Understanding How Offending Prevalence and Frequency Change with Age in the Cambridge Study in Delinquent Development using Bayesian Statistical Models

Abstract

Objectives To provide a detailed understanding of how the prevalence and frequency of offending vary with age in the Cambridge Study in Delinquent Development (CSDD) and to quantify the influence of early childhood risk factors such as high troublesomeness on this variation.

Methods We develop a statistical model for the prevalence and frequency of offending based on the hurdle model and curves called splines that allow smooth variation with age. We use the Bayesian framework to quantify estimation uncertainty. We also test a model that assumes that frequency is constant across all ages.

Results For 346 males from the CSDD for whom the number of offenses at all ages from 10 to 61 are recorded, we found peaks in the prevalence of offending around ages 16 to 18. Whilst there were strong differences in prevalence between males of high troublesomeness and those of lower troublesomeness up to age 45, the level of troublesomeness had a weaker effect on the frequency of offenses, and this lasted only up to age 20. The risk factors of low nonverbal IQ, poor parental supervision and low family income affect how prevalence varies with age in a similar way, but their influence on the variation of frequency with age is considerably weaker. We also provide examples of quantifying the uncertainty associated with estimates of interesting quantities such as variations in offending prevalence across levels of troublesomeness.

Conclusions Our methodology provides a quantified understanding of the effects of risk factors on age-crime curves. Our visualizations allow these to be easily presented and interpreted.

Keywords: Age-crime curves • Hurdle model • Longitudinal study • Troublesomeness • Visualization of prevalence and frequency

Introduction

For a long time there has been much interest amongst criminologists in developing mathematical and statistical models for crime data with the aim of providing a better understanding of criminal behaviour. Part of this effort has focused on the age-crime curve, a crucial topic in developmental and life-course criminology. This paper makes a contribution to learning about and visualizing age-crime curves by presenting Bayesian statistical methodology that can provide a detailed understanding of how offending patterns depend on age and on early risk factors such as troublesomeness.

Criminal Careers, Age-Crime Curves and Our Contribution

In the 1980s there was considerable debate about the value of reseach into criminal careers. A large part of that debate concerned age-crime curves. A simple form of the age-crime curve is a plot of the aggregate number of convictions at each age. It was known even as far back as the 19^{th} Century that the aggregate crime rate peaks in the teenage years. MacLeod et al. (2012), for example, described multiple instances of this. An important question of interest was whether this aggregate crime rate peak was caused by changes with age in prevalence, or in frequency, or in both. The terms prevalence and frequency were clearly defined in Blumstein et al. (1988a), for example: '... *prevalence* (or *participation* ...) refers to the proportion of a population who are active offenders at any given time; it reflects the pervasiveness of offenders in a population. *Frequency* ... refers to the average annual rate at which this sub-group of active offenders commits crimes; it characterizes the intensity, or rate of criminal activity, of *individual* offenders' [p. 3]. Farrington (1986), for example, stated that the age-crime curve generally seems to reflect variations in prevalence rather than frequency.

Gottfredson and Hirschi (1986, 1987), in a criticism of the criminal career approach to criminology, argued that the relationship between age and crime was a law of nature that did not change with social conditions, and that very little was gained by disaggregating this relationship into prevalence and frequency. In response, Blumstein et al. (1988a) explained in detail the construct of a criminal career and discussed carefully why it is critical to investigate whether declines in aggregate offending with age are due to a decline in the number of active offenders, or to a decrease in the frequency of offending by each active offender, or to both. Gottfredson and Hirschi (1988) responded by reaffirming their criticisms of the criminal career approach. Blumstein et al. (1988b) then further replied to these criticisms by explaining how the prevalence-frequency distinction is particularly important when comparing crime in population sub-groups which may have similar aggregate crime levels but quite different prevalences and frequencies. They also explained why it is necessary to take a prevalence-frequency approach when discussing declines in offending with age, in order to understand whether still-active offenders are committing crimes at lower frequencies, or whether prevalence declines with age.

Barnett et al. (1987, 1989) and MacLeod et al. (2012) fitted parsimonious models, assuming constant rates of offending and constant probabilities of career termination at different ages, to criminal career data with offenders stratified into sub-groups designated as ‘frequents’ and ‘occasionals’. The review essay by Piquero et al. (2003) discussed the criminal career paradigm and provided in its Section V an overview of empirical findings generated by criminal career research. As part of this overview, they compared the causes of criminal career dimensions such as onset and frequency. They tentatively concluded that some causes are associated with two or more dimensions, while some are uniquely associated with just one dimension. Also, Smith et al. (1991) studied data from the National Youth Survey (Elliott et al., 1985) and found that, while some variables were related to specific dimensions of delinquency, a core of variables were related to multiple dimensions. In addition, Smith and Brame (1994) found that, while many variables similarly predicted initial and continued involvement in deliquency, other variables predicted only one of these dimensions. Also, Nagin and Farrington (1992) reported that, in the CSDD, low IQ, having criminal parents, and a risk-taking disposition were associated with an initial conviction as well as with subsequent convictions.

Farrington (2019a) compared and contrasted six major developmental and life-course criminology theories in order to explain within-individual changes in offending and antisocial behaviour over time. This work was extended in McGee and Farrington (2019), who described and made key comparisons across the following theories:
integrated cognitive antisocial potential theory (Farrington, 2005), the social developmental model (Cambron et al., 2019), life-course persistent and adolescence-limited antisocial behavior (Moffitt, 1993), the age-graded theory of informal social control (Laub and Sampson, 2003), the situational action theory of crime causation (Wikström, 2006), and interactional theory (Thornberry and Krohn, 2019). McGee and Farrington (2019) also called for empirical testing of these theories.

Rocque et al. (2016) reviewed research on the age-crime curve, referring to the 'great debate in criminology' between Gottfredson and Hirschi (1986, 1988) and Blumstein et al. (1988a, 1988b). Rocque et al. (2016) discussed theoretical explanations and concluded that 'both social and biological factors likely influence the age-crime curve, and policy should be developed accordingly' [their online abstract]. Farrington et al. (2016) suggested that 'the time is ripe to build on simple models of the age-crime curve ... ' and to understand more about how 'particular risk ... factors influence ... criminal careers' [p. 351]. Stander et al. (1989) used Markov chain models to quantify specialization in criminal careers.

In this paper, we follow the suggestion of Farrington et al. (2016) by studying the disaggregated effects of prevalence and frequency on age-crime curves using a statistical technique called the hurdle model introduced by Cragg (1971). The hurdle model assumes that an individual is in the 'offending group' with a certain probability. If the individual is in the offending group, the number of offenses committed is assumed to follow a Poisson distribution that has been modified by being truncated to allow only a non-zero number of offenses. We use smooth curves called splines – rather than specific parametric models – to understand how prevalence and frequency vary with age. We make inferences or learn about the parameters of our model in the Bayesian framework, as this allows us to quantify the uncertainty in unknown quantities in a easily interpretable way. This quantification of uncertainty can be used to assess the effect of childhood risk factors such as troublesomeness on changes in offending with age.

In addition to the statistical model that allows both the probability of offending and the mean number of offenses to vary smoothly with age, we also test the 'constant model' that assumes that the mean number of offenses is totally constant across all ages. We use a state-of-the-art information criterion called WAIC, standing for 'widely applicable information criterion' (Watanabe, 2013), to choose between models. We apply our statistical methodology to data from the Cambridge Study in Delinquent Development (CSDD). In particular, we analyse data on 346 males for whom the number of offenses at each of the 52 ages 10 to 61 is known. We stratify our data by the two levels of the childhood risk factor troublesomeness in order to ascertain the effect of this covariate on relationships with age. We investigate some other risk factors, but the results with troublesomeness are typical, and we discuss them in detail as an example of how our methods can be used to advance knowledge about the age-crime curve. Results for some other risk factors are provided in the Appendix.

The contributions of this paper are as follows:

- We present Bayesian statistical methodology that provides a detailed understanding of how the probability of offending and the mean number of offenses vary with age;
- We show how we can determine whether the mean number of offenses depends smoothly on age, or is totally constant;
- We explain how to quantify the influence of early childhood risk factors on the age dependence of the probability of offending and the mean number of offenses, and we illustrate this using the high troublesomeness risk factor.
- We provide easy to interpret visualizations of our prevalence and frequency age-crime results.

