



UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2024:23

Degree project 30 credits

Master`s Programme in Mathematics

July 2024

Examining the robustness of disproportionality analysis of adverse events for medicinal products

Moritz Gustav Berg

Department of Mathematics, Uppsala University

Supervisor: Oskar Gauffin, Niklas Norén

Subject reviewer: Shaobo Jin

Examiner: Rolf Larsson



UPPSALA
UNIVERSITET

Examining the robustness of disproportionality analysis of adverse events for medicinal products

Author:
Moritz Berg

Supervisors:
Oskar Gauffin
Niklas Noren

July 15, 2024

Abstract

The thesis examines the robustness of the Information Component (IC) disproportionality measure. A model used for screening combinations in pharmacovigilance data bases. The focus lies on investigating structural uncertainty in one component of the IC model. This is of interest since as of now this component is assumed to be constant, which could cause the model to overestimate its true precision and ultimately cause faulty screening. To uncover structural uncertainty, a stratified dataset by year is considered. To estimate the structural uncertainty, the residual variance of a generalized additive model, which is fitted on relative reporting rates, is utilized. This estimated structural uncertainty is then incorporated into the IC in two different models. First, in a hierarchical Bayesian model, a prior distribution is included for the previously constant component. The second model follows a parametric bootstrap approach, creating a bootstrap sample of the component in question on which the distribution of the IC is estimated.

The main finding of this study is that 14% of combinations exhibit a substantially increased variance in the models with structural uncertainty in comparison to the current IC model. From these combinations, a few screening results were altered due to this discrepancy. Further results indicate complex relationships between multiple parameters of the model, which complicates the interpretation of the impact that the structural uncertainty has. It can be concluded that the results demonstrate that the IC model could be improved by including structural uncertainty. Nevertheless, for the majority of combinations, the current model accurately estimates the disproportionality.

Acknowledgements

I would like to extend my deepest gratitude to my two supervisors, Oskar Gauffin and Niklas Noren. Oskar was continually there to help with not only my scientific questions and engaged in insightful discussions that sparked new ideas and guided me through the process. Niklas consistently offered fresh perspectives and constructively challenged my approaches, greatly enhancing the quality of this work. I am also grateful to Shaobo Jin for his mathematical expertise and practical advice, which he shared during numerous meetings throughout all stages of this thesis. A special thank you goes to the data science team at UMC. Their support, inspiration, and the enjoyable moments shared over many Fikas made the last six months a memorable experience.

Contents

Abstract	i
Acknowledgements	iii
1 Introduction	1
1.1 Pharmacovigilance	1
1.2 VigiBase Data	2
1.3 Disproportionality Analysis	3
1.4 The Information Component (IC)	5
1.4.1 The IC as Bayesian model	6
1.4.2 The IC025	7
1.5 Research Question	8
2 Mathematical Background	9
2.1 Bayesian Statistics	9
2.1.1 General setup	9
2.1.2 Conjugate prior	10
2.1.3 Hierarchical Bayesian models	11
2.2 Markov Chain Monte Carlo method	11
2.2.1 Metropolis-Hastings algorithm	12
2.2.2 Hamiltonian Monte Carlo	12
2.2.3 No-U-Turn-Sampler (NUTS)	13
2.2.4 Stan	13
2.3 Bootstrap	13
2.3.1 Non-parametric Bootstrap	13
2.3.2 Parametric Bootstrap	14
2.4 Generalized Additive Models	14

3	Methods	15
3.1	Cumulative and stratified expected counts	15
3.2	Study dataset	16
3.3	Choice of Combinations	17
3.4	Bootstrap model	18
3.4.1	Baseline: bootstrap only observed	18
3.4.2	Bootstrap expected and observed	19
3.5	Structural Uncertainty in the Expected	20
3.5.1	Uncertainty in the proportions	22
3.5.2	GAM residual variance	23
3.5.3	Special case of new drugs or reactions	24
3.5.4	Uncertainty in total proportion	25
3.6	Hierarchical Bayesian Model	27
3.6.1	The Model	27
3.6.2	Model Parameters	28
3.6.3	Modelling using Stan	29
3.7	Parametric Bootstrap	30
4	Results	31
4.1	E_{strat} and E_{cum}	31
4.2	Basic bootstrap analysis	31
4.3	Hierarchical Bayesian accounting for structural uncertainty	36
4.4	Parametric bootstrap accounting for structural uncertainty	41
5	Discussion	48
5.1	E_{strat} and E_{cum}	48
5.2	Basic bootstrap analysis	48
5.3	Hierarchical Bayesian accounting for structural uncertainty	50
5.4	Parametric bootstrap accounting for structural uncertainty	51
6	Conclusion	53
A	Appendix	59

