Accident Analysis and Prevention 123 (2019) 69–78

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A meta-analysis of the crash risk of cannabis-positive drivers in culpability studies—Avoiding interpretational bias

Ole Rogeberg

Frisch Centre, Norway

ARTICLE INFO

Keywords:
Culpability study
Meta-analysis
Cannabis
Crash risks
Bayesian inference

ABSTRACT

Background: Culpability studies, a common study design in the cannabis crash risk literature, typically report odds-ratios (OR) indicating the raised risks of a culpable accident. This parameter is of unclear policy relevance, and is frequently misinterpreted as an estimate of the increased crash risk, a practice that introduces a substantial "interpretational bias".

Methods: A Bayesian statistical model for culpability study counts is developed to provide inference for both culpable and total crash risks, with a hierarchical effect specification to allow for meta-analysis across studies with potentially heterogeneous risk parameter values. The model is assessed in a bootstrap study and applied to data from 13 published culpability studies.

Results: The model outperforms the culpability OR in bootstrap analyses. Used on actual study data, the average increase in crash risk is estimated at 1.28 (1.16–1.40). The pooled increased risk of a culpable crash is estimated as 1.42 (95% credibility interval 1.11–1.75), which is similar to pooled estimates using traditional ORs (1.46, 95% CI: 1.24–1.72). The attributable risk fraction of cannabis impaired driving is estimated to lie below 2% for all but two of the included studies.

Conclusions: Culpability ORs exaggerate risk increases and parameter uncertainty when misinterpreted as total crash ORs. The increased crash risk associated with THC-positive drivers in culpability studies is low.

1. Introduction

The increased crash risks associated with recent cannabis use are receiving increasing scrutiny as "recreational cannabis" is being legalized in Canada and across US states.

Culpability studies are a commonly used observational study design for assessing traffic safety (Kim and Mooney, 2016), and use an odds-ratio to assess whether culpable crashes are associated with driver characteristics in a sample of crash involved drivers. The studies require data on culpability status, which is based on the assessment of trained individuals applying pre-specified criteria to information on the circumstances of the crash. The scorer is (ideally) blinded to the type of driver scored (i.e. positive/negative status). Under the *identifying assumption* that non-culpable drivers can be considered a random sample of the driver population on the road at the time of the crash, the culpability odds ratio (OR) can be interpreted as the relative odds of culpable crashes for positive relative to negative drivers. If confounding is balanced across driver types, the culpability OR can be interpreted as a causal parameter that expresses the impairment-attributable risk increase for culpable crashes. This follows since the driver types under this assumption will have identical average crash risks in the absence of impairment.

The commonly noted drawbacks of the culpability design focus on the quality of the responsibility assessment and the strong assumption that nonculpable crashes can be considered a random sample from the driver population at the time of a crash (Kim and Mooney, 2016). Largely unnoticed is the *interpretational bias* caused by researchers treating culpability ORs as equivalent to crash ORs. Since culpability ORs relates to *culpable* crashes only, the change in total crash risk will necessarily be smaller. The magnitude of the bias can be considerable (see numerical illustration in supporting materials), and the misinterpretation appears widespread in the cannabis crash risk literature: a high profile meta-analysis in the British Medical Journal pooled estimates of culpability and crash OR as though they were exchangeable (Asbridge et al., 2012) while a study in the BMJ used culpability ORs to calculate the attributable risk fraction (Laumon et al., 2005). Despite the issue being highlighted in a later meta-review (Rogeberg and Elvik, 2016a), researchers both responding to (Gjerde and Mørland, 2016) and referring to (Martin et al., 2017) this article persisted in the misinterpretation. The issue appears to be poorly understood, and improved inference models are needed to estimate the increased crash risk from culpability study counts.

A Bayesian model, defined in terms of intuitive structural parameters, is presented that allows for inference regarding both crash risk

E-mail address: ole.rogeberg@frisch.uio.no.

https://doi.org/10.1016/j.aap.2018.11.011
Received 14 August 2018; Received in revised form 17 October 2018; Accepted 12 November 2018
Available online 20 November 2018

and culpable crash risk increases. The model avoids the known upwards bias in odds-ratios computed from small sample and sparse count data (Greenland, 2000; Greenland and Schwartzbaum, 2000; Nemes et al., 2009), as well as the need for a log-normal approximation to compute confidence intervals. To allow for meta-analysis, the model is formulated with a hierarchical culpability risk parameter, improving inference by reducing the impact of random sampling variation on parameter estimates.

The performance of the inference model is demonstrated in a bootstrap exercise using realistic sample sizes and data drawn from 13 published culpability studies on the risk associated with cannabis-positive drivers.

Having confirmed the performance of the Bayesian inference model on bootstrapped studies, crash risk increases are estimated on data from 13 identified culpability studies with counts from no-alcohol subsamples. Estimates of overall crash risk increase are compared to the culpability odds identified using traditional culpability ORs to illustrate the magnitude of the interpretational bias in the literature. Finally, the model is used to infer the attributable risk fraction, which states the percentage change in total crashes that would follow from an elimination of cannabis-positive driving in a road population.

It is important to note that the current meta-analysis has a scope limited to culpability studies only, while a full assessment of the crash risks associated with cannabis would also need to assess the evidence from other designs, e.g., case control studies and experimental driving studies. All model code is written in Stan, a programming language for probabilistic modelling that can be estimated using open-source software in R and Python on a variety of computing platforms. The model code is included in the supporting online materials.

2. Data and methods

2.1. Data

Count data come from 13 published culpability studies of the increased average culpable crash risk associated with cannabis positive drivers. The data consists of no-alcohol counts from all culpability studies identified in a recently published systematic meta-analysis (Rogeberg and Elvik, 2016a), supplemented with counts from three culpability studies published since 2016 (Martin et al., 2017; Li et al., 2017; Romano et al., 2017). Some of the data counts are directly reported in the studies, some can be inferred from the reported data and results, and some was provided directly from the researchers involved. The sources and details are provided in the supplementary information.

From each included study, we use the four counts used by the odds-ratio estimator: The number of culpable and non-culpable individuals who were respectively positive (THC levels above the study's threshold level) and negative. The studies and the data used, along with descriptive information (country, crash type), are shown in Table 1.

Table 2
Underlying population shares expressed using model parameters.

	Negative	Positive
Culpable	$(1 - s_e) \times s_c$	$\alpha \times s_e \times s_c$
Nonculpable	$(1 - s_e) \times (1 - s_c)$	$s_e \times (1 - s_c)$

2.2. Inference models

The standard approach to analyzing culpability studies involves the use of an odds-ratio $\widetilde{\theta}$. Using $+/-$ in the superscript to distinguish between positive and negative drivers, this estimator can be stated as

$$\widetilde{\theta} = \frac{culp^{+}/nonculp^{+}}{culp^{-}/nonculp^{-}}$$

where *culp* and *nonculp* refers to counts of drivers assessed as culpable and nonculpable.

In large sample studies without sparse cells, this estimator will be approximately lognormally distributed with standard errors

$$se = \sqrt{\frac{1}{culp^{+}} + \frac{1}{culp^{-}} + \frac{1}{nonculp^{+}} + \frac{1}{nonculp^{-}}}$$

This is commonly used to calculate confidence intervals irrespective of sample size, though the OR estimator has a known upwards bias in small sample sparse count samples (Greenland, 2000; Greenland and Schwartzbaum, 2000; Nemes et al., 2009).

More fundamentally, the problem with $\widetilde{\theta}$ is the fact that it measures the increased risk of *culpable* crashes rather than of *all crashes* – the risk increase parameter that is typically of interest.