The Hurdle Model

The hurdle model was introduced by Cragg (1971) and the version that we use appears in Welsh et al. (1996). Our discussion is based on Section 5.6 of Stan Development Team (2022a), which also covers the related class of zero-inflated models (Lambert, 1992). Estévez-Soto et al. (2021) presented a recent use of the hurdle model in criminology. To clarify this model let us assume, for the moment, that our data are the total number of offenses $y_1, \ldots, y_n$ committed by a sample of n males, where $y_i \in \{0, 1, 2, \ldots\}$, $i = 1, \ldots, n$. As illustrated in Fig. 1, the i-th male is assumed to be either in the offending group with probability p, $0 \leq p \leq 1$, or not in the offending group with probability $1 - p$. If he is not in the offending group, he commits no offense and $y_i = 0$. If he is in the offending group, he commits a non-zero number of offenses y_i, where y_i is assumed to follow a truncated Poisson distribution (with no probability mass at $y_i = 0$) with parameter $\lambda > 0$:

$$\Pr[y_i \,|\, \text{offending group}, \lambda] = C \, \frac{e^{-\lambda} \, \lambda^{y_i}}{y_i!},$$

in which $C = 1/(1 - e^{-\lambda})$ ensures that the probabilities sum to 1:

$$\sum_{y_i=1}^{\infty} \Pr[y_i \,|\, \text{offending group}, \lambda] = 1.$$

The notation '$|$' means 'conditional on' or 'given that'. It follows from this that the mean value of y_i for the offending group takes the form

$$\mathrm{E}[y_i \,|\, \text{offending group}, \lambda] = \frac{\lambda}{1 - e^{-\lambda}}.$$

We shall refer to this model as the hurdle model with parameters p and λ.

Fig. 1 about here.

From Fig. 1 it follows that

$$\Pr[y_i \,|\, p, \lambda] = \begin{cases} 1 - p & \text{if } y_i = 0 \\ \\ p \times C \, \dfrac{e^{-\lambda} \, \lambda^{y_i}}{y_i!} & \text{if } y_i \in \{1, 2, \ldots\}. \end{cases} \tag{1}$$

The unknown model parameters p and λ have to be estimated from the data $y_1, \ldots, y_n$.

The Bayesian Framework for Estimating Model Parameters

A standard approach to estimating model parameters is to write down the 'likelihood' L and then to maximize L over the parameters to obtain maximum likelihood estimates. In the case of the hurdle model depicted in Fig. 1, the likelihood L takes the form

$$L(\underbrace{y_1, \ldots, y_n}_{\text{data}} \,|\, \underbrace{p, \lambda}_{\text{parameters}}) = \prod_{i=1}^{n} \Pr[y_i \,|\, p, \lambda],$$

in which $\Pr[y_i \,|\, p, \lambda]$ is defined in Eq. (1). Approximate 95% confidence intervals can be obtained for p and λ using L. Confidence intervals have a 'frequent sampling' interpretation, and so this way of learning about model parameters from data is referred to as 'frequentist statistical inference'; see Section 6.2 of Williams (2001). The 'Bayesian approach' to statistical inference (Bayes, 1763; Laplace, 1785, 1810) has become popular in recent years, partly because of the increasing availability of computing power; see Gelman et al. (2014) and Efron and Hastie (2016) for excellent modern coverages, and Stander et al. (2018) for an illustrative example based on survival analysis.

In the Bayesian approach, inference is based on the distribution of the model parameters given the data, rather than on the distribution of the data given the parameters represented by the likelihood $L\,(y_1,\ldots,y_n \,|\, p,\lambda)$. In particular, in the Bayesian approach we work with $\pi(p,\lambda \,|\, y_1,\ldots,y_n)$, that is the 'posterior probability density function' of the parameters p and λ given the data. Many statisticians feel that it is more natural to work with the distribution of the parameters given the data, rather than the distribution of the data given the parameters, because we are indeed 'given the data'. Under standard assumptions, it can be shown using Bayes' Theorem that

$$\pi(\underbrace{p,\lambda}_{\text{parameters}} \,|\, \underbrace{y_1,\ldots,y_n}_{\text{data}}) \propto L\,(y_1,\ldots,y_n \,|\, p,\lambda) \times \pi(p) \times \pi(\lambda),$$

in which the probability density functions $\pi(p)$ and $\pi(\lambda)$ quantify our beliefs about p and λ before seeing the data and, therefore, are referred to as 'prior probability density functions'.

Nowadays, it is common to use simulation methods, rather than mathematical analyses, to understand the posterior probability density function. In particular, computer programs such as `BUGS` (Lunn et al., 2013) or `Stan` (Stan Development Team, 2022a; McElreath, 2020) make the task of sampling parameter values according to the posterior probability density function routine. The `R` (`R` Core Team, 2022) packages `rstanarm` (Goodrich et al., 2020) and `brms` (Bürkner, 2017) provide simple interfaces to `Stan` that allow a vast range of statistical models to be easily fitted in the Bayesian framework; Gelman et al. (2021) provide an excellent book-length discussion. One important advantage of summarizing the posterior probability density function by simulating values from it is that we can easily quantify the uncertainty associated with our inferences.

Bayesian Statistical Inference in Criminology

Recent examples of the use of Bayesian statistical inference in criminology include Blattenberger et al. (2010), Anwar and Loughran (2011), Levine and Block (2011), Kreager and Matsueda (2014), Aljumily (2017), Hodges (2018, 2019), and Dennison et al. (2020). Marchant et al. (2018) emphasized the advantages of the Bayesian approach for quantifying the uncertainty associated with estimation and for updating inferences as new data are collected. Kaimi et al. (2010), Marchant et al. (2018) and Mahfoud et al. (2021) discussed spatial or spatio-temporal modelling, as did Vicente et al. (2021), who used univariate and bivariate splines and employed the INLA algorithm (Rue et al., 2009) to derive inferences.

Cubic Splines

An important ingredient of our model for the CSDD data is the cubic spline, which is discussed in detail in Hastie et al. (2001), James et al. (2013) and Wood (2017), for example. We base our implementation on Kharratzadeh (2017). The general mathematical form of a straight line is $y = \beta_0 + \beta_1 x$, while the equation of a cubic curve is $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. A cubic spline is a continuous function (that is, a function with no gaps) defined over a range of x values. The range of x values is divided into adjacent sections and the cubic spline is required to be a cubic curve in each section. All these cubic curves fit together in a smooth way at the section boundaries which are called knots. It turns out that a cubic spline can be expressed as $y = a_0 + a_1 B_1(x) + a_2 B_2(x) + \cdots + a_N B_N(x)$, where $B_1(x), \ldots, B_N(x)$ are referred to as B-spline basis functions, and each of the parameters $a_0, a_1, \ldots, a_N$ can take any value. The value of N depends on the number of knots. These basis functions $B_1(x), \ldots, B_N(x)$ can be evaluated using the `bs` function of `R`'s `spline` package (`R` Core Team, 2022). We shall use the general notation $f(x \mid a_0, a_1, \ldots, a_N)$ to represent a spline: $f(x \mid a_0, a_1, \ldots, a_N) = a_0 + a_1 B_1(x) + a_2 B_2(x) + \cdots + a_N B_N(x)$.

Method

The Cambridge Study in Delinquent Development (CSDD)

Our analyses are based on the CSDD, which is a prospective longitudinal survey of 411 London males. Farrington et al. (2021) presented the most up-to-date summary of the CSDD and provide a detailed description of who is in the cohort. They also discuss the key findings of the study and its main strengths and weaknesses. Data collection began in 1961–62 when most of the boys were aged 8–9 years, and continued up to 2017. These males comprised a complete population of males of that age living in a working-class area of South London at that time. Information is now available about all offenses committed up to age 61.99 years. CSDD is almost unique in its long follow-up period. The only comparable long-term follow-up study of a large community sample is Le Blanc (2021); see his Table 5.4 [p. 141]. The results of the CSDD have been described in many other publications including Farrington (1995, 2003, 2019b, 2019c, 2020, 2021), Farrington and West (1981, 1990), Farrington et al. (2009, 2013), Piquero et al. (2007), West (1969, 1982), and West and Farrington (1973, 1977). We worked with data on 346 males for whom the number of 'standard list' (more serious) offenses at each of the 52 ages from 10 (the minimum age of criminal responsibility in England and Wales) to 61 were known. For 65 of the original 411 males, the number of offenses was not recorded at all 52 ages because of death or emigration.