Chapter 1

Introduction

1.1 Pharmacovigilance

Pharmacovigilance encompasses the science and activities related to the detection, assessment, understanding, and prevention of adverse effects, commonly known as side effects, or any other drug-related problems [WHO]. Understanding and documenting potential side effects is a fundamental aspect of ensuring the safety of a medicine. The pharmacovigilance process spans both pre-market research, which includes clinical trials, and post-market surveillance but emphasis is on the latter. While the clinical trials are restricted to a rather small number of exposed patients, up to several thousands, which may not represent all groups of patients in the real world. In the post-market surveillance a much larger number of patients is exposed to the drug including all groups, what gives the surveillance its importance. For example patients with comorbidities or pregnant women are typically not included in clinical trials but are exposed to the drugs once they are on the market. Following the approval of a drug by the regulator, monitoring primarily relies on individual case safety reports, from now on simply case reports, which provide detailed information about patients and their adverse events. Adverse events include different types of events not only suspected reactions of the patient but also quality issues of the medicine or medication errors etc. To simplify notation throughout the thesis the term reaction is used instead of adverse event. The reports are collected spontaneously which means that there is no routine follow-up of all exposed patients, instead, reports are submitted once a patient experiences a suspected adverse event to the national health centers or pharmaceutical companies. From there the reports are shared within the WHO Program for International Drug Monitoring (PIDM) [WHO] between all member states and regions which cover around 99% of

the worlds population.

Pharmacovigilance as Uppsala Monitoring Center (UMC) does it focuses on the post-market monitoring and relies primarily on case reports. Initially these reports are screened for drug-event combinations of potential interest for further examination, using disproportionality analysis which will be explained later. This process identifies combinations of a medication with reported reactions that are reported more often than expected. Subsequently, a team of specialists reviews the reports associated with flagged combinations and determines whether the evidence is strong enough to communicate the combination as a signal to the PIDM members and the affected pharmaceutical companies who ultimately after further assessment may take regulatory action to manage possible risks, such as updating the Summary of Product Characteristics (SmPC) and include the reaction as an adverse drug reaction (ADR). The initial screening is important because it allows the specialists to focus their attention on the critical combinations as many reported adverse events may be coincidental or due to underlying disease, or other drugs used at the same time.

1.2 VigiBase Data

VigiBase is the WHO global database of adverse event (AE) reports for medicines and vaccines. It contains all the reports since January 1st, 1967. The reports are shared by the national health organisations from more than 130 countries that collect them for instance from health care professionals, patients or pharmaceutical companies. Each report contains information about the adverse event and treatments of one patient. The information may include information about the patient, like age, sex, weight, etc. and information about the report like the country or date. Additionally information about the treatment, which is coded using the WHODrug dictionary [Lag+20], and the suspected reaction, that is coded using the Medical Dictionary for Regulatory Activities (MedDRA) is included. The main bit of information for the disproportionality analysis is the medicines and vaccines, from now on for simplicity referred to as drugs, that the patient took and the reactions they had [Cou10].

The data comes with some weaknesses. One of them is under-reporting which means that not for every patient a report exists because not every adverse event is recognised as such and of those who are recognised not all are reported. This may lead to a bias towards more severe reactions. Another problem is the lack of

information on the exposed population since a treatment is only reported once an adverse event happened [Cou10]. Another issue are duplicates where the same case is reported multiple times for example by the physician and the pharmacy company. This violates the independence assumption for the reports. For this case there exists an algorithm classifying those duplicate reports so they can be filtered out for further analysis.

In total the database contains more than 35 million reports of more than 4 million different combinations and only in 2023 around 3.2 million new reports were added. This large amount of data makes it not feasible to assess all combinations manually. Hence, a prioritisation of combinations is needed. This prioritisation is done through disproportionality analysis which is the main theme of this thesis. This concept is explained in the next section.

1.3 Disproportionality Analysis

The most widely used approach to statistical signal detection in large collections of individual case reports is disproportionality analysis. As they are the main topic of this thesis, the general concept of disproportionality analysis will be introduced first [Mon+11].

The fundamental idea of disproportionality analysis is to highlight combinations that are reported disproportionately often compared to other combinations. This disproportionality can be measured with many different point estimates that all follow from the components of a two by two contingency table. Let drug X be the drug of interest and reaction Y be the reaction of interest then the contingency table for the combination of X and Y is the following

Table 1.1: Contingency table for a combination of drug X and reaction Y .

	reaction Y	not reaction Y	
drug X	a	b	$n_d = a + b$
not drug X	c	d	$c + d$
	$n_r = a + c$	$b + d$	$n_{\text{tot}} = a + b + c + d$

Note that all counts in this contingency table are on the report level, which means that even if a report contains multiple drugs or reactions it only appears in one count in the table. Based on the four values in this contingency table, there are several point estimators for disproportionality [Cou10]. One of them is the

proportional reporting ratio (PRR) which has the following form

$$PRR = \frac{a/(a+b)}{c/(c+d)}.$$

Here the numerator is the proportion of the ADR of interest among all reports with the drug of interest and the denominator is the same proportion of the ADR but for all reports without the drug of interest. A PRR not equal to 1 indicates an association in the reporting between the drug and suspected reaction. In practice UMC focuses on the positive association indicated by a PRR larger than 1, which happens if the ADR is reported more frequently with the drug of interest because the proportion in the numerator will be larger than the one in the denominator.

