL.

3.2 All crashes

To answer the question about whether a change in crash rates among 16-year-olds may be plausibly associated with the introduction of graduated licensing, we consider the posterior mean of the crash rate and the posterior changepoint location given in Figure 1. Examination of Figure 1 indicates that, in contrast to simply determining the number of changepoints, there is little sensitivity to the prior with respect to the shape of the mean function and the presence of changepoints both at the introduction of graduated licensing and approximately one year after its introduction. The three- and four-changepoint models indicated as plausible under the Poisson prior pick up the additional structure of an uptick or an uptick and downtick in crash rates before the introduction of graduated licensing respectively, while the five-changepoint model finds additional change after the one-year-post-GDL changepoint. The posterior probability of any GDL effect on all types of crash among Michigan 16 year-olds is .61 under the Poisson changepoint prior and .76 under the BIC changepoint prior, that is, there is a better than 60% chance that a change in crash rates occured within the two effect windows of three months before or after the introduction of GDL and again within 6-18 months of GDL introduction. The posterior mean of GDL-relatable crash rate changes under the Poisson changepoint prior was -9.6

53

crashes per 1,000 MI 16-year-olds (95% PPI=-2.6,-16.5), a decline of 24.9% (95% PPI=7.0%,38.2%); under the BIC changepoint prior the posterior mean change in crash rates was -11.5(-5.6,-16.7) per 1,000 or a decline of 29.2%(19.5%,37.4%). (See Table 3.)

Positive values of the rebound effect β_3^* are indicative of a trend toward increasing crash rates. The posterior mean of the rebound effect conditional on plausible GDL-related changepoints is 1.1 crashes per 1,000 MI 16-year-olds per year with a 95% PPI of (-.4,2.9) under the Poisson changepoint prior and 1.0 crashes per 1,000 MI 16-year-olds per year with a 95% PPI of (-0.5,2.9) under the BIC changepoint prior.

3.3 Nighttime crashes

Nighttime crashes are of substantive interest since unsupervised driving between midnight and 5 am is no longer allowed for 16 year-olds under Michigan GDL. Figure 2 summarizes the posterior mean crash rates and changepoint locations for nighttime crashes among 16-year-old Michigan drivers. Evidence of a plausible GDL-related effect is weak; it appears that the downward trend among nighttime crashes began more than a year in advance of the introduction of GDL and continued as a linear trend throughout the GDL introduction period, perhaps ending in mid-to-late 1999. $P(12 \leq \theta_j \leq 14, 15 \leq \theta_{j+1} \leq 19 \mid y)$=.06 under the Poission changepoint prior and <.01 under the BIC changepoint prior (see Table 3), indicating that changes in nighttime crash rates among 16-year-old Michigan drivers cannot be linked to GDL with a reasonable degree of certainty.