Of the 346 males with complete records, 198 had no recorded offenses, while 148 had one or more recorded offenses; corresponding figures for the 65 males with incomplete records are 35 and 30. Therefore, out of the complete sample of 411 males, there were 178 with one or more offenses. Offenses had to be committed on different days to be counted, so that each offense arose from a separate incident. Lesser crimes such as minor traffic infractions and simple drunkenness were excluded; see Farrington (2019b) for more details.

No corrections were made for periods of imprisonment because these periods were generally quite short; for example, average sentences and average time served are much less in England and Wales than in the USA (see Farrington et al., 2004). The usual time served was two-thirds of the sentence up to 1991 and one-half after 1991, due to changes introduced by the Criminal Justice Act 1991. The CSDD records specified sentence length, and the above fractions were used to estimate the actual time served. Of the 178 males in the complete sample who were convicted, 45 (representing just over 25%) were imprisoned at some stage. Their average total time served was only 1.37 years. Sixteen males served up to 6 months, 14 served 7–12 months, 9 served over one year up to two years, and only 6 served over two years.

Fig. 2 provides plots of some data summaries. The proportion of the 346 males who are in the offending group at each age is shown in the top graph of Fig. 2. The bottom graph shows the number of offenses committed by each male in the offending group at each age, together with the mean number of offenses that they commit at each age. The proportion of males in the offending group generally increases until age 17, after which there is a slow, but noisy decline. The pattern for the number of offenses committed is less clear. The modelling approach that we discuss in the next section provides us with a much better understanding of the underlying dependence between the mean number of offenses committed and age.

Fig. 2 about here.

Many covariates were measured for the CSDD males at different ages, and these are discussed in a codebook prepared by Farrington (1999). To illustrate our methods, we considered the childhood risk factor of troublesomeness that was assessed at ages 8 to 10 based on teacher and peer ratings of who gets into trouble most.

Troublesomeness has two levels: (1) high for 75 males and (2) lower for 271 males. Data summaries for each level of troublesomeness are shown in Fig. 3.

We obtained quite similar results for some other risk factors, namely low nonverbal IQ, poor parental supervision and low family income, and these are given in the Appendix.

The data y_{ij} from the CSDD that we will now analyze are the number of offenses committed by Male$_i$, $i = 1, \ldots, n = 346$, at age j, $j = 10, \ldots, 61$ years, together with the early childhood risk factor high troublesomeness for each male. We now present what we refer to as a Bayesian Generalized Additive Hurdle Model for these data, and also discuss a simpler model.

A Bayesian Generalized Additive Hurdle Model and a Simpler Model

We assume that our data y_{ij} come from a hurdle model with parameters p_j and λ_j, which themselves are modelled as functions of age j, $j = 10, \ldots, 61$ years. The parameter p_j is the probability of offending at age j, or prevalence. As probabilities take values between 0 and 1, it is usual to model $\log\{p_j/(1-p_j)\}$ (the inverse-logistic or logit function of p_j) rather than p_j. The inverse-logistic function of p_j can take any real value and its use is familiar from the generalized linear model (Nelder and Wedderburn, 1972). If $\log\{p_j/(1-p_j)\} = \eta$, then we can easily recover p_j as $p_j = e^{\eta}/(1+e^{\eta})$. We model the inverse-logistic function of p_j as a smooth function of age j using a spline f:

$$\log\left(\frac{p_j}{1-p_j}\right) = f(j \,|\, a_0, a_1, \ldots, a_N),\ j = 10, \ldots, 61. \tag{2}$$

This ensures that our model for p_j is sufficiently flexible to explain the peak in the relationship between the probability of offending and age.

As the parameter λ_j takes positive values, it is usual to model $\log(\lambda_j)$ instead of λ_j, again following the generalized linear model, as $\log(\lambda_j)$ can take any value. If $\log(\lambda_j) = \xi$, we can recover λ_j as $\lambda_j = e^{\xi}$. We consider two models for $\log(\lambda_j)$. Our first model is similar to Eq. (2) as it is based on a spline g:

$$\log(\lambda_j) = g(j \,|\, b_0, b_1, \ldots, b_N),\ j = 10, \ldots, 61, \tag{3}$$

in which $b_0, b_1, \ldots, b_N$ are the coefficients for g such that $g(x \,|\, b_0, b_1, \ldots, b_N) = b_0 + b_1 B_1(x) + b_2 B_2(x) + \cdots + b_N B_N(x)$. We refer to this as the 'spline model' for λ_j. As the model defined through Eq. (2) and Eq. (3) involves transformations that are used in the generalized linear model and splines that can appear in additive models, we call the whole model a Bayesian Generalized Additive Hurdle Model.

We also work with a model that assumes that $\log(\lambda_j)$ (and hence λ_j) is constant for all $j = 10, \ldots, 61$:

$$\log(\lambda_j) = \xi, \; j = 10, \ldots, 61. \tag{4}$$

We refer to this as the 'constant model' for λ_j.

The quantities $a_0, a_1, \ldots, a_N$, $b_0, b_1, \ldots, b_N$, and ξ are the unknown parameters of our models. As we are working in the Bayesian framework, we need to specify prior probability density functions for these parameters. For $a_0, a_1, \ldots, a_N$ and for $b_0, b_1, \ldots, b_N$ we follow Kharratzadeh (2017), who explains that, if $a_0, a_1, \ldots, a_N$ take similar values, then the resulting spline f has limited local variability (that is, is not wiggly). Therefore, as Kharratzadeh (2017) suggests, we adopt a random-walk prior distribution for $a_0, a_1, \ldots, a_N$, the step sizes of which are controlled by a standard deviation parameter σ_p. This means that, before seeing any data, we expect the parameters $a_0, a_1, \ldots, a_N$ to take similar values and the spline f not to be wiggly. We further assume, before seeing any data, that σ_p can take a wide range of values by assigning a Cauchy prior probability density function to it. Our experience is that results are quite robust to the specification of this Cauchy distribution. A similar approach is used to specify the random-walk prior distribution for the parameters $b_0, b_1, \ldots, b_N$ that define the spline g, with the step sizes being controlled by σ_λ. We have found that specifying the model using random-walk priors means that results tend to be quite robust to the choice of the number of knots. If the number of knots is increased, there is more scope for the spline to have local variability, but this is counter-balanced by the random-walk having smaller step size. This simplifies the choice of the number of knots. We take the prior probability density functions for ξ to be normal with a mean of 0 and a large standard deviation of 100.

Drawing Inferences and Quantifying Uncertainty

We wrote `Stan` code to make Bayesian statistical inferences about all unknown model parameters. We ran our code in `R` using the `rstan` package (Stan Development Team, 2022b). Although the hurdle model can be fitted in two stages, for convenience our code draws inferences about all model parameters together. `Stan` code is based on blocks called `data`, `parameters`, `transformed parameters`, `model` and `generated quantities`. The `generated quantities` block allows us to define and work with quantities that are transformations of other model quantities. This means that we can easily draw inferences about the mean number of offenses $\mu_j = \lambda_j/(1 - e^{-\lambda_j})$ that a male commits at age j, as well as about the probability p_j that a male is in the offending group at age j, $j = 10, \ldots, 61$ years. We use p_j^{median} and μ_j^{median} as our one number summaries or point estimates of p_j and μ_j, where p_j^{median} , for example, has the property that $\Pr\left[p_j \le p_j^{\text{median}} \,|\, \text{data}\right] = 0.5$. Plotting p_j^{median} and μ_j^{median} against age j provides us with a visualization of how the probability of offending and the mean number of offenses change with age.

`Stan` also allows us to quantify the uncertainty associated with our estimation by reporting 95% 'credible intervals'. Credible intervals are similar to confidence intervals, but have an easier interpretation in terms of probability. For example, if the interval (lower, higher) is a 95% credible interval for a parameter θ, say, then

$$\Pr\left[\text{lower} \le \theta \le \text{upper} \,|\, \text{data}\right] = 0.95.$$

Many of the covariates recorded on the CSDD males have only two values. We consider troublesomeness, which has levels (1) high troublesomeness (approximately the 'worse' quarter) and (2) lower troublesomeness. It is therefore of interest to see whether there is a difference between the values of p_j and the values of μ_j at each covariate level. Let p_{1j}, $j = 10, \ldots, 61$, be the values of p_j for covariate level 1 and p_{2j}

be the values of p_j for covariate level 2. `Stan` allows us to quantify the uncertainty associated with differences $p_{1j} - p_{2j}$ by computing 95% credible intervals for these quantities. We can visualize these credible intervals for $p_{1j} - p_{2j}$ as ribbons on our plots. If the credible interval at age j intersects with zero, we may conclude that there is no real difference between p_{1j} and p_{2j} at age j; otherwise we may conclude that there is a real difference. Similar considerations apply to μ_{1j}, μ_{2j} and $\mu_{1j} - \mu_{2j}$.