Another estimator for the disproportionality is the reporting odds ratio (ROR) which has the form

$$ROR = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}.$$

This estimator can be interpreted as the numerator is the odds of the drug of interest to all other drugs in all reports with the ADR of interest and the denominator is the same odds for all reports without the ADR of interest. Similar to the PRR a ROR value other than 1 indicates association between the drug and reaction of interest.

A third point estimator is the relative reporting ratio (RRR), which can be seen as the foundation of the statistical model this thesis focuses on, namely the information component (IC). The relative reporting rate is defined as

$$RRR = \frac{a/(a+b)}{(a+c)/(a+b+c+d)} \tag{1.1}$$

$$= \frac{a}{n_d \cdot n_r / n_{\text{tot}}}, \tag{1.2}$$

where

$$n_d = (a+b), \quad n_r = (a+c), \quad n_{\text{tot}} = (a+b+c+d).$$

By writing the RRR like such the interpretation of the numerator becomes the observed number of reports with the combination. The denominator can be interpreted as the expected number of reports for the combination because $\frac{n_r}{n_{\text{tot}}}$ is the proportion of reports with the reaction of interest in the overall database which is multiplied by the number of reports with the drug of interest. The representation in terms of the values of the contingency table emphasizes that the independence assumption for the reports is required.

One important property of all three disproportionality estimates is that they

measure only the strength of association between the drug and reaction of interest. They don't measure the amount of data support or in other words the statistical significance. Consider the example of the reporting odds ratio for two sets of counts $a_1 = 10$, $b_1 = 30$, $c_1 = 20$, $d_1 = 50$ and $a_2 = 100$, $b_2 = 300$, $c_2 = 200$, $d_2 = 500$. Both have the same ROR with

$$ROR = \frac{10 \cdot 50}{30 \cdot 20} = \frac{100 \cdot 500}{300 \cdot 200} = \frac{5}{6}$$

even though the second set has 10 times the amount of data supporting the estimator. This makes the models well suited for the task because when combined with a measure of statistical significance they work both for common and rare combinations. Especially the rare combinations are interesting in practice since they are often unknown to the pharmacovigilance experts [NHB13].

1.4 The Information Component (IC)

The information component (IC) [HN12] was developed and is routinely used at UMC. This model is also the one this thesis focuses on and hence I will introduce it in detail in this section.

The formula for the IC based on the RRR is the following

$$IC = \log_2 \frac{a + \frac{1}{2}}{\frac{(a+b)(a+c)}{a+b+c+d} + \frac{1}{2}}.$$

The main difference to the RRR is the term $\frac{1}{2}$ that is added both in the numerator and denominator of the IC. It is a so called shrinkage parameter and includes a measure of statistical significance into the estimator. Its main purpose is to make the model more robust for low values of O and E as without the shrinkage small changes in the expected could have large impact on the IC. It is also worth noting that by adding the same value in the numerator and denominator the quotient gets shrunken towards 1. This means the model is shrunken towards the base assumption of no disproportionality.

Another representation of the IC in terms of O , the number of observed cases, and E , the number of expected cases. From now on also referred to as observed (O) and expected (E) is

$$IC = \log_2 \frac{O + \frac{1}{2}}{E + \frac{1}{2}},$$

where O and E are defined as

$$O = a, \quad (1.3)$$

$$E = \frac{(a+b)(a+c)}{a+b+c+d} = \frac{n_d \cdot n_r}{n_{\text{tot}}}. \quad (1.4)$$

This means that O is the number of reports with that combination and E is the quotient of number of reports with the reaction times number of reports with the drug divided by the total number of reports. Note that for the definitions in 1.3 and 1.4 the independence assumption for the reports is required.

The reason for taking the logarithm with base 2 is mainly historical and to make it easier to visualize several IC-estimates on a joint scale. It therefore will be omitted for large parts of this thesis as it does not change the statistical properties.

1.4.1 The IC as Bayesian model

Bayesian statistics provides a framework for incorporating prior knowledge and updating it with data to form posterior distributions. If one omits the logarithm one can represent the IC as the posterior mean of a Bayesian Poisson-Gamma model [DuM99], from now on also referred to as model A. This allows to construct credibility intervals of the IC. In this subsection the Bayesian Poisson-Gamma model will be introduced and motivated. For further background on Bayesian statistics see chapter 2.