To provide inference regarding *overall* crash risk increases based on culpability count data we formulate a statistical model in terms of three underlying parameters:

1 The true share of drivers on the road who are positive and negative on some substance or risk factor (ideally in a sample conditioned on scoring negative on other substances). We denote the positive share s_e (for exposed).
2 The baseline culpability probability amongst negative drivers (denoted s_c), and
3 the relative risk of having a culpable accident given that you are positive (denoted α). This is the parameter estimated by a culpability OR.

Using these parameters we can express the expected relative share of drivers in each of the four cells, and model the data as a multinomial draw with sampling probabilities proportional to these (Table 2). Intuitively, the table states that the positive share of nonculpable drivers equals the positive share in the driver population (due to the identifying

Table 1
Culpability studies included.

Study	Country	Crash type	Culp +	Culp-	Nonculp +	Nonculp-
Terhune (1982)	USA	Injury	4	94	4	157
Williams et al. (1985)	USA	Fatal	10	55	9	23
Terhune et al. (1992)	USA	Fatal	11	541	8	258
Longo et al. (2000)	Australia	Injury	21	996	23	891
Lowenstein and Koziol-McLain (2001)	USA	Injury	4	114	6	126
Drummer et al. (2004)	Australia	Fatal	51	1214	5	376
Laumon et al. (2005)	France	Fatal	319	4386	131	3585
Soderstrom et al. (2005)	USA	Injury	126	980	59	540
Bédard et al. (2007)	USA	Fatal	1106	18405	541	12491
Poulsen et al. (2014)	New Zealand	Fatal	74	403	18	128
Li et al. (2017)	USA	Fatal	910	9663	716	12595
	France	Fatal	122	1686	44	1410
Romano et al. (2017)	USA	Fatal	64	1398	37	1085

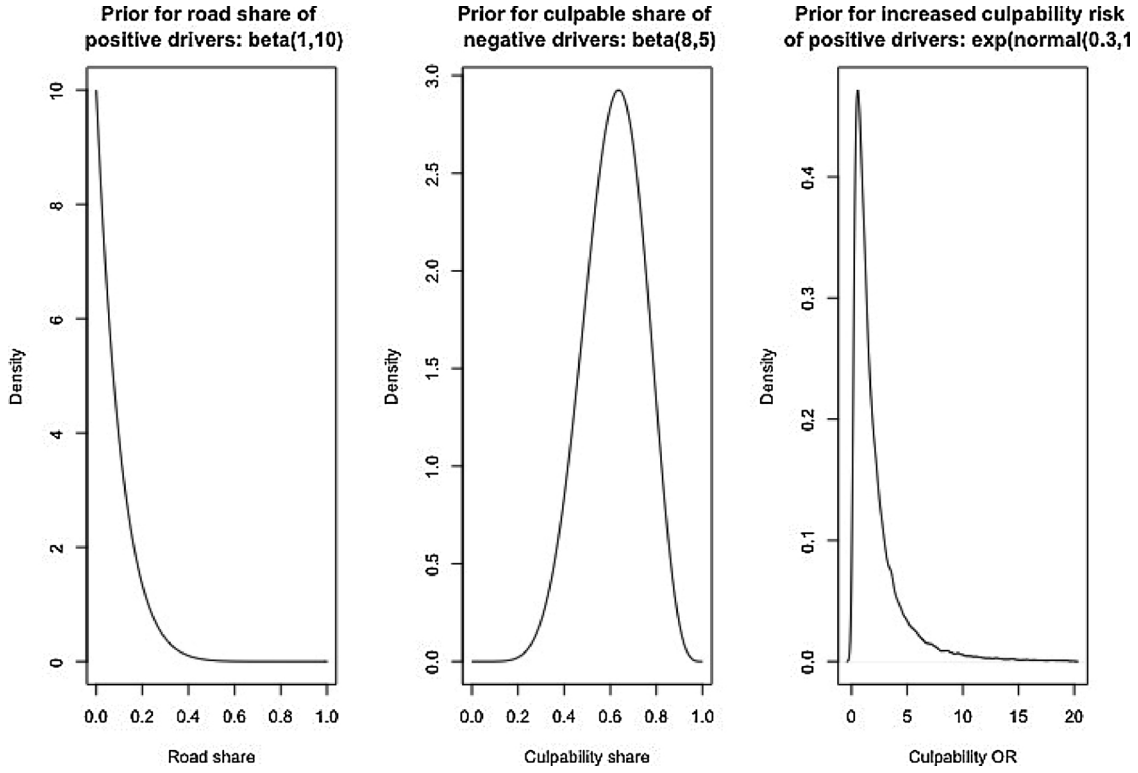

Fig. 1. Prior distributions for model parameters.

assumption); among negative drivers the culpable share is equal to the baseline culpability share; amongst culpable positive drivers, the numbers are altered by the relative risk parameter α. When α differs from one, the cell expressions will fail to sum to 1, requiring a normalization to express the probabilities.

The benefit of this approach is that it allows for statistical inference on any magnitude of interest that can be expressed in terms of the structural parameters. The total increase in crashes for the positive drivers due to α is found by summing the two positive driver cells (an expression we can simplify to $(\alpha - 1) \times s_e \times s_c + s_e$) and dividing this by the sum they would have in the absence of impairing effects (which simplifies to s_e). This gives us the expression for the *relative crash risk increase* of positive drivers (for culpable *and* nonculpable crashes) as $(1 + (\alpha - 1) \times s_c)$.

Similarly, we can get an expression for the attributable risk factor, the percentage change to the total crash rate that would occur under a causal interpretation of α if no drivers were positive. The four cells would then sum to 1, while they now sum to $1 + (\alpha - 1) \times s_e \times s_c$. The impairment attributable share of crashes as a share of all crashes can therefore be written as

$$\frac{(\alpha - 1) \times s_e \times s_c}{1 + (\alpha - 1) \times s_e \times s_c}$$

To estimate the model we use Bayesian maximum likelihood and the probabilistic programming language Stan (Carpenter et al., 2016; Stan Development Team, 2016) with weakly informative priors on the underlying parameters. Estimation conditions on observed study counts, and allows us to draw a representative sample of parameter values from the posterior distribution of the model. These draws tell us how the "beliefs" embedded in the prior are optimally updated in light of the observed data. To perform inference on a parameter or a combination of parameters, we simply assess the distribution of this parameter (or parameter combination) across the posterior samples. This marginalizes across the uncertainty we have regarding the true values of all other parameters, giving us the probability that the parameter will take various values conditional on the imposed model structure, the assigned priors, and the observed data.

The priors for the share of positive drivers on the road and for the culpable share of the negative drivers are both specified using the beta-distribution. The increased culpability risk of positive drivers is given a lognormal prior. The specific parameter values used will depend on context, and we choose values that weakly reflect the typically low prevalence of cannabis positive drivers on the road and the typically high culpability shares found for drug negative drivers, while allowing for large but not extreme and implausible risk increases (Fig. 1).

Note that the prevalence prior is on a sample conditioned on scoring negative on alcohol and other drugs. This conditional prevalence will necessarily be higher than the actual on-the-road prevalence in the total driver population. If, e.g., 6% of drivers score positive for cannabis-use alone and 40% score positive on alcohol or other drugs (with or without cannabis), then the conditional prevalence would be 0.06/0.6 = 10%. For this reason, we let the prior assign non-negligible probability to prevalence values up to about 30%.

2.3. Assessing methods

Performance of the inference model is assessed using a bootstrap methodology: The counts from each of the 13 culpability studies constitutes a bootstrap source sample, from which we draw 500 new samples of size identical to the original study. Each sample is randomly drawn with replacement from the study's observed counts, and both estimation approaches are applied to each of the bootstrap samples. Since the traditional culpability OR is undefined when one of the cell counts is 0, bootstrap samples with zero counts in any cells are discarded.