Model Choice

It is usual in the Bayesian framework to choose between models based on information criteria or related quantities. These Bayesian model choice quantities have similar interpretations to Akaike's information criterion or AIC (Akaike, 1974), which is used in frequentist statistical inference mentioned above. AIC is a type of penalized 'badness-of-fit' statistic and so the model with the smallest value of AIC is preferred; see Venables and Ripley (2002), for example. The deviance information criterion or DIC (Spiegelhalter et al., 2002, 2014) is a Bayesian version of AIC and has been used very successfully for model comparisons. Nowadays, WAIC (Watanabe, 2013) is considered to be a generally better, but more computationally expensive, alternative to DIC, as it provides a more fully Bayesian approach for assessing the out-of-sample predictive performance of a model. Mahfoud et al. (2021) used WAIC to compare the performance of their models in the context of residential burglary. Stander et al. (2019) provided an example in which the values of DIC and WAIC are very similar. The leave-one-out cross-validation score or LOO (Vehtari et al., 2017) is an alternative way to assess the predictive performance of a model and is discussed in detail in Gelman et al. (2014). The `R` package `loo` (Vehtari et al., 2020) supplies functions to calculate WAIC and LOO. We used WAIC to help us choose between our models. We also calculated LOO and found that it yielded almost identical results to WAIC.

Results

We present the results of our analyses using visualizations that we believe are easily interpretable and that we produced with `R`'s `ggplot2` package (Wickham, 2016). In Fig. 3 we show, as curves for each level (1 and 2) of troublesomeness, the median of the posterior distribution of the prevalences or probabilities p_{1j} and p_{2j} that a male is in the offending group at age j, $j = 10, \ldots, 61$ years, and of the frequencies or mean numbers of offenses μ_{1j} and μ_{2j} that he commits. We also provide 95% credible intervals for the differences $p_{1j} - p_{2j}$ and $\mu_{1j} - \mu_{2j}$ as ribbons to quantify the uncertainty associated with these quantities.

Fig. 3 about here.

For the probability of offending, the curve for the high troublesomeness group lies above the curve for the lower troublesomeness group, although there is, of course, uncertainty associated with these estimates. For both groups the peak of the probability curves occurs around ages 16 to 18. This is consistent with the discussion in MacLeod et al. (2012) and Farrington (1986). We do not add credible intervals to these plots to avoid making them too complicated. Instead, we present credible intervals for the differences $p_{1j} - p_{2j}$, $j = 10, \ldots, 61$ years. These credible intervals quantify our uncertainty about $p_{1j} - p_{2j}$. They do not intersect 0 until around 45 years, indicating that there is a real difference due to the level of troublesomeness over a long time period, with high troublesomeness males being more likely to offend.

We can see from the data summaries shown in Figs. 2 and 3 (and in Figs. 4–6 in the Appendix) that our spline-based curves have smoothed out a lot of the variability in the data. This, together with the difference curves, has led to a much clearer picture of the underlying relationships with age. This is a considerable strength of our method.

WAIC can be used to help us choose between our models. Approximate values of WAIC were 5275 for the spline model Eq. (3) and 5285 for the constant model Eq. (4), meaning that we opt for the spline model.

For the frequency or mean number of offenses, the posterior median levels for the high troublesomeness males lie above the levels for the lower troublesomeness males, but as usual there is uncertainty associated with these estimates. The credible intervals for the differences $\mu_{1j} - \mu_{2j}$ suggest that there may be a real difference due to the level of troublesomeness up to around age 20, with a larger mean number of offenses being committed by high troublesomeness males. This suggests that the effect of high troublesomeness on the frequency of offending is felt over quite a limited time period.

We obtained quite similar results for some other important childhood risk factors: low nonverbal IQ, poor parental supervision and low family income. These are shown in the Appendix. For all three risk factors there is a peak in prevalence around ages 16 to 18 and clear differences in the prevalence curves due to the levels of the risk factors. The effect of these risk factors on the frequency curves is considerably less.

A great advantage of working in the Bayesian framework using `Stan` is that we can find the posterior distribution of quantities of interest by defining them in the `generated quantities` block. For example, if there is interest in understanding how the probability of offending changes between ages 14 to 17 years across levels of a covariate, then the increases $\Delta p_1 = p_{1,17} - p_{1,14}$ and $\Delta p_2 = p_{2,17} - p_{2,14}$ can be included in the `generated quantities` block. For troublesomeness, these quantities have posterior medians of around 0.03 (high troublesomeness) and 0.02 (lower troublesomeness), but the full posterior distrbution is available. We can also estimate event probabilities related to the difference $\Delta p_1 - \Delta p_2$ in these changes between covariate levels such as $\Pr[\Delta p_1 - \Delta p_2 > 0.025 \,|\, \text{data}]$ (posterior median of $\Delta p_1 - \Delta p_2$ is around 0.016; value of the probability is approximately 0.33).

We can perform simlar calculations for changes in the mean number of offenses and their difference such as $\Delta\mu_1 = \mu_{1,30} - \mu_{1,16}$, $\Delta\mu_2 = \mu_{2,30} - \mu_{2,16}$, $\Delta\mu_1 - \Delta\mu_2$ (posterior medians of -0.3, -0.1 and -0.2 for $\Delta\mu_1$, $\Delta\mu_2$ and $\Delta\mu_1 - \Delta\mu_2$; value of the probability $\Pr[\Delta\mu_1 - \Delta\mu_2 < -0.4 \,|\, \text{data}]$ is approximately 0.05).

In addition, we can work with predictions by computing the so-called posterior predictive distribution. For example, for a male in the offending group, we can compute $\Pr[y \,|\, \text{data}]$, where y is number of offenses committed and takes values $1, 2, \ldots$. The posterior predictive distribution takes proper account of the uncertainty associated with parameter estimation. Computation of $\Pr[y \,|\, \text{data}]$ involves sampling from the truncated Poisson distribution, which we do outside `Stan` using the `R` function `rtpois` from the `extraDistr` pacakage (Wolodzko, 2020). We can then report quantities such as

$$\Pr\left[\text{Two or more offenses are committed} \,|\, \text{data}\right] = \Pr[y \geq 2 \,|\, \text{data}]$$

for males aged 18 in the offending group of high or lower troublesomeness (the approximate value of $\Pr[y \geq 2 \,|\, \text{data}]$ is 0.39 or 0.24 for males of high or lower troublesomeness).

The ability to compute quantities such as those given in the above examples is important because it means that we have a powerful tool that can help researchers and policy-makers to obtain answers to many questions of interest. This tool can also quantify the uncertainty associated with the answers.

Conclusions and Further Work

In this paper, we have presented a statistical methodology based on the hurdle model that uses splines to allow model parameters to depend smoothly on age. We also considered a model that was totally constant across the age range. Our modeling results provide insights about how the prevalence of offending (the probability of being in the offending group) and the frequency of offending (the mean number of offenses per offender) change with age. We worked in the Bayesian framework, an advantage of which is that the uncertainty associated with the estimation of unknown quantities can be quantified in a straightforward and easily interpretable way. We applied our methodology to data from the CSDD stratified by levels of troublesomeness, and we provided visualizations of our results. We found that there was a peak in the probability of offending around ages 16 to 18. There were clear differences between males with high troublesomeness and those with lower troublesomeness in the prevalence of offending up to around age 45. Differences in the frequency of offending were less strong and lasted up to around age 20.

The demonstration of different effects of different explanatory risk factors at different ages has important implications for criminological theories. Major criminological theories do not pay sufficient attention to this issue. For example, Moffitt (1993) postulated that cognitive deficits, an under-controlled temperament, hyperactivity, poor parenting, disrupted families, teenage parents, poverty, and low socioeconomic status influence whether a child becomes a life-course-persistent offender, but she did not specify how these factors have different effects at different ages. Similarly, she did not discuss how these factors might have different effects on participation compared with the frequency of offending. Our method, and our results, could assist theorists in making more specific predictions from their theories.