For the Bayesian model it is assumed that the observed count O is Poisson distributed with the rate parameter $\mu \cdot E$ where E is seen as a constant. This is based on the assumption that the variance of E is negligible because it is estimated using the marginal counts n_d and n_r which tend to be much more common than their combination and thus their estimates tend to have lower variance. μ is a hyper parameter in the Bayesian model for which a Gamma prior is assumed. Hence, the Bayesian model can be written as follows

$$\text{likelihood:} \quad O|\mu, E \sim \text{Pois}(\mu \cdot E) \quad (1.5)$$

$$\text{prior:} \quad \mu \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \quad (1.6)$$

$$\text{posterior:} \quad \mu|O \sim \text{Gamma}\left(O + \frac{1}{2}, E + \frac{1}{2}\right). \quad (1.7)$$

The choice of the Poisson distribution for the O count can be motivated by the assumption that all reports are independent and so is the time until one report is

shared with UMC independent of when the last report was shared with UMC. This assumption is both true when considering all reports or when restricting to only reports of one combination. The latter case is the one relevant for this model.

By choosing the conjugate Gamma prior for the Poisson likelihood the existence of a closed form posterior is guaranteed. The mathematical derivation of the posterior can be found in chapter 2. Now the quotient in the IC formula can be represented as the mean value of the posterior distribution like

$$\mathbb{E}(\mu|O) = \frac{O + \frac{1}{2}}{E + \frac{1}{2}}. \quad (1.8)$$

In this model μ can be interpreted as rate in which observations occur with respect to the time it takes to accumulate one expected. So it basically represents the quotient $\frac{O}{E}$. By multiplying μ with E in the Poisson distribution the time scale of the Poisson distribution is changed from the time it takes to accumulate one expected to the time it takes to accumulate E expected. The choice of the parameters of the Gamma prior as $\frac{1}{2}$ has a direct impact on the posterior distribution in the sense that a prior $Gamma(\alpha, \beta)$ will result in the posterior $Gamma(O + \alpha, E + \beta)$. This framework allows for better interpretation of the shrinkage because the prior reflects apriori believes about the parameter.

1.4.2 The IC025

Now that the IC is no longer only a point estimate but with the posterior density also has a distribution, one can consider credible intervals of the IC. The most commonly used one is the credibility interval to the 95% credibility level. One important value that arises through this is the IC025 which is the lower bound of the 95% credible interval or in other words the 2.5% quantile.

Figure 1.1 shows a density of a gamma distribution (the blue line) and the 95% credible interval is highlighted in green. Then the yellow line marks the IC value corresponding with this gamma distribution and then the red line shows the IC025.

In practise the IC025 is the metric that is mainly used for screening the adverse event database. This is because by considering a credibility interval the statistical significance of the disproportionality is considered. As previously indicated, the IC point estimates are designed as a balance between measuring the strength of association and the statistical significance. By taking the IC025 another form of accounting for statistical significance is included. Therefore, using the 95% credible interval is a more conservative approach, as it generally provides a lower estimate.

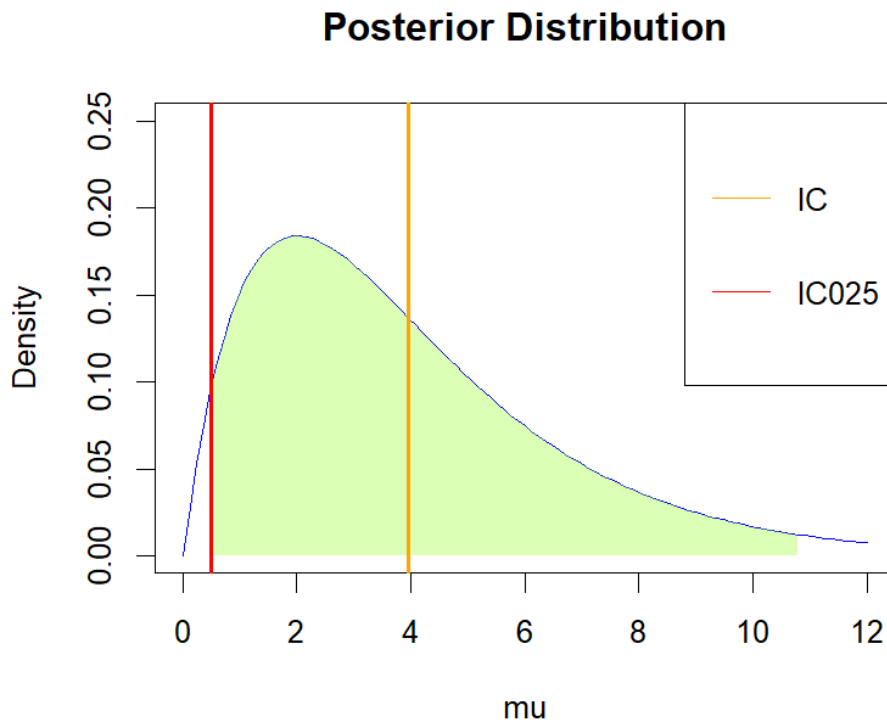


Figure 1.1: Example of posterior density with highlighted credible interval, IC, and IC025.