Under the identifying assumption of culpability studies that nonculpable drivers are a random draw from the driver population, we can view the nonculpable counts in the bootstrap source sample as representative of the relevant driver population. This defines a "true" relative risk of all car crashes for positive drivers equal to

$$\theta = \frac{(culp^{+} + nonculp +)/nonculp^{+}}{(culp^{-} + nonculp^{-})/nonculp^{-}}$$

For each of the studies, we then assess:

- **Sampling distribution of the point estimate.** For each of the simulated samples we extract the median under the posterior from the Bayesian model. We display the distribution of these point estimates across the bootstrap samples along with a vertical line indicating the average estimate across the samples and the true relative risk. For comparison, we include the distribution of culpability ORs estimated using the traditional odds-ratio estimator.
- **Sampling distribution of the confidence/credibility interval widths.** For each of the simulated samples we calculate the width, measured in risk increase percentage points, of a 95% Bayesian credibility interval. Conditional on the model and the prior, this interval should have a 95% probability of including the true parameter value. For comparison we also plot the width distribution of the 95% confidence intervals of the traditional culpability OR estimator. These express a range that should cover the true parameter value 95% of the time in repeatedly sampled data from the same population.
- **Calibration test of confidence intervals.** For each of the simulated samples we calculate a set of credibility intervals for different probabilities (5%, 10%, …, 95%) and assess whether each of these contains the true parameter value. This allows us to draw a calibration curve that shows the percentage of bootstrap samples in which the different credibility intervals actually contained the true parameter value. We compare this to the analogous calibration curve for the confidence intervals of the standard culpability OR. This allows us to assess whether a $p\%$ confidence or credibility interval typically contains the true parameter value in $p\%$ of the bootstrapped samples.

2.4. Meta-analytic approaches

Sampling variation generates substantial variation in effect estimates when samples are small, but the underlying (true) effect sizes that are estimated may themselves differ when data are drawn from different populations. In studies of crash risk increase from cannabis, for instance, the actual effect in different study populations would differ across studies if:

- The impairing effects of THC increases the risk of crashes in specific types of road or traffic conditions, and these are more common in some of the traffic systems sampled.
- Individuals driving with THC in their blood differ in their baseline (unimpaired) crash risk across studies due to selection into use and into impaired driving, and this confounding is stronger in some of the driver populations sampled.
- The average level of THC in the THC-positive driver population differs across regions and periods, leading to different average doses, impairment and risk.
- The scoring used to determine culpability differs across studies, making the non-culpable drivers more representative of the underlying driver population in some studies than others.

While the baseline model performs independent inference on each study, we extend the model with a hierarchical effect specification to estimate the *distribution* of the culpability risk increase parameter across studies. We assume a prior for the across-study heterogeneity that implies that 95% of studies will have effect sizes that range from half to double the mean culpability risk increase. The hierarchical effect priors (and the posterior after estimation) are displayed in the results section below.

To test this meta-analytic model's performance, we estimate it using 10 bootleg samples drawn from two bootleg source samples with sparse counts – one where the implied risk increase in the bootstrap source data is low (Williams et al., 1985) and one where the implied risk increase is high (Drummer et al., 2004). We compare the statistical inference possible when each of the 10 bootstrap samples from a study is assessed separately (using the baseline model) and collectively (using the hierarchical model), comparing both to the true parameter values implied by each of the two bootstrap source samples.

2.5. Crash risk inference – meta-analysis

While the assessment of the inference model compared estimates from resampled study counts to the "true" effect as defined by the source data, we now view the source data as being itself a random draw from some unobserved underlying population with unknown risk parameters.

We assess the risks associated with cannabis in culpability study data by four methods:

1. Single study estimates
   a The baseline Bayesian model, which estimates each study in isolation from the others.
   b The traditional culpability OR
2. Meta-analysis
   a The hierarchical Bayesian model.
   b Random Effects Meta-analysis of the traditional culpability odds-ratios and their associated CIs.

The results from the Random Effects Meta-analysis is presented as a forest plot with a pooled effect estimate and compared to the hierarchical effect estimates for the culpability risk increase from the Bayesian model. To assess the magnitude of the interpretational bias that comes from misinterpreting culpability ORs as "all crash" ORs, we compare the Bayesian estimates of culpability risk to that of all crash. Finally, we present results for the attributable risk fraction based on the hierarchical Bayesian model to assess the overall effect of cannabis impairment on road traffic crashes.

The hierarchical specification allows for study level differences in the underlying risk parameter, though it is worth noting that this variation may itself be related to study level factors. Crash severity, exposure measure (e.g., blood, saliva, urine, self-report) and threshold, country, time period etc. may all systematically affect the size (and interpretation) of the underlying risk parameter. In principle, this can be assessed by adding further hierarchical effects, representing e.g., country or time-period differences, to the model. As these factors operate at the study level, however, the number of "observations" is substantially reduced, increasing the risk of data mining and specification search unless a principled approach is followed.

2.6. Software

All analyses were done within RStudio, running R (version 3.4.2) on a Macintosh computer. Culpability ORs and confidence intervals were calculated using standard odds ratios and the lognormal approximation. The meta-analysis was performed using the DerSimonian-Laird estimator of the Metafor R-package (Viechtbauer et al., 2010). The Bayesian models were specified in the Stan programming language and estimated using Hamiltonian Monte Carlo and the NUTS (No U-Turn) sampler using the Rstan package (version 2.16.2). Three chains with 10,000 iterations each were used for each model estimation, of which the first 3000 samples from each chain tune the model estimation and were discarded from the posterior draws used for inference. Full model code for the inference model is available in the supporting materials.

Plots were made using the ggplot2 R-package (Wickham, 2016) and the ggridges add-on (version 0.4.1) from the CRAN repository.