3.4 Single-vehicle crash

Single vehicle crashes are of particular interest since they tend to be "purer" measures of driver behavior than multiple-vehicle crashes, where culpability may reside with drivers other than the target 16-year-old. There is both considerable sensitivity to the type of prior and "disagreement" as to the number of changepoints in the single-vehicle crash rate, although there is, on average, substantial evidence that single-vehicle crash rates peaked in late 1995 and declined afterwards (see Figure 3). There is weak evidence
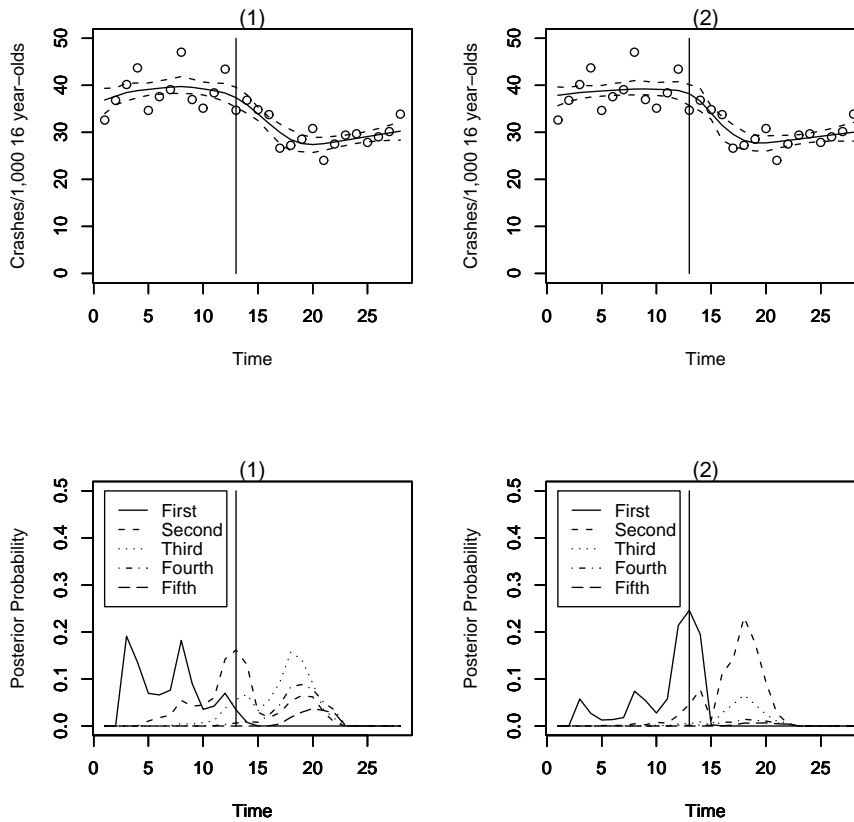
Figure 1. Top row: observed all-crash rates for 16 year-olds and seasonally-adjusted posterior mean rate using Bayesian changepoint model (dotted line=90% PPI for posterior mean). Bottom row: posterior probability of changepoint location for all-crash rates ("first"=first changepoint, "second"=second changepoint, etc.). (1) under $K \sim POI(2)$ and $P(-3 < \beta_k < 3) = .95$ (Poisson changepoint prior) and (2) under $p(k) \propto T^{-k/2}$ and $P(-10 < \beta_k < 10) = .95$ (BIC changepoint prior). Time = quarter-year intervals since January 1994. Vertical line indicates last time point prior to introduction of GDL.
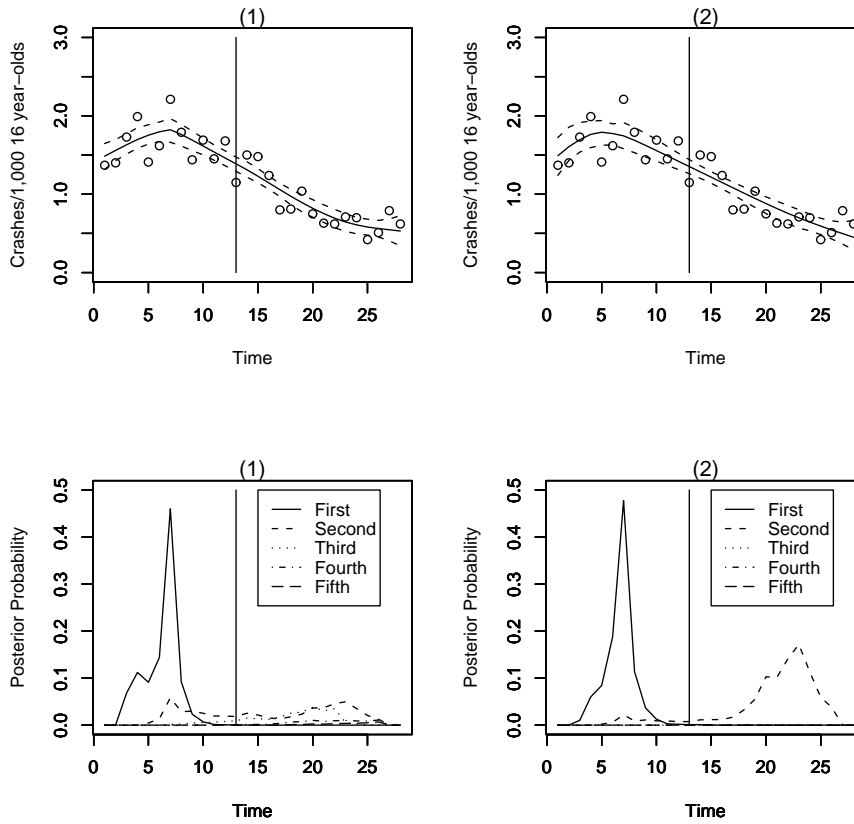
55

Figure 2. Top row: observed nighttime crash rates (midnight-5 am) for 16 year-olds and seasonally-adjusted posterior mean rate using Bayesian changepoint model (dotted line=90% PPI for posterior mean). Bottom row: posterior probability of changepoint location for nighttime crash rates ("first"=first changepoint, "second"=second changepoint, etc.). (1) under $K \sim POI(2)$ and $P(-3 < \beta_k < 3) =$ .95 (Poisson changepoint prior) and (2) under $p(k) \propto T^{-k/2}$ and $P(-10 < \beta_k < 10) = .95$ (BIC changepoint prior). Time = quarter-year intervals since January 1994. Vertical line indicates last time point prior to introduction of GDL.