We hope that our research has illustrated some of the advantages of working in the Bayesian framework and has pointed towards useful software tools and techniques.

Other count data generated by the mechanism described in Fig. 1 together with continuous covariates could be modelled using our approach. Possible examples are provided by Wallace et al. (2015), who considered the number of health care institutions in neighborhoods in the context of the release of prisoners, Hester and Hartmann (2017), who analyzed prison terms in months, and Rydberg and Carkin (2017), who worked with the number of adult arrests, together with a range of covariates that can be treated as continuous. Zero-inflated models (Lambert, 1992) should also be considered.

Our findings apply to the cohort of males who grew up from the 1950s to the 2010s. As Farrington et al. (2021) point out, the single cohort design of the CSDD makes it difficult to distinguish between ageing and period effects. They explain that, for example, between ages 14 and 18, the percentage of males who had taken drugs increased from less than 1% to 31%, but that this was probably more influenced by broader social changes during the 1967 to 1971 period rather than reflecting the effect of ageing. Different time periods may, of course, affect participation and frequency in different ways, although the underlying pattern of the relationship between participation and age will probably remain similar. Our method could assist in investigating how participation and frequency vary in different time periods.

Further work could include modifying the model to allow each male to have his own probability of offending and mean number of offenses curves, in the spirit of Farrington (1986). This would have to be done in such a way that the computational cost does not become enormous. One approach could be to use the 'SuperImposition by Translation And Rotation' (SITAR) methodology of Cole et al. (2010), developed in the context of growth curve modeling. We would replace Eq. (2), for example, by

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = f_i(j),$$

where p_{ij} is the probability that Male$_i$ is in the offending group at age j and

$$f_i(j) = \alpha_i + f((j - \beta_i)\, e^{\gamma_i} \mid a_0, a_1, \ldots, a_N),$$

in which α_i, β_i and γ_i are random effects for each male; see also Cole (2020). This is similar to the approach of Baker (1997), who developed models for engine failures based on non-homogeneous Poisson processes with intensity functions $\lambda_i(t) = e^{\beta_{0i}+\beta_{1i}t}$ in which t is time, i indexes engine and β_{0i} and β_{1i} are engine random effects; see also Maiorano (2001) for an application of this random effects approach in criminology. There are issues in drawing inferences from these models when the response variable is discrete and sample sizes are small. Britt (2019) also refers to growth curve models in the context of age and crime.

It would also be interesting to perform similar analyses on data about self-reported offending.

Acknowledgement For funding the CSDD, we are grateful to the Home Office, the Department of Health, the Department of Education, the Rayne Foundation, the Barrow Cadbury Trust, and the Smith-Richardson Foundation. For carrying out criminal record searches, we are grateful to Gwen Gundry in the 1960s and 1970s, Lynda Morley in the 1980s, Sandra Lambert in the 1990s, Debbie Wilson in the 2000s, Owen Thomas in 2011–12 and Lisa Robinson in 2017. We also thank Tim Cole, UCL GOS Institute of Child Health, for introducing us to his work on SITAR models, and Mario Cortina-Borja, Hugh Davies, John Eales, and David Teague for interesting conversations about statistical inference and criminal behavior.

Appendix: Further Analysis of CSDD Risk Factors

Results for Some Other Risk Factors

Fig. 4, 5 and 6 show results similar to Fig. 3 for low nonverbal IQ, poor parental supervision and low family income.

Fig. 4 about here.

Fig. 5 about here.

Fig. 6 about here.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716–723

Aljumily R (2017) Quantitative criminology: Bayesian statistics for measuring the 'Dark Figure' of crime. SSRN Electronic Journal https://doi.org/10.2139/ssrn.2999280

Anwar S, Loughran TA (2011) Testing a Bayesian learning theory of deterrence among serious juvenile offenders. Criminology 49(3): 667–698

Baker J (1997) Random effect models for repairable system reliability. PhD thesis, University of Plymouth

Barnett A, Blumstein A, Farrington DP (1987) Probabilistic models of youthful criminal careers. Criminology 25: 83–107

Barnett A, Blumstein A, Farrington DP (1989) A prospective test of a criminal career model. Criminology 27: 373–388

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society: 330–418. Reprinted with biographical note by Barnard, GA (1958) in Biometrika 45: 293–315

Blattenberger G, Fowles R, Krantz J (2010) Bayesian models to predict the return to prison. Section on Bayesian Statistical Science, JSM 2010 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.6919&rep=rep1&type=pdf

Blumstein A, Cohen J, Farrington DP (1988a) Criminal career research: its value for criminology. Criminology 26: 1–35

Blumstein A, Cohen J, Farrington DP (1988b) Longitudinal and criminal career research: further clarifications. Criminology 26: 57–74

Britt CL (2019) Age and crime. In: Farrington DP, Kazemian L, Piquero AR (eds) The Oxford handbook of developmental and life-course criminology. Oxford University Press, New York, pp 13–33

Bürkner P-C (2017) `brms`: An `R` Package for Bayesian multilevel models using `Stan`. Journal of Statistical Software 80(1): 1–28

Cambron C, Catalano RF, Hawkins, JD (2019) The social development model. In; Farrington DP, Kazemian L, Piquero AR (eds) The Oxford handbook of developmental and life-course criminology. Oxford University Press, New York, pp 224–247

Cole TJ (2020) Tanner's tempo of growth in adolescence: recent SITAR insights with the Harpenden Growth Study and ALSPAC. Annals of Human Biology 47(2): 181–198

Cole TJ, Donaldson MDC, Ben-Shlomo Y (2010) SITAR – a useful instrument for growth curve analysis. International Journal of Epidemiology 39: 1558–1566

Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39(5): 829–844

Dennison S, Besemer K, Low-Choy S (2020) Maternal parenting stress following paternal or close family incarceration: Bayesian model-based profiling using the HILDA longitudinal survey. Journal of Quantitative Criminology 36: 753–778

Efron B, Hastie T (2016) Computer age statistical inference: algorithms, evidence, and data science. Cambridge University Press

Elliott DS, Huizinga D, Ageton SS (1985) Explaining delinquency and drug use. Sage, Beverly Hills, California

Estévez-Soto PR, Johnson SD, Tilley N (2021) Are repeatedly extorted businesses different? A multilevel hurdle model of extortion victimization. Journal of Quantitative Criminology 37: 1115–1157

Farrington DP (1986) Age and crime. In: Tonry M, Morris N (eds) Crime and justice: an annual review of research, vol 7. Chicago University Press, Chicago, pp 189–250

Farrington DP (1995). The development of offending and antisocial behaviour from childhood: Key findings from the Cambridge Study in Delinquent Development. Journal of Child Psychology and Psychiatry 36: 929–964

Farrington DP (1999) Cambridge study in delinquent development [Great Britain], 1961–1981. Inter-university Consortium for Political and Social Research. ICPSR 8488. September 1999

Farrington DP (2003) Key results from the first 40 years of the Cambridge Study in Delinquent Development. In: Thornberry TP, Krohn MD (eds) Taking stock of delinquency: an overview of findings from contemporary longitudinal studies. Kluwer/Plenum, New York, pp 137–183

Farrington DP (2005) The integrated cognitive antisocial potential (ICAP) theory. In: Farrington DP (ed) Integrated developmental and life-course theories of offending: advances in criminological theory. Transaction: New Brunswick, NJ, vol 14, pp 73–92

Farrington DP (2019a) Developmental theories of crime. In: Morgan, RD (ed) The Sage encyclopedia of criminal psychology. Sage, Thousand Oaks, California, pp 367–371

Farrington DP (2019b) The Cambridge Study in Delinquent Development. In: Eaves D, Webster CD, Haque Q, Eaves-Thalken J (eds) Risk rules: a practical guide to structured professional judgment and violence prevention. Pavilion Publishing, Hove, pp 225–233

Farrington DP (2019c) The duration of criminal careers: how many offenders do not desist up to Age 61? Journal of Developmental and Life-Course Criminology 5:

4–21

Farrington DP (2020) Childhood risk factors for criminal career duration: comparisons with prevalence, onset, frequency and recidivism. Crim. Behav. Ment. Health 30: 159–171

Farrington DP (2021) New findings in the Cambridge Study in Delinquent Development. In: Barnes JC, Forde DR (eds) The encyclopedia of research methods in criminology and criminal justice, vol. 1. Wiley, Hoboken, NJ, pp 96–103