1.5 Research Question

This thesis investigates the robustness of the Information Component (IC) disproportionality measure, specifically whether the model overestimates its precision by ignoring structural variability in the expected E . This is of interest because it is suspected that there may be greater variability in the expected than would be expected from the assumed independence between reports. It seeks to explore the underlying uncertainty in the expected derived from the empirically observed variability in the relative reporting rates of adverse events over time and its impact on the variance of the IC. This work builds on previous work in [Nor03] where the effect of various parameters on an earlier version of the IC was investigated without accounting for structural uncertainty.

Chapter 2

Mathematical Background

2.1 Bayesian Statistics

2.1.1 General setup

Bayesian statistics represents a principal branch within the field of statistics. It offers an alternative approach to frequentist statistics and is based on Bayes' theorem, which states that for any two elements, A and B , defined in a sample space Ω , with the marginal probabilities, $P(A)$, $P(B)$ and the conditional probability $P(B|A)$, the following holds:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (2.1)$$

To introduce the wording of Bayesian statistics one can say that the unconditional **prior** probability of A is updated into the conditional **posterior** probability of A using knowledge of B .

This theorem serves as the foundation for the Bayesian model, which performs inference on a parameter θ . In this context, the theorem is represented in terms of densities. Let X be the data that was generated from a density $f(x|\theta)$, $\theta \in \Theta$. θ is the unknown parameter of interest. What was the prior probability $P(A)$ in the theorem 2.1 is now the prior density of $\pi(\theta)$. The knowledge with which the prior is updated is now the likelihood function $f(x|\theta)$. This way Bayes theorem gives

$$\pi(\theta|X) \propto f(x|\theta)\pi(\theta), \quad (2.2)$$

where $\pi(\theta|X)$, the posterior density, is the updated density of the parameter θ . It should be noted that the posterior density is only proportional to the right hand

side in equation 2.2 because the right hand side must be normalized by $\int f(x|\theta) d\theta$ for it to be equal.

One key idea of Bayesian models is that the prior density reflects the prior knowledge about the parameter of interest. By selecting a prior distribution, one can incorporate prior knowledge or beliefs into the model. It is crucial to maintain a balance between the prior and the likelihood, or in other words, between the a priori knowledge and the data evidence.

2.1.2 Conjugate prior

One special choice for a prior is a conjugate prior. For a given likelihood function, the posterior density will belong to the same family of probability distributions as the prior from the conjugate class. In practice, this means that if the conjugate prior is, for example, Gamma distributed, then the posterior will also follow a Gamma distribution. Mathematically, it is defined through the concept of a conjugate family.

Definition 1 (Conjugate family). *For a given statistical model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ a family \mathcal{F} of probability distributions on Θ is called conjugate if for every prior $\pi \in \mathcal{F}$ and for every x the posterior $\pi(\cdot|x) \in \mathcal{F}$.*

A conjugate prior is a prior that is an element of a conjugate family. The benefit of a Bayesian model using a conjugate prior is that the posterior density has a closed form. This offers many advantages, such as exact analytical solutions, straightforward inference, and easier interpretation and understanding of the model.

One example of a conjugate Bayesian model which will be used in this thesis is a Poisson-Gamma model. In this model the likelihood function is the density of the Poisson distribution $\ell(x|\theta) = \frac{\theta^x}{x!} e^{-\theta}$ and the prior is a gamma density function $\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$ with the prior hyper parameters α , β . To deduce that the Gamma distribution is the conjugate prior for the Poisson likelihood one writes out the equation for the posterior density like in equation 2.2 with the according likelihood and prior like follows

$$\pi(\theta|x) \propto \frac{\theta^x}{x!} e^{-\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}. \quad (2.3)$$

Dropping all terms that don't depend on θ yields

$$\pi(\theta|x) \propto \theta^{(\alpha+x)-1} e^{-(\beta+1)\theta}. \quad (2.4)$$

Since the term in equation 2.4 as a function of θ is proportional to a density of a Gamma distributions with the parameters $\alpha + x$ and $\beta + 1$, $\pi(\theta|x)$ has to be the density of $Gamma(\alpha + x, \beta + 1)$. Note that the parameters of the posterior density are different to the parameters of the prior and they are called posterior hyper parameters.