3. Results – model assessment

3.1. Assessing the baseline model

The estimated increase in the risk of a culpable crash is largely the

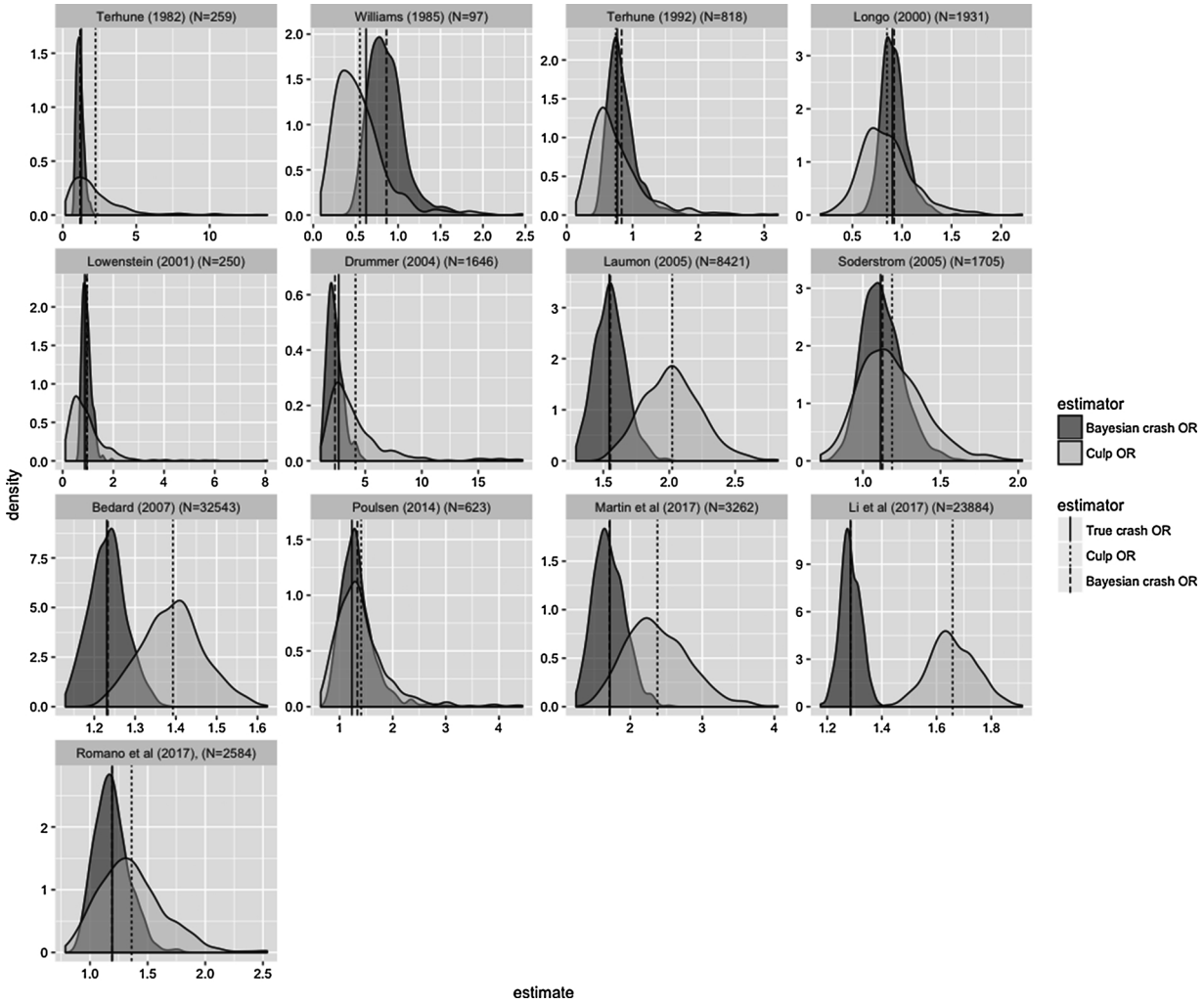

Fig. 2. Culpability and crash OR - distribution of point estimates. Results from 500 bootstrapped samples drawn from – and of equal size to – the original data. Culpability ORs and Bayesian crash ORs are assessed relative to the true crash OR in the bootstrapped data.

same whether we use the Bayesian model or culpability ORs: Averaged across multiple bootstrapped samples the point estimate from both methods average close to the true value, confidence/credibility intervals are similar, and the confidence/credibility intervals are well calibrated (figures S1-S3 in the supporting materials). The priors of the Bayesian model help avoid some of the most extreme and implausible point estimates suggested by the culpability OR, but the differences are minor. For one small sample study with an extreme share of positive drivers amongst the nonculpable (Williams et al., 1985), the nonculpable counts are insufficient to shift the prior, resulting in exaggerated risk increase estimates. In this study, the road share parameter of the bootstrap source sample is 28%, but this is based on 9 positive and 23 negative nonculpable drivers.

Turning to the overall crash risk increase and the interpretational bias, we next compare the estimated increase in total car crash risk to the traditional culpability OR. Comparing the distribution of point estimates across resampled study counts, the culpability ORs show substantially more variation and are typically centered away from the true crash risk increase, illustrating the interpretational bias issue (Fig. 2). The issue is especially prominent for the large sample studies, where the average culpability OR is about twice the underlying total crash risk increase. The Bayesian model recovers the underlying risk parameter both in cases where the risk increase is low (as in Longo et al. (2000)) and high (as in Martin et al (2017)).The exception is Williams et al. (1985), where the sample size used (n = 97) is very small and contains insufficient information to shift the prior sufficiently towards the parameter value implied by the bootstrap source data.

Turning to the inferential uncertainty, the 95% confidence intervals are substantially larger than the 95% credibility intervals of the Bayesian model (Fig. 3). The Bayesian credibility intervals for the crash increase are well-calibrated except for the Williams study with its small sample, while the confidence intervals of the culpability ORs fail to encompass the true risk increase value in 4 of the larger sample studies (Fig. 4).

3.2. Assessing the hierarchical model

To assess the hierarchical (meta-analytic) Bayesian model we use two studies with sparse counts, Drummer et al. (2004) and Williams et al. (1985). These bootstrap source samples implied some of the most extreme risk parameter values, with implied relative risks of all crashes at 2.65 and 0.62 respectively. We now draw 10 samples from each of these studies, and use the hierarchical and baseline models on each set of 10 samples. The hierarchical specification of the culpability risk parameter improves inference by pooling across studies, as it correctly infers that there is little evidence in the data that the underlying risks differ across the samples (Fig. 5). Note that the plotted distributions now show statistical uncertainty, while earlier plots showed the distribution of point estimates and uncertainty intervals across multiple bootleg replications.

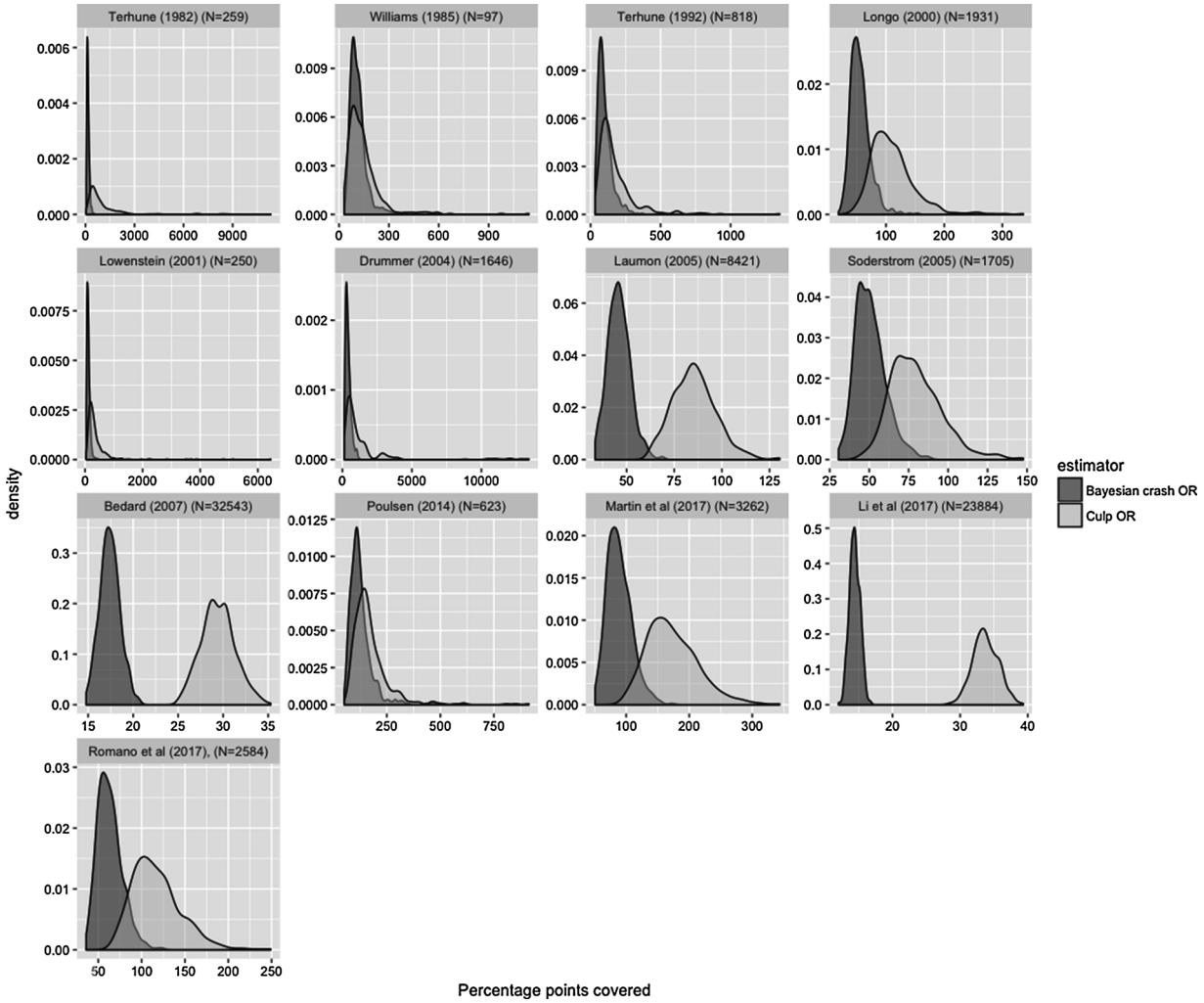

Fig. 3. Width distribution of confidence intervals and credibility intervals. Results from 500 bootstrapped samples drawn from – and of equal size to – the original data.