56

Table 3: Posterior probability and posterior mean of any GDL-related change

		Poisson changepoint prior		BIC changepoint prior
(a)	(b)	(c)	(d)	(e)
All	.61	24.9%(7.0%-38.2%)	.76	29.2%(16.4%-38.6%)
Night	.06	N/A	<.01	N/A
SV	.23	23.9%(-14.8%-46.0%)	.09	N/A

(a)= crash type, (b)= posterior prob. of GDL effect, (c) posterior mean of GDL-related crash rate change, (d)= posterior prob. of GDL effext, and (e)= posterior mean of GDL-related crash rate change.

that this decline accelerated and then slowed under the Poisson changepoint prior (posterior probability of GDL-related changepoints=.23), while under the BIC-type changepoint prior either no trend or a 1995 drop and 1998 rebound are most probable. (posterior probability of GDL-related change-points=.09). The posterior mean of GDL-relatable single-vehicle crash rate change was -2.4 crashes per 1,000 MI 16-year-olds (95% PPI=.8,-4.9), a decline of 23.9% (95% PPI=-14.8%,46.0%) under the Poisson changepoint prior. (See Table 3.) There is no evidence of a rebound effect among single-vehicle crashes: the post-GDL slope under the 2-changepoint model is es-timated as .2(-.3,.8) crashes per 1,000 MI 16-year-olds per year under the Poission changepoint prior. We do not attempt to estimate GDL effects un-der the BIC prior since the posterior probability of relatable effects is so low.

3.5 Model Checking Via Posterior Predictive Distributions

To determine whether the normal model is a reasonable approximation for the error terms, we utilize posterior predictive distributions (PPD) (Gel-man, Meng, and Stern 1996). The PPD p-value represents the probability that the observed statistic (which can be a function of both the data y and the parameters θ) is more extreme than the replicated statistic, conditional on the observed data: $P(T(y, \theta) \leq T(y^{rep}, \theta) \mid y)$. Specifically, for each
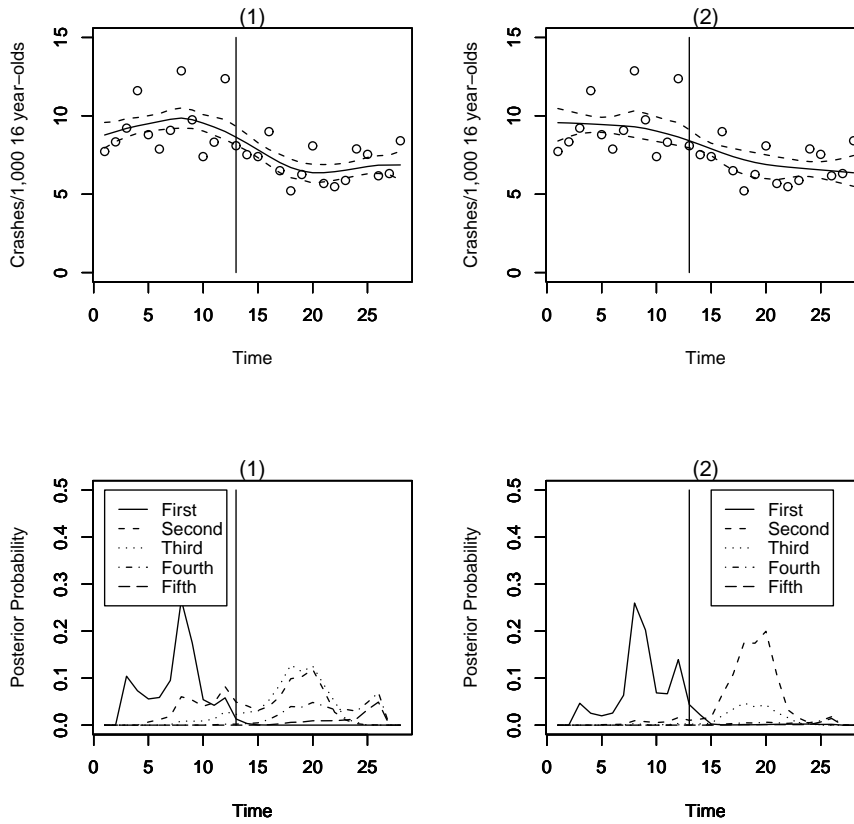
Figure 3. Top row: observed single-vehicle crash rates for 16 year-olds and seasonally-adjusted posterior mean rate using Bayesian changepoint model (dotted line=90% PPI for posterior mean). Bottom row: posterior probability of changepoint location for single-vehicle crash rates ("first"=first changepoint, "second"=second changepoint, etc.). (1) under $K \sim POI(2)$ and $P(-3 < \beta_k < 3) = .95$ (Poisson changepoint prior) and (2) under $p(k) \propto T^{-k/2}$ and $P(-10 < \beta_k < 10) = .95$ (BIC changepoint prior). Time = quarter-year intervals since January 1994. Vertical line indicates last time point prior to introduction of GDL.