Farrington DP, Coid JW, West DJ (2009). The development of offending from age 8 to age 50: recent results from the Cambridge Study in Delinquent Development. Monatsschrift Fur Kriminologie Und Strafrechtsreform (Journal of Criminology and Penal Reform) 92: 160–173

Farrington DP, Jolliffe D, Coid JW (2021) Cohort profile: The Cambridge Study in delinquent development. Journal of Developmental and Life-Course Criminology 7: 278–291

Farrington DP, Langan PA, Tonry M (2004, eds) Cross-national studies in crime and justice. Washington, D.C.: U.S. Bureau of Justice Statistics (NCJ 200988)

Farrington DP, MacLeod JF, Piquero AR (2016) Mathematical models of criminal careers: Deriving and testing quantitative predictions. Journal of Research in Crime and Delinquency 53(3): 336–355

Farrington DP, Piquero AR, Jennings WG (2013) Offending from childhood to late middle age: recent results from the Cambridge Study in Delinquent Development. Springer, New York

Farrington DP, West DJ (1981) The Cambridge Study in Delinquent Development (United Kingdom). In: Mednick SA, Baert AE (eds) Prospective longitudinal research. Oxford University Press, Oxford, pp 137–145

Farrington DP, West DJ (1990). The Cambridge Study in Delinquent Development:

A long-term follow-up of 411 London males. In: Kerner H-J, Kaiser G (eds) Kriminalitat: personlichkeit, lebensgeschichte und verhalten (Criminality: personality, behavior and life history). Springer-Verlag, Berlin, pp 115–138

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) Bayesian Data Analysis, third edition, CRC Press, Boca Raton

Gelman A, Hill J, Vehtari A (2021) Regression and other stories, Cambridge University Press, Cambridge

Goodrich B, Gabry J, Ali I, Brilleman S (2020) `rstanarm`: Bayesian applied regression modeling via `Stan`. `R` package version 2.21.1. `https://mc-stan.org/rstanarm`

Gottfredson M, Hirschi T (1986) The true value of lamba would appear to be zero: An essay on career criminals, criminal careers, selective incapacitation, cohort studies, and related topics. Criminology 24: 213–233

Gottfredson M, Hirschi T (1987) The methodological adequacy of longitudinal research on crime. Criminology 25: 581–614

Gottfredson M, Hirschi T (1988) Science, public policy, and the career paradigm. Criminology 26: 37–55

Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference, and prediction, Springer, New York

Hester R, Hartmann TK (2017) Conditional race disparities in criminal sentencing: A test of the liberation hypothesis from a non-guidelines state. Journal of Quantitative Criminology 33 (1): 77–100. `https://doi.org/10.1007/s10940-016-9283-z`

Hodges H (2018) Towards a Bayesian approach in criminoogy: a case study of risk assessment in youth justice. PhD thesis, University of Swansea

Hodges H (2019) Have we got the right data to solve the problem? Welsh Policy and Politics in Unprecedented Times. Wales Centre for Public Policy.

https://www.wcpp.org.uk/wp-content/uploads/2018/12/Helen-Hodges-Have-we-got-the-pdf

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with applications in R, Springer, New York

Kaimi I, Diggle PJ, Rodrigues A (2010) Analysing the spatio-temporal distribution of crime. MethodsNews Summer 2010, 2–3. http://eprints.ncrm.ac.uk/1307/1/MethodsNews_summer2010.pdf

Kharratzadeh M (2017) Splines in Stan. https://mc-stan.org/users/documentation/case-studies/splines_in_stan.html

Kreager DA , Matsueda RL (2014) Bayesian updating and crime. In: Bruinsma G, Weisburd D (eds) Encyclopedia of criminology and criminal justice. Springer, New York

Lambert D (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34(1): 1–14

Laplace PS (1785) Mémoire sur les formules qui sont fonctions de très grands nombres. In: Mémoires de l'Académie Royal des Sciences

Laplace PS (1810) Mémoire sur les formules qui sont fonctions de très grands nombres et sur leurs applications aux probabilités. In: Mémoires de l'Académie des Sciences de Paris

Laub JH, Sampson RJ (2003) Shared beginnings, divergent lives: delinquent boys to age 70, Cambridge MA: Harvard University Press

Le Blanc M (2021) The development of antisocial behavior and crime, Springer, Cham, Switzerland

Levine N, Block R (2011) Bayesian journey-to-crime estimation: an improvement in geographical profiling methodology. The Professional Geographer 63(2): 213–229

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter DJ (2013) The BUGS book: a practical introduction to Bayesian analysis. CRC Press, Boca Raton

MacLeod JF, Grove PG, Farrington DP (2012) Explaining criminal careers: implications for justice policy, Oxford University Press, Oxford

McElreath E (2020) Statistical rethinking: a Bayesian course with examples in `R` and `Stan`, second edition, CRC Press, Boca Raton

McGee, TR, Farrington DP (2019) Developmental and life-course explanations of offending. Psychology, Crime & Law 25 (6): 609–625

Mahfoud M, Bernasco W, Bhulai S, van der Mei R (2021) Forecasting spatio-temporal variation in residential burglary with the integrated Laplace approximation framework: Effects of crime generators, street networks and prior crimes. Journal of Quantitative Criminology 37: 835–862

Maiorano F (2001) Metodi Bayesiani nell'analisi statistica delle carriere criminali (Bayesian methods in the statistical analysis of criminal careers). Master's Thesis, University of Bologna

Marchant R, Haan S, Clancey G, Cripps S (2018) Applying machine learning to criminology: semi-parametric spatial-demographic Bayesian regression. Security Informatics 7: Article number 1

Moffitt TE (1993) Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. Psychological Review, 100 (4): 674–701

Nagin DS, Farrington DP (1992) The onset and persistence of offending. Criminology 30: 501–523

Nelder J, Wedderburn R (1972) Generalized linear models. Journal of the Royal Statistical Society Series A (General) 135(3): 370–384

Piquero AR, Farrington DP, Blumstein A (2003) The criminal career paradigm. Crime and Justice, 30: 359–506.

Piquero AR, Farrington DP, Blumstein A (2007) Key issues in criminal career research: new analyses of the Cambridge Study in Delinquent Development, Cambridge University Press, Cambridge

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rocque M, Posick C, Hoyle J (2016) Age and crime. In: Jennings WG (ed) The encyclopedia of crime and punishment, first edition, John Wiley & Sons

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society Series B: Statistical Methodology 71: 319–392

Rydberg J, Carkin DM (2017) Utilizing alternate models for analyzing count outcomes. Crime & Delinquency, 63 (1): 61–76

Smith DA, Brame R (1994) On the initiation and continuation of delinquency. Criminology, 32: 607–630

Smith DA, Visher CA, Jarjoura GR (1991) Dimensions of delinquency: exploring the correlates of participation, frequency, and persistence of delinquent behavior. Journal of Research in Crime and Deliquency, 28: 6–32

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society Series B: Statistical Methodology 64: 583–639

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2014) The deviance information criterion: 12 years on. Journal of the Royal Statistical Society Series B: Statistical Methodology 76: 485–493

Stan Development Team (2022a) Stan User's Guide, Version 2.29. https://mc-stan.org

Stan Development Team (2022b). rstan: the R interface to Stan. R package version 2.26.4. https://CRAN.R-project.org/package=rstan

Stander J, Dalla Valle L, Cortina Borja M (2018) A Bayesian survival analysis of a historical dataset: How long do popes live? The American Statistician 72: 368–375

Stander J, Dalla Valle L, Taglioni C, Liseo B, Wade A, Cortina-Borja M (2019) Analysis of paediatric visual acuity using Bayesian copula models with sinh-arcsinh marginal densities. Statistics in Medicine 38: 3421–3443

Stander J, Farrington DP, Hill G, Altham PME (1989) Markov chain analysis and specialization in criminal careers. British Journal of Criminology 29: 317–335

Thornberry TP, Krohn MD (2019) Interactional theory. In: Farrington DP, Kazemian L, Piquero AR (eds) The Oxford handbook of developmental and life-course criminology. Oxford University Press, New York, pp 248–271

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P, Paananen T, Gelman A (2020). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.4.1. https://mc-stan.org/loo/

Vehtari A, Gelman A, Gabry J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing 27: 1413–1432