Note also that the partially pooled relative risk inference contains the true risk parameter, although the median value (which we used as a point estimate when testing the baseline model) is above the true value for the low-risk Williams samples and below for the high-risk Drummer sample. This reflects the prior for the road prevalence, combined with the sparse counts identifying the extreme values of this parameter in the data. Since we do not impose a hierarchical specification on the road share, the typically sparse counts identifying this parameter in the samples is insufficient to credibly shift the prior all the way towards the true value. We can see this by repeating the exercise with two other studies where the sparse counts issue for the road share is less extreme: relative risk estimates are now centered directly on the true effect, while the non-pooling Baseline model suggests an excessive degree of across-study effect heterogeneity (see figure S4 in the supporting info).

4. Results - assessing the average crash risk of cannabis

Independent inference on each of the 13 sets of study counts, using culpability ORs and the Bayesian baseline model, finds that the Bayesian estimates of total crash increase tend to be both lower and more precisely estimated than indicated by (misinterpreted) culpability ORs (Fig. 6), with particularly large discrepancies in large sample studies.

Pooling the evidence across studies using a random-effects meta-analysis of culpability odds-ratios, the 13 data samples give a pooled odds ratio of 1.46 (95% CI: 1.24–1.72) (see forest plot in the supporting information, S5), and a forest plot fails to suggest positive publication bias (supporting information, figure S6). The random-effects pooled effect is essentially the same as the pooled culpability risk increase in the Bayesian meta-analytic model, which averages at 1.42 (95% credibility interval 1.11–1.75).

Comparing the study level estimates of total and culpable crash risk change from the meta-analytic Bayesian model, we see how changes to total crash risk are systematically smaller and more precisely estimated (Table 3). The importance of the interpretational bias can be seen in Table 3 by noting that the estimated average total crash risk increase at the study level is below the lower bound of the random effects culpability OR meta-analytic pooled effect for 8 of the 13 studies. Taking the average crash risk across the 13 study samples for each of the posterior draws, we can extract the mean and the 2.5% and 97.5% quantile values to characterize the average crash risk increase overall, giving a point estimate of 1.28 (1.16–1.40).

Relative to the baseline Bayesian model, the hierarchical specification reduces the dispersion in estimates across studies (Supporting materials, figure S7). In particular, the model suggests that sampling variability is the most plausible explanation for the substantial risk-reducing culpability RRs found using data from Williams et al. (1985), Terhune et al. (1992), Longo et al. (2000) and Lowenstein et al. (2001), as well as the large risk increase estimated from the data of Drummer et al. (2004).

Finally, the posterior distribution for the attributable risk fraction, which gives the percentage change in car crashes involving no-alcohol drivers that would be expected if all cannabis use ceased (Fig. 7), finds

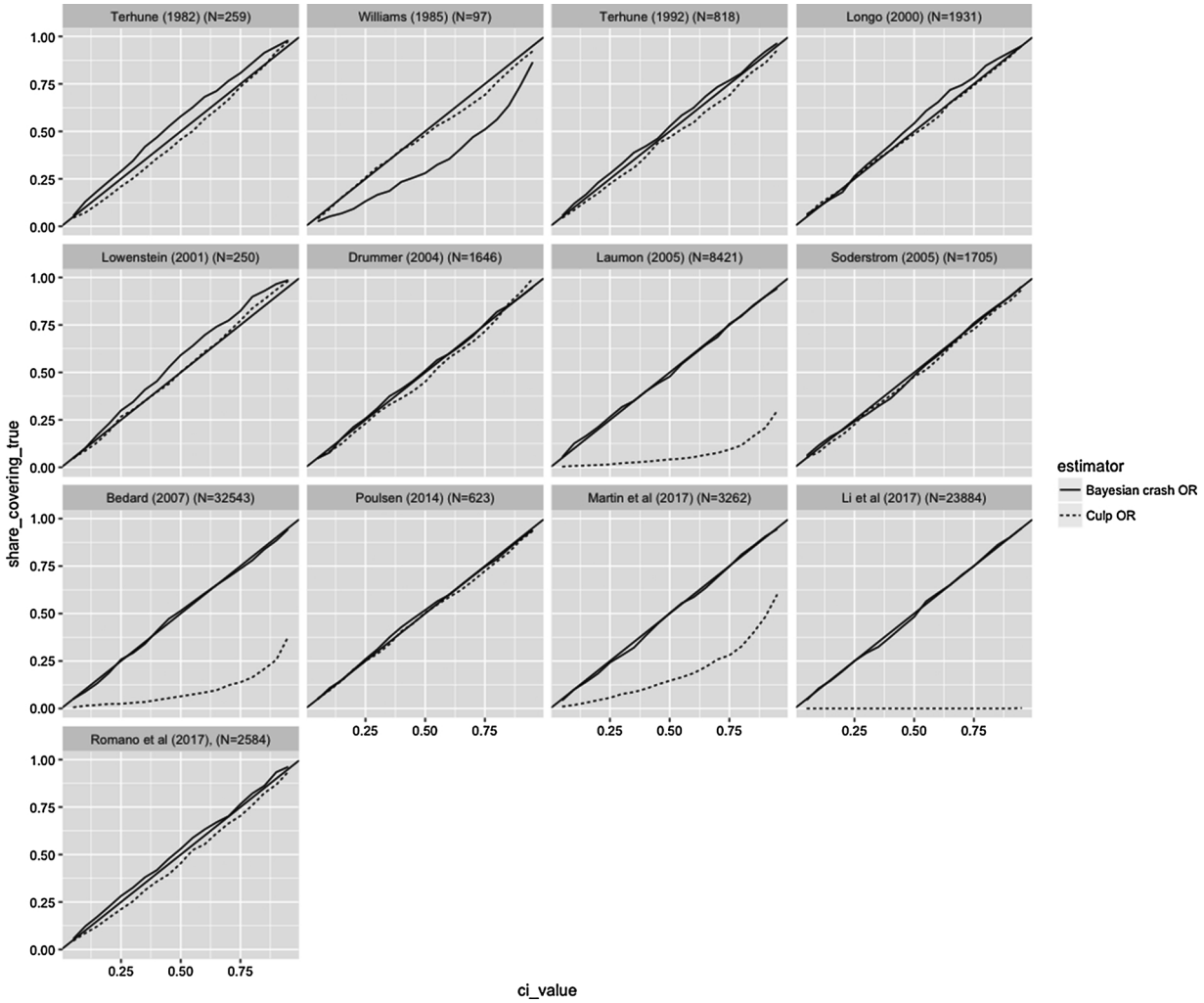

Fig. 4. Calibration test of confidence/uncertainty intervals. Results from 500 bootstrapped samples drawn from – and of equal size to – the original data. Culpability and crash ORs are here assessed relative to the true crash OR in the bootstrapped data.

that the mean attributable risk fraction evaluated across the posterior is below 2% for all but two studies. These results can be compared to reported attributable risk fractions of 2.5% (Laumon et al., 2005) and 4.2% (Martin et al., 2017) erroneously estimated using culpability ORs. Importantly, any inference on attributable risk fractions assumes that a causal interpretation is appropriate, an assumption we return to in the next section.