2.1.3 Hierarchical Bayesian models

Hierarchical Bayesian models extend the Bayesian models described in section 2.1.1 by treating hyperparameters as random variables and assuming prior distributions for them. This means that a hierarchical Bayesian model consists of a likelihood and more than one prior for different parameters. To expand equation 2.2 into a hierarchical model we assume that the prior of θ depends on another parameter α , $\pi_\theta(\cdot|\alpha)$. For α we assume now a prior distribution $\pi_\alpha(\cdot)$ as well. Following Bayes' theorem, the joint posterior is

$$\pi(\theta, \alpha|x) \propto \ell(\theta|x)\pi_\theta(\theta|\alpha)\pi_\alpha(\alpha). \quad (2.5)$$

From this joint posterior the marginal posterior density for θ can be deduced as follows

$$\pi(\theta|x) \propto \ell(\theta|x)\pi(\theta) \quad (2.6)$$

where

$$\pi(\theta) = \int \pi_\theta(\theta|\alpha)\pi_\alpha(\alpha) d\alpha. \quad (2.7)$$

Note that this is only one example of a hierarchical model, and this class of statistical models is very flexible. It is possible to include even more priors at different levels or have multiple priors at the same level for different parameters.

2.2 Markov Chain Monte Carlo method

For modeling complex Bayesian models and hierarchical Bayesian models, Markov Chain Monte Carlo (MCMC) methods provide a general approach to solving the integrals required for calculating marginal posterior densities and the normalizing constant if they cannot be solved analytically.

The idea behind the MCMC method is to use a Monte Carlo method to approximate an integral $\int f(x)\pi(x)dx$, where $\pi(x)$ is a probability density, and to obtain samples from π as elements of a Markov chain. Realizing the samples recursively in this manner is more efficient than sampling directly from the density. However, producing a Markov chain with the stationary distribution $\pi(x)$ is the main challenge in the MCMC method.

2.2.1 Metropolis-Hastings algorithm

The first algorithm capable of producing such a Markov chain for a given distribution was the Metropolis-Hastings algorithm, first published in Metropolis et al. [Met+53] and generalized by Hastings in Hastings [Has70]. The idea is to draw a candidate for the next element of the Markov chain from a trial distribution that depends on the previous element of the chain. In a second step, a test is carried out to check if the candidate fits into the Markov chain. If the test accepts the candidate, it becomes the new element; otherwise, the previous element is repeated. This algorithm was improved by changing the way the candidate is drawn, giving rise to subsequent Monte Carlo methods.

2.2.2 Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo (HMC) [Nea11] method improves the algorithm for building the Markov chain. Instead of sampling candidates from a trial distribution, it simulates a Hamiltonian dynamical system where the candidates are the simulated particles of the system. Because, in addition to the candidates, an auxiliary momentum is also updated in each step, the model is more likely to make larger steps between the candidates while still proposing candidates that are likely under the given stationary distribution. This reduces the correlation between successive Markov chain samples, making the method more efficient because the Markov chain converges faster to the stationary distribution, and fewer samples are needed for the Monte Carlo approximation.

A downside of this method is that it needs to be optimized for each density by selecting a step length and the number of steps for the simulation of the Hamiltonian dynamical system. This optimization must be done manually and usually requires multiple preliminary runs of the algorithm.

2.2.3 No-U-Turn-Sampler (NUTS)

The No-U-Turn-Sampler (NUTS) [HG14] algorithm is then another optimization of the Hamiltonian Monte Carlo algorithm. It automatically tunes the step length and number of steps during the warm up period of the algorithm. The number of warm up steps W needs to be specified in all of the three algorithms that produce a Markov chain and it means that the first W elements of the of the produced Markov chain are discarded.

2.2.4 Stan

Stan [Car+17] is the probabilistic programming language used in this thesis for the Bayesian statistical modeling. It allows for simple definitions of complex Bayesian models and utilises the NUTS algorithm for sampling. The program offers more tools beside NUTS but this is the only one used for this thesis.

2.3 Bootstrap

Bootstrapping [Efr79] is a widely used statistical method that because of its simplicity is applicable to many models. One can distinguish between two kinds of bootstrap methods the non-parametric and parametric one both are introduced shortly in the following sections [ZM19].

2.3.1 Non-parametric Bootstrap

The non-parametric bootstrap method is appealing because of its minimal assumptions that it makes about the model. However, one assumption is that the observed data sufficiently represents the population. The idea of the bootstrap method is to produce many versions of the observed data the so called bootstrap samples. Each bootstrap sample has the same size as the observed data and is produced by sampling with replacement from the observed data. This sampling can also be viewed as drawing from the empirical distribution of the observed data. Usually the number B of bootstrap samples is rather large with above 1000 to assure a reasonable accuracy. For each of the B bootstrap samples the statistic of interest is calculated. Then the inference is made on the B bootstrap versions of the statistic and for example the mean, variance, or quantiles can be estimated.