Venables WN, Ripley BD (2002) Modern applied statistics with S, second edition, Springer, New York

Vicente G, Goicoa T, Ugarte MD (2021) Multivariate Bayesian spatio-temporal P-spline models to analyze crimes against women. Biostatistics. https://doi.org/10.1093/biostatistics/kxab042

Wallace D, Eason JM, Lindsey AM (2015) The influence of incarceration and re-entry on the availability of health care organizations in Arkansas. Health & Justice 3 (3): 1–11

Watanabe S (2013) A widely applicable Bayesian information criterion. Journal of Machine Learning Research 14: 867–897

Welsh AH, Cunningham RB, Donnelly CF, Lindenmayer DB (1996) Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88 (Issues 1–3): 297–308

West DJ (1969) Present conduct and future delinquency, Heinemann, London

West DJ (1982) Delinquency: its roots, careers and prospects, Heinemann, London

West DJ, Farrington DP (1973) Who becomes delinquent? Heinemann, London

West DJ, Farrington DP (1977) The delinquent way of life, Heinemann, London

Wickham H (2016) `ggplot2`: Elegant graphics for data analysis, Springer-Verlag

Wikström P-O,H (2006) Individuals, settings, and acts of crime: Situational mechanisms and the explanation of crime. In: Wikström P-O,H, Sampson RJ (eds) The explanation of crime: context, mechanisms and development, Cambridge University Press, Cambridge

Williams D (2001) Weighing the odds: a course in probability and statistics, Cambridge University Press, Cambridge

Wolodzko, T (2020) `extraDistr`: Additional univariate and multivariate distributions. R package version 1.9.1. `https://CRAN.R-project.org/package=extraDistr`

Wood SN (2017) Generalized additive models: an introduction with `R`, second edition, CRC Press, Boca Raton

Fig. 1: The hurdle model with parameters p and λ

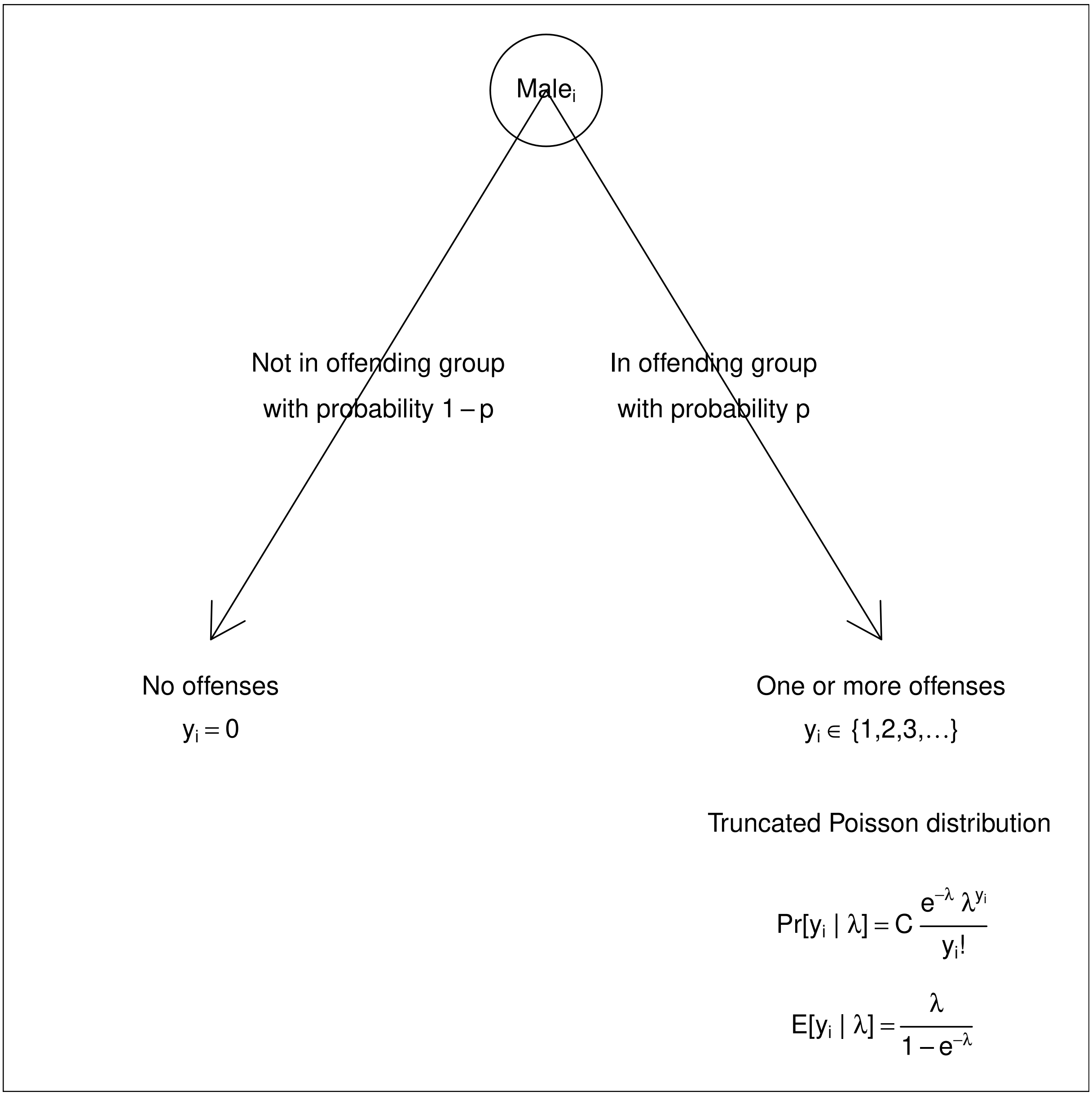
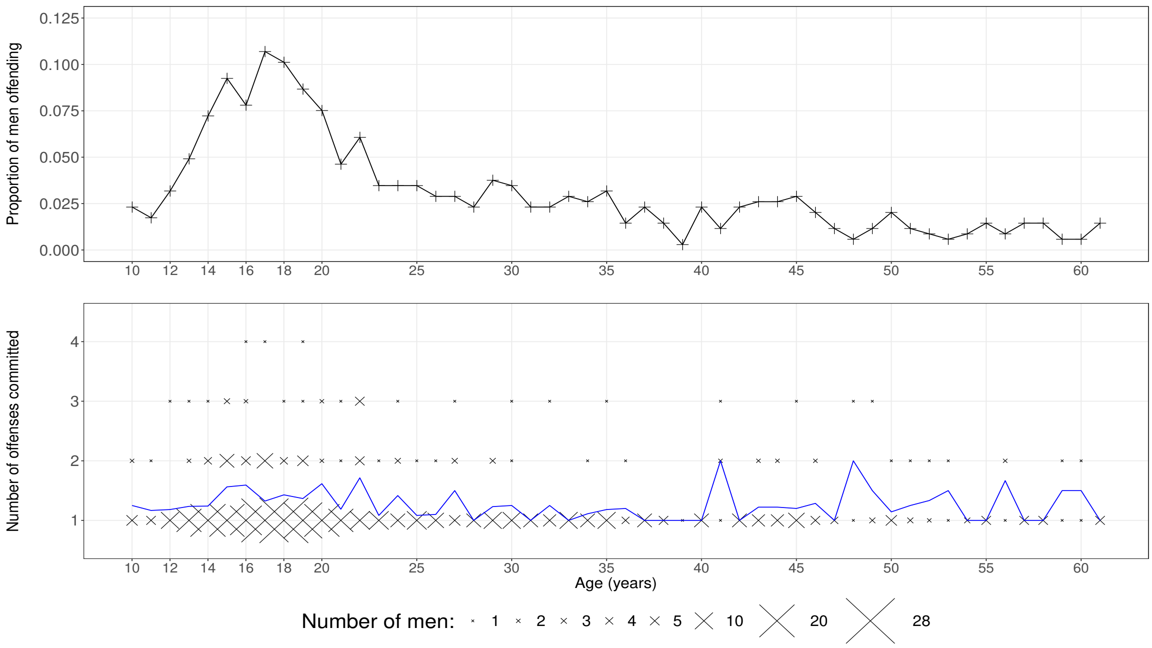
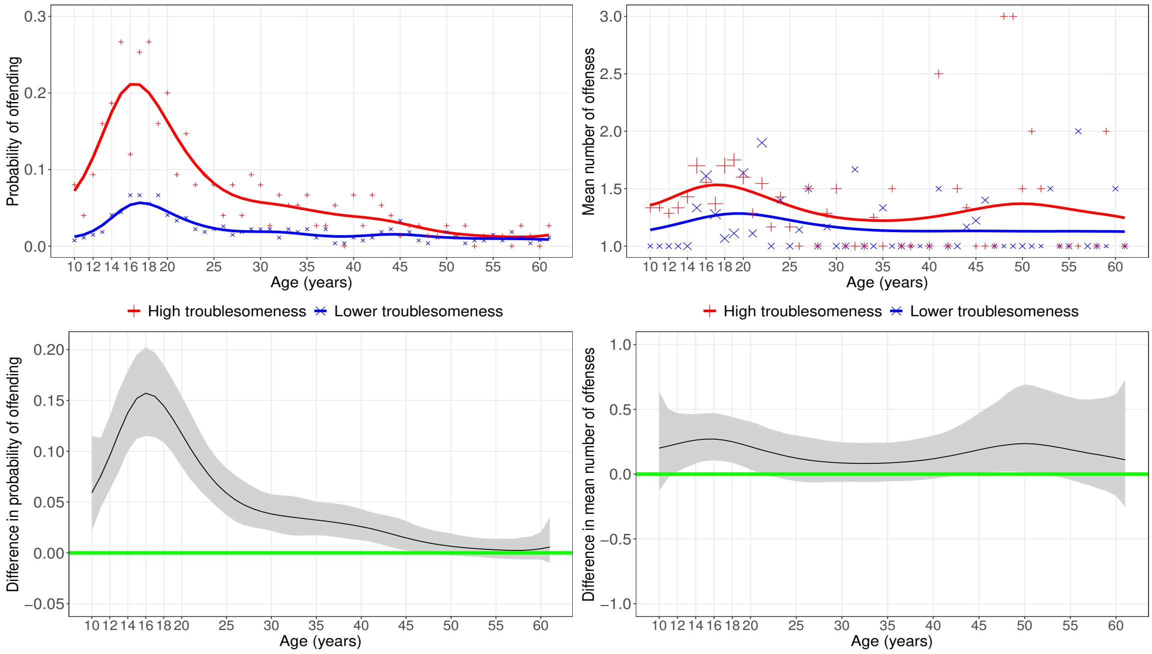