5. Discussion

A Bayesian model specified from intuitive structural parameters was shown to provide credible inference regarding the overall crash risk implied by culpability study data. Applied on data from 13 published culpability studies, average crash risk increases of cannabis-positive drivers with a Bayesian credibility interval are substantially lower and more precisely estimated than the culpability odds based on a traditional odds-ratio estimator. Interpretational bias has led to an exaggerated impression of the crash risks associated with cannabis in data from culpability crash studies.

While the Bayesian model improves the quality of inference from these data, the strength of the results can be no stronger than the strength of the underlying assumptions and the quality of the individual study data. These limitations are shared with the culpability OR estimator. Specifically, the identification of the increased crash risk hinges on the assumption that nonculpable crash-involved drivers can be viewed as a random sample from the underlying driver population, and the causal interpretation of the risk parameter further depends on the assumption that confounding is balanced across positive and negative drivers. This latter assumption is important, in that the magnitudes of the estimated risk increase is sufficiently small that we cannot rule out residual confounding. Adjustment for confounding was found to have a substantial impact on estimates by Rogeberg and Elvik (2016a, b), who document that the use of unadjusted estimates largely explained the high risks of cannabis impaired driving reported in earlier meta-analyses. Confounding is further indicated by the known higher prevalence of cannabis-impaired driving in groups with higher baseline risks such as young men and individuals with unsafe driving practices and attitudes (Fergusson and Horwood, 2001; Richer and Bergeron, 2009; Bergeron et al., 2014; Bergeron and Paquette, 2014).

The use of prior distributions for the parameters and a Bayesian modeling framework may be unusual for researchers used to working with “off-the-shelf” frequentist estimators like the culpability odds-ratio. A common concern is that priors introduce an element of arbitrary subjectivity in the analysis. While it is certainly possible to impose overly strong priors that predetermine the results of an analysis, however, the point of the priors is rather to reduce excessive sensitivity to sampling variation by statistically encoding knowledge about the plausibility of different regions of parameter space, thus improving the robustness of inference. As shown by the bootstrap analysis, the Bayesian inference model largely coincides with the culpability OR

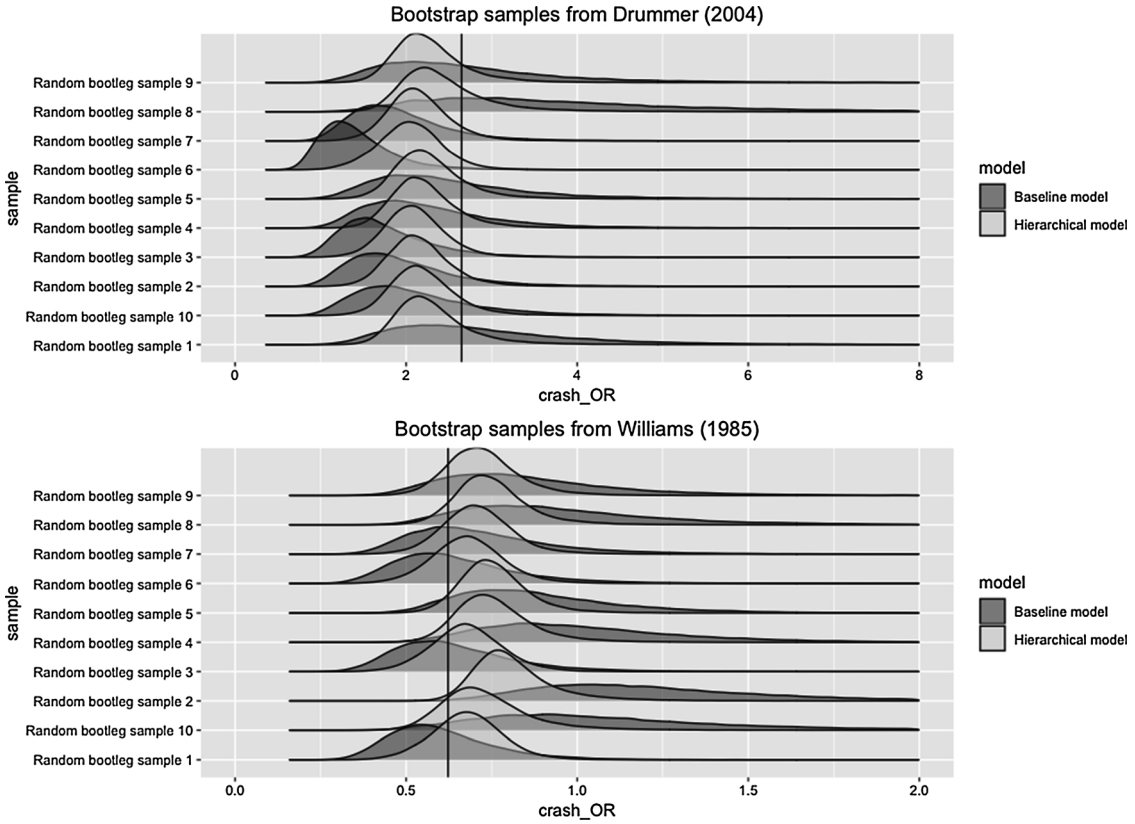

Fig. 5. The posterior distribution of inferred crash ORs – comparison of baseline and hierarchical model. Each of the two panels show the posterior distribution of estimates from two models analysing 10 resampled sets of counts from a low-powered culpability study. One model (baseline) analyses the data from each of the 10 samples independently, the other users a hierarchical effect specification to pool studies.

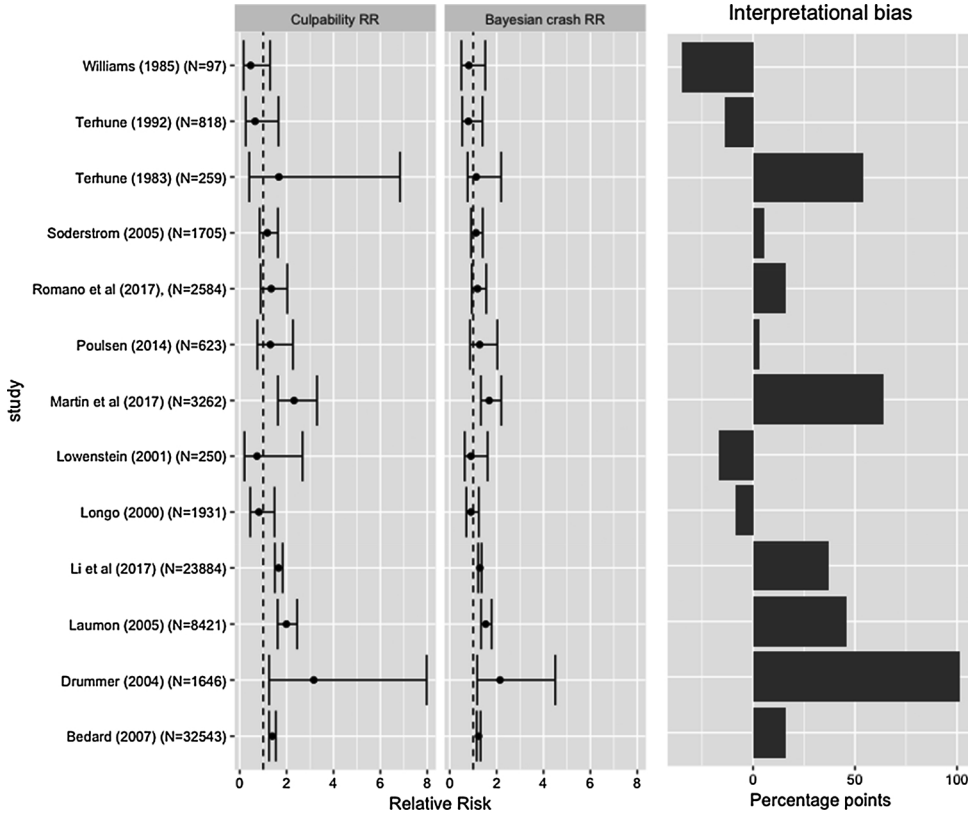

Fig. 6. Culpability and Crash risks compared – The figure shows point estimates with 95% crediblity intervals for culpability and crash risk increases associated with THC-positive drivers, as well as the percentage point difference in risk between the two parameters, corresponding to the interpretational bias resulting from viewing culpability ORs as interchangeable with crash ORs.