58

crash type, we consider the chi-square-type statistic

$$S(y, \eta, \sigma) = \sum_{t=1}^{T} \frac{(y_t - \eta_t)^2}{\sigma^2} \qquad (3.1)$$

where $S^{obs} = S(y^{obs}, \eta^{rep}, \sigma^{2rep})$ is then compared with $S^{rep} = S(y^{rep}, \eta^{rep}, \sigma^{2rep})$ where y^{rep} is generated from $N(\eta^{rep}, \sigma^{2rep})$; consequently $S^{rep} = \sum_t \epsilon_t^2$ for a $\epsilon_t \sim N(0, 1)$. That is, S^{obs} should follow a χ^2_{28} distribution if the model proposed in (2.1) with $K \leq 5$ and normally-distributed error terms is correct. Comparison of 200 values of S^{obs} in a quantile-quantile plot with a χ^2_{28} indicates excellent model fit for all crash types considered, with Kolmogorov-Smirnov goodness-of-fit test p-values of .45, .34, and .86 for all crashes, nighttime crashes, and single vehicle crashes respectively.

4. Summary

A Bayesian linear spline model with an unknown number of knots and knot locations provides a better means than previous approaches of describing time trends in 1994-2000 crash rates among Michigan 16-year-old drivers and in linking these trends to the introduction of graduated driver's licensing in Michigan in April 1997. Averaging across models with differing number of knots and knot locations provides 1) mean estimates that have the flexibility of loess or cubic spline models, while retaining the simple interpretability of linear regression slopes, 2) posterior probabilities that changes in rate trends can be linked to GDL effects, and 3) a method of taking into account the uncertainty as to when GDL effects began and ended together with previous trends in estimation of these effects, if any. We considered models of up to 5 changepoints with two prior distributions on the changepoint: a BIC-type prior that *a priori* downweighted many-changepoint models, and a less restrictive Poisson prior with a mean of 2. The posterior probability of the number of changepoints was sensitive to the choice of prior. Despite the sensitivity of the changepoint count to priors, key posterior distributions such as the location of changepoints and the change in crash rates between GDL-relevant changepoints did not differ greatly under the different priors considered.

When all types of crashes combined are considered, GDL appears more likely than not to have had a substantial effect, dropping crash rates among Michigan 16-year-olds by approximately 25% during the 12-18 months following its implementation. The effect on single vehicle crashes among MI 16 year-olds is less certain, with a posterior probability of .23 that a changepoints appeared within three months of GDL introduction and again 6 to 18 months after GDL introduction under a Poisson prior and only .09 under the BIC prior. Conditional on changes within these windows, the plausible GDL effect on single vehicle crash rates was a decline of 24%. Although the law specifically prohibits all 16 year-olds from driving between midnight and 5 am except under special pre-approved circumstances, it is unclear if GDL has had any impact on late night crashes. While the crash rate for that type of crash declined nearly 50% between 1996 and 1998, this analysis suggests that this drop is part of a decline that began in 1995 or 1996 and its rate was largely unaffected by the implementation of the graduated licensing law April 1, 1997.

Comparing our results with those obtained by a simple comparision of 1996 to 1998 rates (Shope et al. 2001) shows that our point estimate of effect size is similar for all crashes combined, although the confidence interval is wider under the Bayesian changepoint analysis since it incorporates uncertainty in the location of exactly when the GDL effects begin and end. In the case of single-vehicle crashes, the Bayesian changepoint analysis suggests that a previous downward trend might explain a modest part of the change between 1996 and 1998 single-vehicle crash rates. Similarly, a simple analysis of 1996 versus 1998 nighttime crash rates does not capture the fact that GDL caused no significant break in the trend of downward crash rates among this subset of (already uncommon) crashes for 16 year-olds.