Fig. 2: Proportion of men who offend and the number of offenses that they commit by age.



Notes: Top graph: The proportion of the 346 males who are in the offending group at each age. Bottom graph: the number of offenses committed by each male in the offending group at each age. The number of offenses takes the values 1, 2, 3 and 4 and the size of the plotting symbol depends on the number of men with each value. The average number of offenses committed at each age is shown in blue.

Fig. 3: Age-crime curves for troublesomeness



Notes: Top graphs: posterior medians of prevalences p_{1j} and p_{2j} (left) and frequencies μ_{1j} and μ_{2j} (right) plotted against age j, $j = 10, \ldots, 61$ years, for each level of the early childhood risk factor troublesomeness. Data summaries are shown for each level of troublesomeness. Bottom graphs: posterior medians (curves) and 95% credible intervals (ribbons) for the differences $p_{1j} - p_{2j}$ (left) and $\mu_{1j} - \mu_{2j}$ (right).

Fig. 4: Age-crime curves for low nonverbal IQ

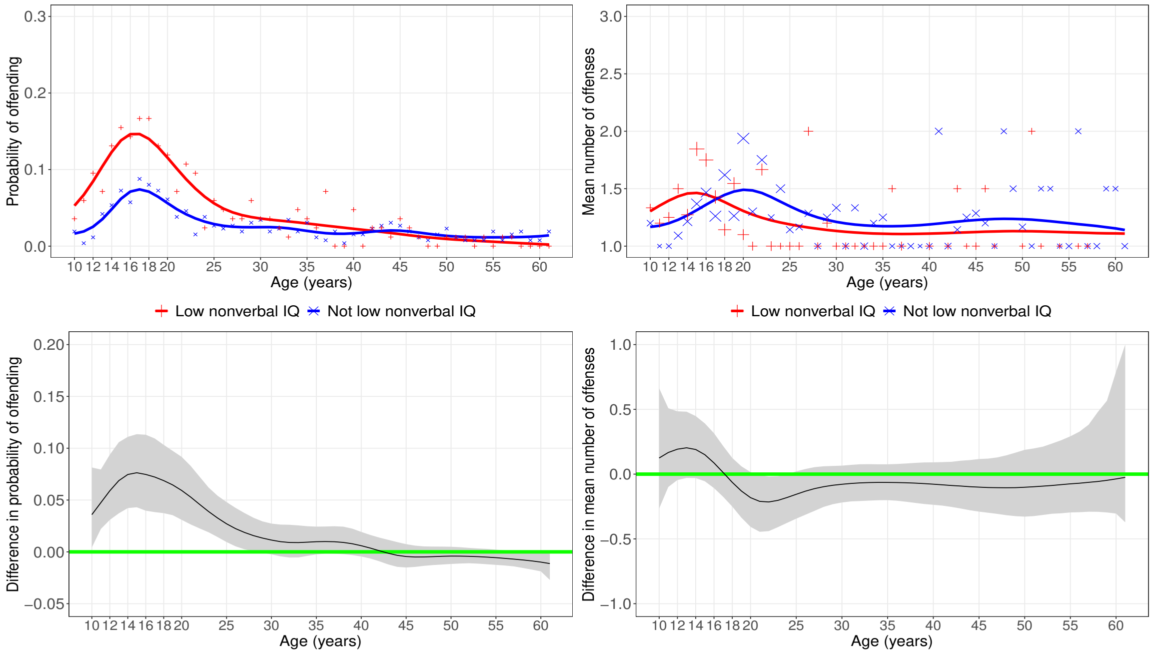


Fig. 5: Age-crime curves for poor parental supervision

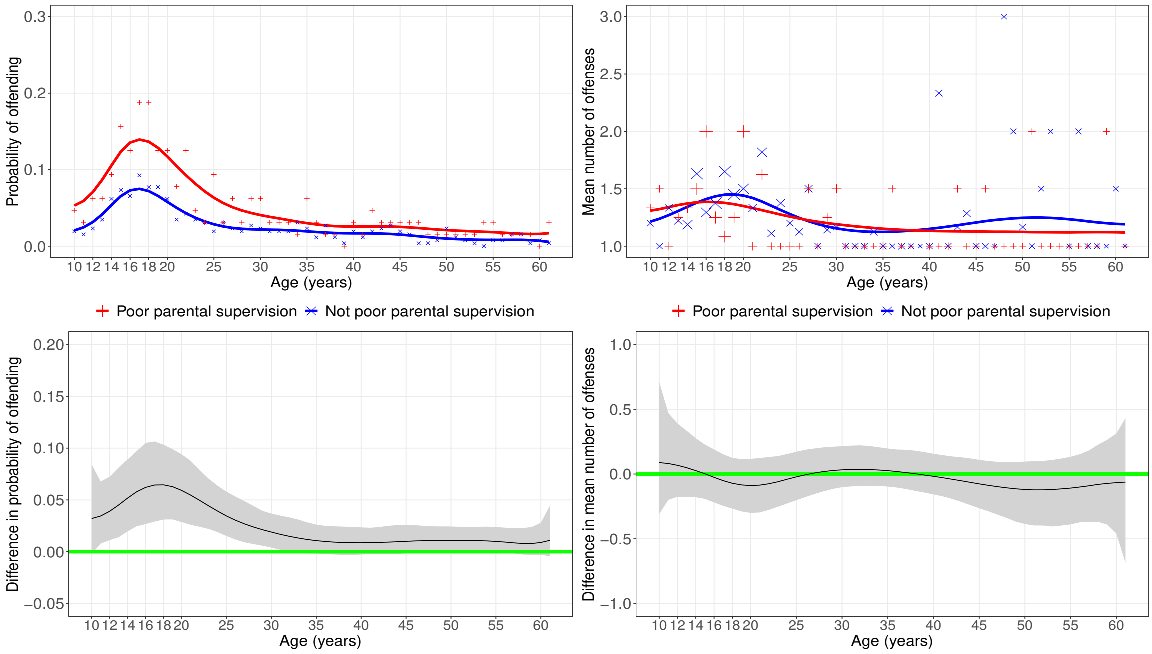


Fig. 6: Age-crime curves for low family income