Table 3
Crash and culpability relative risk estimates from hierarchical Bayesian model. The table shows the average value of the crash risk and culpability relative risk across the posterior distribution, along with the bounds of the 95% credibility interval.

Study	Crash relative risk	Culpability relative risk
Terhune (1982) (N = 259)	1.18 (0.91-1.56)	1.48 (0.78-2.46)
Williams et al. (1985) (N = 97)	1.15 (0.74-1.59)	1.23 (0.61-1.92)
Terhune et al. (1992) (N = 818)	1.13 (0.73-1.58)	1.2 (0.6-1.87)
Longo et al. (2000) (N = 1931)	1.08 (0.82-1.36)	1.15 (0.66-1.69)
Lowenstein Kozial-Mclain (2001) (N = 250)	1.14 (0.83-1.51)	1.3 (0.65-2.06)
Drummer et al. (2004) (N = 1646)	1.62 (1.12-2.56)	1.82 (1.15-3.04)
Laumon et al. (2005) (N = 8421)	1.49 (1.3-1.72)	1.89 (1.54-2.31)
Soderstrom et al. (2005) (N = 1705)	1.18 (0.96-1.43)	1.28 (0.94-1.67)
Bédard et al. (2007) (N = 32,543)	1.23 (1.15-1.32)	1.39 (1.25-1.55)
Poulsen et al. (2014) (N = 623)	1.33 (0.96-1.8)	1.44 (0.94-2.07)
Li et al (2017) (N = 23,884)	1.28 (1.21-1.35)	1.65 (1.49-1.82)
Martin et al (2017) (N = 3262)	1.54 (1.25-1.95)	1.98 (1.46-2.74)
Romano et al (2017) (N = 2584)	1.22 (0.99-1.5)	1.39 (0.98-1.89)

when used to assess the relative risk increase for culpable crashes (figure S1 in the supporting materials), and the estimated mean of the culpability risk increases in the hierarchical model largely agrees with a random effect meta-regression on culpability ORs. This shows that the main benefit of the Bayesian approach is that it provides inference regarding total crash increases in addition to culpable crash risk increases, as well as its increased stability and precision in small count samples. Conversely, we could also ask why the scientific community should take confidence intervals for culpability ORs seriously when these are based on small-sample studies whose confidence intervals include ORs of 40 or more that are clearly ruled out by the evidence base as a whole. The goal of epidemiological traffic crash research is to develop a credible evidence base that informs us about the actual risks associated with different substances and traits, and the Bayesian approach to inferring crash risk increases appears to do this in a better way than the culpability odds-ratio estimator.

6. Conclusion

Culpability studies typically report the raised odds of culpable crashes associated with some factor, but these estimates provide information on a risk parameter with an unintuitive interpretation and unclear policy relevance, often leading to interpretational bias as researchers treat culpability ORs as direct estimates of crash ORs. A Bayesian model provides accurate inference on the relative risks of a crash, the parameter of interest to researchers and policy makers. Tested using a bootstrap procedure, the model accurately recovers the underlying crash risk increase and produces well-calibrated credibility intervals, and inference can be further improved by using a hierarchical effect specification to capture across-study heterogeneity in the culpability risk parameter.

The main contribution of this study is to explain, document and make available an improved inference model for use in future culpability studies, allowing researchers to assess and report estimates of

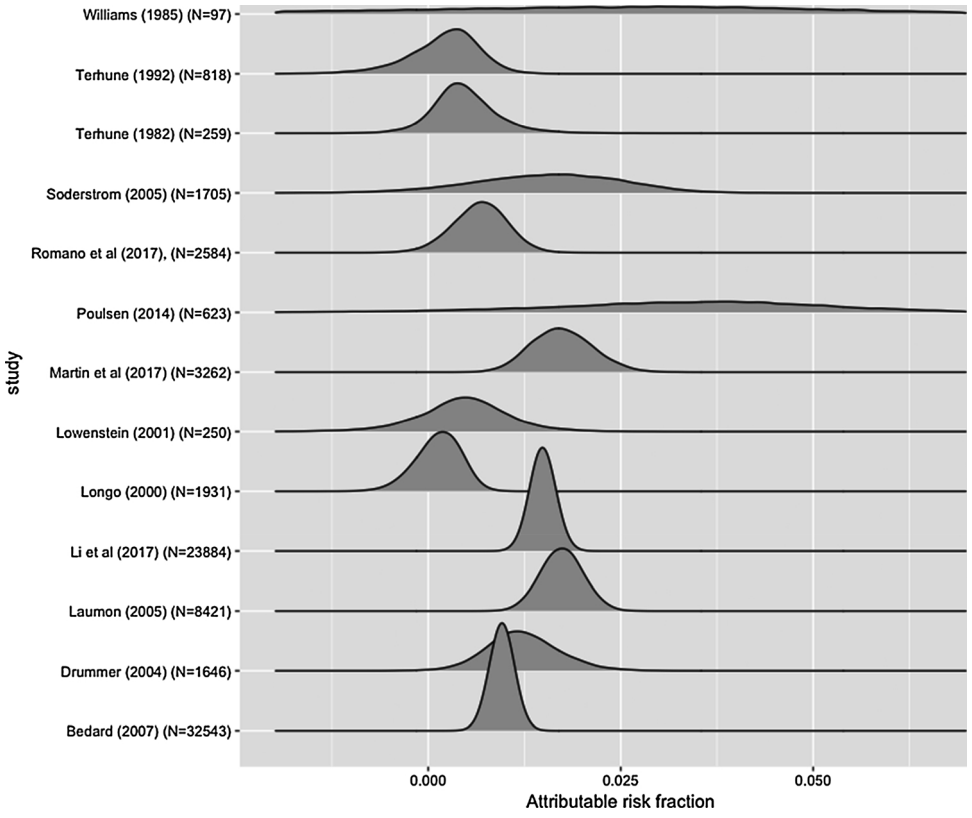

Fig. 7. Posterior distribution of the attributable risk fraction of cannabis – estimated using the hierarchical specification of the Bayesian model.

total crash risk from culpability study counts, and reducing the future risk of interpretational bias.

As a whole, the evidence from the 13 sets of no-alcohol culpability study counts imply that the raised crash risks associated with cannabis are low on average for drivers with THC-values above typical study thresholds. As with culpability odds-ratios the credibility of the reported associations rest on the assumption that nonculpable drivers can be viewed as a random sample from the underlying driver population, while the causal interpretation of estimates as impairment-induced risk increases hinge on the (likely false) assumption that confounding is balanced across positive and negative drivers. Comparing the estimates to those reported using traditional meta-analytic methods, the average crash risk increase of 1.28 (1.16–1.40) is above that reported for culpability study estimates adjusting for individual level factors (age, sex, etc.) (Rogeberg et al., 2018)[1]. This is consistent with residual confounding of the kind individual level controls are meant to reduce. More interesting, however, the pooled estimate from 15 case control studies is reported as 1.82 (1.19, 2.79), suggesting that case control studies *adjusting for individual level confounders* indicate higher risks than culpability studies *without* confounder controls. Understanding this discrepancy should be a priority for future research. The estimated ranges overlap, so the difference could be random and due to (unsystematic) sampling variation. It could also, however, indicate that culpability studies systematically underestimate risk increases, for instance by misclassifying culpable, intoxicated drivers as nonculpable. Conversely, it could indicate that case control studies systematically exaggerate risk increases, for instance by failing to account for the likely systematically reduced participation rates of intoxicated drivers asked to participate in control samples.