The analysis presented here is conservative in that we have restricted "GDL effects" to lie in somewhat narrow windows around the beginning and conclusion of the introduction of graduated drivers licensing. Lack of trends among nighttime and, to a lesser extent, single vehicle crashes that clearly relate to GDL does not eliminate the possibility that GDL had substantial impact on these crash types. All three crash types considered show evidence of declines in rates beginning in late 1995 or early 1996. While this may be due to changes in enforcement patterns or driving habits irrespective of

the introduction of GDL, it may also be evidence of the effects of publicity about the upcoming law that changed parents' restrictions on 16-year-old late night driving well in advance of the law.

Another issue of interest is whether or not crash rates have begun to climb after a leveling-off period, due to non-adherence to the requirements of 50 hours of supervised driving or midnight to 5 am driving restrictions, two requirements which are entirely or largely self-enforced. We considered this possibility by analyzing 2-changepoint models that provided an estimate of linear change following the second changepoint, conditional on the 2 changepoints being GDL-relevant. There is evidence of the beginning of "rebound" in the rate for all crashes combined, with a posterior probability of a rebound effect of .93 under the Poisson changepoint prior and .89 under the BIC changepoint prior. Somewhat weaker evidence was present in the case of single vehicle crashes, where the posterior probability of a rebound was estimated at .79. The total effect thus far, however, appears quite small relative to the crash rate drop associated with GDL, on the order of about 10% of the decline. We have followed subjects for a relatively short period of time since the introduction of graduated driver's licensing; further observation may be required to fully describe any rebound effect, if present.

Acknowledgements

References

Berry D., Hartigan J.A. (1993). A Bayesian analysis for changepoint problems. *Journal of the American Statistical Association*, **88**, 309-319.

Boase P., Tasca L. (1998). Graduated Licensing System Evaluation: Interm Report '98. Safety Policy Branch, Ministry of Transportation of Ontario, Toronto, Ontario.

Box G.E.P., Jenkins G.M. (1976). *Time Series Analysis: Forecasting and Control.* Holden Day, San Francisco.

Carlin B., Gelfand A.E., Smith A.F.M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, **41**, 389-405.

Carlin B.P., Louis T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Edition. Chapman & Hall/CRC, London.

Chipman, H., George E.I., McCulloch, R.E. (2001). Bayesian Treed Models. Technical report: gsbrem.uchicago.edu/papers/treedModels.pdf.

Cleveland, W.S. (1979). Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.

Gelman A., Meng X-L, and Stern H.S. (1996). Posterior Predictive Assessment of Model Fitness via Realized Discrepancies (with discussion), *Statistica Sinica*, **6**, 733-807.

Hastie, T.J., Tibshirani, R.J. (1993). *Generalized Additive Models*, London: Champman and Hall.

Langley J.D., Wagenaar A.C., Begg D.J. (1996). An Evaluation of the New Zealand Graduated Licensing System. *Accident Analysis and Prevention*, **28**, 139-146.

Mayhew D.R., Simpson H.M., des Groseilliers M. (1999). Impact of the Graduated Driver Licensing Program in Nova Scotia. Traffic Injury Research Foundation, Ottawa, Ontario.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

Shope J.T., Molnar L.J., Elliott M.R., Waller P.F. (2001). Graduated Driver Licensing in Michigan: Early Impact on Motor Vehicle Crashes Among 16-Year-Old Drivers. *Journal of the American Medical Association*, **286**, 1593-1598.

Stephens, D.A. (1994). Bayesian Retrospective Multiple-changepoint Identification. *Applied Statistics*, **43**, 159-178.

Ulmer R.G., Preusser D.F., Williams A.F., Ferguson S.A., Farmer C.M. (1999). Effect of the Graduated Licensing Program on Crashes of Teen-age Drivers. Insurance Institute for Highway Safety, Arlington, VA.

US Census Bureau (2001). Populations Estimates Program, Population Division. Available at www.state.mi/dmb/mic.

Michael R. Elliott
Department of Biostatistics and Epidemiology
Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania School of Medicine
612 Blockley Hall, 423 Guardian Drive
Philadelphia, PA 19104-6021 USA
melliott@cceb.upenn.edu

Jean T. Shope
University of Michigan Transportation Research Institute
2901 Baxter Road,
Ann Arbor, Michigan 48109-2150 USA
jshope@umich.edu