2.3.2 Parametric Bootstrap

What sets the parametric bootstrap apart from the non-parametric one is that the observed data is assumed to follow a distribution which often will be fitted to the data by estimating the mean or variance of it. Instead of resampling from the observed data the bootstrap samples are obtained by sampling from the assumed distribution. This way B bootstrap samples are sampled and inference for the statistic is made like in the non-parametric bootstrap method. One advantage of the parametric bootstrap method is that it can be more efficient and often requires fewer resamples to achieve the same accuracy as the non-parametric one if the assumed distribution is correct. The downside is that the model relies on the good fit of the assumed distribution to the data.

2.4 Generalized Additive Models

Generalized additive models (GAM) [Woo17] are an extension of generalized linear models (GLM). GLMs are themselves an extension of linear regression that allows for non-normally distributed response variables through the use of a link-function. GAMs extend this further by enabling a greater flexibility in modeling relationships between the response variables and the prediction variables. While GLMs assume a linear relationship after a potential transformation through the link function, GAMs relax this assumption by replacing the linear predictor by a sum of smooth functions of predictors. The general form of a GAM is

$$g(\mathbb{E}(Y)) = f_1(x_1) + f_2(x_2) + f_2(x_2) + \dots + f_p(x_p), \quad (2.8)$$

where the form of f is unknown. GAMs approximate f by a linear combination of basis functions where the parameters of the linear combination are fitted to the data. The basis functions are specified beforehand and a few popular sets of basis functions include cubic regression spline basis, cubic B-spline basis, Fourier basis, and thin plate spline basis. Depending on how many basis functions are included the GAM is more or less flexible. Because the GAM includes a penalising term for wiggleness to the loss function, overfitting and unnecessary complicated models are prevented.

Chapter 3

Methods

The main contribution of this chapter is to present three models of IC that incorporate uncertainty in the expected value. After some necessary definitions and a description of the study dataset used for all models, a first model is introduced. Then, a method for estimating structural uncertainty is presented, which is incorporated in the remaining two models.

3.1 Cumulative and stratified expected counts

Before delving into the discussion of the three models, let's closely examine the expected value E and the method by which it is derived from the data. In the current IC model, E most commonly has the formula $E = \frac{n_d \cdot n_r}{n_{\text{tot}}}$ (see equation 1.4). In this context all three parameters n_d , n_r , and n_{tot} are cumulative counts. These counts aggregate all reports within the database up to the date at which the IC is calculated. Another approach to computing E involves stratifying the data into M strata and calculating an expected value E_i for stratum i , then summing these expected values to obtain the total expected as follows

$$E = \sum_{i=1}^M E_i = \sum_{i=1}^M \frac{n_{d,i} \cdot n_{r,i}}{n_{\text{tot},i}}, \quad (3.1)$$

where the three parameters indexed with i denote the counts in the i th stratum. Throughout this chapter, unless stated otherwise, the strata represent years [Sea+16].

It's noteworthy that the two methods of calculating the expected value generally yield different results for E . Therefore, we define

$$E_{\text{cum}} = \frac{n_d \cdot n_r}{n_{\text{tot}}}, \quad (3.2)$$

$$E_{\text{strat}} = \sum_{i=1}^M \frac{n_{d,i} \cdot n_{r,i}}{n_{\text{tot},i}}. \quad (3.3)$$

The cumulative expected count has the benefit that it is smoother than the stratified one because especially for rare combinations the counts will be unstable over different strata. The stratified expected on the other hand has the advantage of being more sensitive and enabling the analysis of the change over time. To ensure the comparability of results between different models, the same E needs to be used consistently across all of them. Therefore, to enable the models to account for structural uncertainty between the strata, from now on, E_{strat} will be utilized.

3.2 Study dataset

The models described in the following will all be examined on a study dataset, which is a cleaned version of a frozen instance of the VigiBase from December 31st, 2023. This study dataset contains all reports from VigiBase from the years 1999 to 2019, hence all reports from earlier than 1999 or later than 2019 are filtered out. Furthermore, suspected duplicate reports are removed as well as reports with missing value of the gender of the patient which leaves 18164047 reports in the study dataset. This dataset is then stratified by year into 21 strata. The filtering needs to be done due to two constraints of the VigiBase data.

One constraint is that because of the extensive reporting of Covid-19 vaccines in the years 2020 to 2022 the data is heavily influenced by the Covid-19 pandemic. It is likely that reports on Covid-19 vaccines dominate the data in these years and other drugs are under reported. This could lead to varying biases in the data for different combinations. Therefore, it became common practice to exclude these years when doing analysis on a wide range of combinations.