Combining the prevalence estimates of cannabis-positive driving and the low (but likely positive) risks, cannabis impaired driving as a whole is estimated to have a minor impact on the total number of crashes – with the mean attributable risk fraction in the majority of study samples well below 1% and most likely below 2.5%. While this indicates that the overall public health impact of cannabis impaired driving is minor relative to that of alcohol-impaired driving, it does not imply that cannabis impaired driving is safe, and the low average is consistent with the presence of a smaller group of high-dose drivers with more substantially raised risks (Rogeberg and Elvik, 2016b). An improved understanding of how dose-response effects affect the crash risks of cannabis impaired driving, and how this risk variation can be identified using road-side tests or THC "breathalyzers" remains a priority.

Acknowledgments

This work was supported by the Norwegian Research Council, grant number 240 235. Michael White provided invaluable assistance in checking and identifying the relevant case and control counts from the underlying studies. The authors of Laumon et al. (2005) and Li et al. (2017) are commended for contributing no-alcohol subsample counts from their data to the author and Michael White respectively. Useful comments from Hans Olav Melberg, Anja Grinde Myrann and Rune Elvik are gratefully acknowledged.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2018.11.011.

References

Asbridge, M., Hayden, J.A., Cartwright, J.L., 2012. Acute cannabis consumption and motor vehicle collision risk: systematic review of observational studies and meta-analysis. Bmj 344, e536.
Bédard, M., Dubois, S., Weaver, B., 2007. The impact of cannabis on driving. Can. J Publ. Heal Can Santee Publique 6–11.
Bergeron, J., Paquette, M., 2014. Relationships between frequency of driving under the influence of cannabis, self-reported reckless driving and risk-taking behavior observed in a driving simulator. J. Saf. Res. 49 (19), e1–24. https://doi.org/10.1016/j.jsr.2014.02.002.
Bergeron, J., Langlois, J., Cheang, H.S., 2014. An examination of the relationships between cannabis use, driving under the influence of cannabis and risk-taking on the road. Rev. Eur. Psychol. Appliquée Eur. Rev. Appl. Psychol. 64, 101–109. https://doi.org/10.1016/j.erap.2014.04.001.
Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al., 2016. Stan: a probabilistic programming language. J. Stat. Softw.
Drummer, O.H., Gerostamoulos, J., Batziris, H., Chu, M., Caplehorn, J., Robertson, M.D., et al., 2004. The involvement of drugs in drivers of motor vehicles killed in Australian road traffic crashes. Accid. Anal. Prev. 36, 239–248.
Fergusson, D.M., Horwood, L.J., 2001. Cannabis use and traffic accidents in a birth cohort of young adults. Accid. Anal. Prev. 33, 703–711.
Gjerde, H., Mørland, J., 2016. Risk for involvement in road traffic crash during acute cannabis intoxication. Addiction 111, 1492–1495.
Greenland, S., 2000. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. Biostatistics 1, 113–122.
Greenland, S., Schwartzbaum, J.A., 2000. Finkle WD, others. Problems due to small samples and sparse data in conditional logistic regression analysis. Am. J. Epidemiol. 151, 531.
Kim, J.H., Mooney, S.J., 2016. The epidemiologic principles underlying traffic safety study designs. Int. J. Epidemiol. 45, 1668–1675. https://doi.org/10.1093/ije/dyw172.
Laumon, B., Gadegbeku, B., Martin, J.-L., Biecheler, M.-B., 2005. Cannabis intoxication and fatal road crashes in France: population based case-control study. Bmj 331, 1371.
Li, G., Chihuri, S., Brady, J.E., 2017. Role of alcohol and marijuana use in the initiation of fatal two-vehicle crashes. Ann. Epidemiol. 27, 342–347. https://doi.org/10.1016/j.annepidem.2017.05.003. e1.
Longo, M.C., Hunter, C.E., Lokan, R.J., White, J.M., White, M.A., 2000. The prevalence of alcohol, cannabinoids, benzodiazepines and stimulants amongst injured drivers and their role in driver culpability: part ii: the relationship between drug prevalence and drug concentration, and driver culpability. Accid. Anal. Prev. 32, 623–632.
Lowenstein, S.R., Koziol-McLain, J., 2001. Drugs and traffic crash responsibility: a study of injured motorists in Colorado. J. Trauma Acute Care Surg. 50, 313–320.
Martin, J.-L., Gadegbeku, B., Wu, D., Viallon, V., Laumon, B., 2017. Cannabis, alcohol and fatal road accidents. PLoS One 12, e0187320. https://doi.org/10.1371/journal.pone.0187320.
Nemes, S., Jonasson, J.M., Genell, A., Steineck, G., 2009. Bias in odds ratios by logistic regression modelling and sample size. BMC Med. Res. Methodol. 9, 56.
Poulsen, H., Moar, R., Pirie, R., 2014. The culpability of drivers killed in New Zealand road crashes and their use of alcohol and other drugs. Accid. Anal. Prev. 67, 119–128.
Richer, I., Bergeron, J., 2009. Driving under the influence of cannabis: links with dangerous driving, psychological predictors, and accident involvement. Accid. Anal. Prev. 41, 299–307.
Rogeberg, O., Elvik, R., 2016a. The effects of cannabis intoxication on motor vehicle collision revisited and revised. Addiction 111, 1348–1359. https://doi.org/10.1111/add.13347.
Rogeberg, O., Elvik, R., 2016b. Cannabis intoxication, recent use and road traffic crash risk. Addiction 111, 1495–1498.
Rogeberg, O., Elvik, R., White, M., 2018. Correction to: 'The effects of cannabis intoxication on motor vehicle collision revisited and revised' (2016). Addiction 113, 967–969. https://doi.org/10.1111/add.14140.
Romano, E., Voas, R.B., Camp, B., 2017. Cannabis and crash responsibility while driving below the alcohol per se legal limit. Accid. Anal. Prev. 108, 37–43. https://doi.org/10.1016/j.aap.2017.08.003.
Soderstrom, C.A., Dischinger, P.C., Kufera, J.A., Ho, S.M., Shepard, A., 2005. Crash culpability relative to age and sex for injured drivers using alcohol, marijuana or cocaine. Annu. Proc. Assoc. Adv. Automot. Med. 49, 327–341.
Stan Development Team, 2016. Stan Modeling Language - User's Guide and Reference Manual.Stan Modeling Language - User's Guide and Reference Manual.
Terhune, K.W., 1982. The Role of Alcohol, Marijuana, and Other Drugs in the Accidents of Injured Drivers. Volume 1 - Findings. US Department of Transportation, National Highway Traffic Safety Administration.
Terhune, K.W., Ippolito, C.A., Hendricks, D.L., Michalovic, J.G., Bogema, S.C., Santinga, P., et al., 1992. The Incidence and Role of Drugs in Fatally Injured Drivers. U.S. Department of Transportation - National Highway Traffic Safety Administration.
Viechtbauer, W., et al., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36, 1–48.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.
Williams, A.F., Peat, M.A., Crouch, D.J., Wells, J.K., Finkle, B.S., 1985. Drugs in fatally injured young male drivers. Public Health Rep. 100, 19.

[1] Note that this previous meta-analysis adjusted the point estimate of culpability studies to avoid interpretational bias, but used the culpability OR standard errors. This inflated the statistical uncertainty of the culpability estimates included in their analysis.