Another constraint is due to the stratification of the dataset. This approach often results in numerous combinations having zero counts for certain strata, either due to the absence of reports on the drug, or reaction. This problem is particularly pronounced in the earlier years of VigiBase when fewer reports were gathered. Hence, the further filtration to include only years after 1998. In summary, the study dataset used for all models comprises VigiBase data from the years 1999 to 2019, stratified

into 21 strata based on years.

3.3 Choice of Combinations

Throughout the different models three combinations will be analysed more closely. These combinations are fentanyl-nausea, lornoxicam-angioedema and magnesium-"low birth weight baby". They are chosen because all of them are flagged as a signal of disproportionate reporting but they are reported with quite different frequency. The first combination of fentanyl-nausea was reported 48139 times in the study dataset which makes it one of the most common combinations. The reaction nausea was reported 1046211 times and the drug fentanyl 139081 times. The combination has an IC value of 2.587 and without taking the logarithm the value is 6.009. The second combination is of the drug lornoxicam and the reaction angioedema. In total this combination was reported 16 times in the study dataset. The reaction angioedema occurred in 113698 reports, this is not particularly rare but the drug lornoxicam occurred only 1852 times. The combination has an IC value of 0.509 with the current IC model (without taking the logarithm the value is 1.423). The third combination is between the drug magnesium and the reaction low birth weight baby and is even less frequently reported than the first one. It occurred only 5 times in the database. Here the drug was reported 1046 times and the reaction 1633 times. So while the drug was reported similarly often the suspected reaction is reported notably less frequent compared to the first combination. The third combination has an IC value of 5.870 which is 58.487 without the logarithm.

The bootstrap models in 3.4 are computationally expensive because they have to collect the counts for each bootstrap sample, which contains more than 18 million rows. Therefore, the analysis of these models is only feasible to do for single combinations. The two combinations chosen for analysis lornoxicam-angioedema and magnesium-"low birth weight baby" are quite rare, because then fewer data supports the combinations, which could make the variance of the expected larger. The reasoning behind this selection of combinations is that for rare combinations the influence of uncertainty in the expected might be greater than for common combinations.

For the hierarchical Bayesian model in section 3.6 and parametric bootstrap model in section 3.7 an analysis of a larger number of combinations is feasible. Thus for both models additional to the analysis of the three combinations from above random samples of combinations with at least three observations are considered.

For model D a sample of 100 combinations is considered. The sample of random combinations on which the model E is applied can be considerably larger than the 100 combinations from model D, with a total of 1997 combinations. This is possible due to the fact that the parametric bootstrap is considerably faster and requires less computational power. This implies that the sample has been altered, and therefore, the identical combinations as those employed in model D are not required to be included in this sample.

3.4 Bootstrap model

The first model to include uncertainty in the expected is a bootstrap model. The idea is to bootstrap the database on report level and use the different bootstrap samples both for computing the observed O and the expected E .

On the study dataset two bootstrap models are built of which one is used as a baseline comparator. The models mainly differ in the way they use the bootstrapped data because the comparator only uses it to get the observed and the other model for both the observed and the expected. This allows to compare one model without uncertainty in the expected with one that includes uncertainty in the expected. Both models bootstrap on the report level and create bootstrap samples of the study dataset. These bootstrap samples are then split into the yearly strata again. Note that this is slightly different to the usual definition of stratified bootstrap because the size of the strata can vary between the different samples.

3.4.1 Baseline: bootstrap only observed

This model, call it model B, only uses the bootstrap samples for calculating the observed and takes a constant expected. Hence, it does not include uncertainty in the expected. This model deals as a reference model to compare it to the upcoming model C. For the model N bootstrap samples of the study dataset are created. The only parameter taken from the bootstrap samples is the observed

$$O_j^*, \quad \text{observed count across all strata} \quad (3.4)$$

for all samples $j = 1, \dots, N$. The expected is taken from the counts of the original filtered VigiBase data like known

$$E = \sum_{i=1}^M E_i = \sum_{i=1}^M \frac{n_{d,i} n_{r,i}}{n_{\text{tot},i}}. \quad (3.5)$$

This expected is then used in the calculation of the IC for each sample

$$IC_j^* = \frac{O_j^* + 0.5}{E + 0.5}. \quad (3.6)$$

3.4.2 Bootstrap expected and observed

The aim of this model, call it model C, is to include uncertainty in the expected together with the uncertainty in the observed that is present both in model B and C. For that N bootstrap samples of the study dataset are created, like in model B. Each bootstrap sample has the following counts of reports, let $j = 1, \dots, N$ denote the sample and $i = 1, \dots, M$ denote the strata

$$O_j^*, \quad \text{observed count across all strata;} \quad (3.7)$$

$$n_{d,i,j}^*, \quad \text{drug count for one strata;} \quad (3.8)$$

$$n_{r,i,j}^*, \quad \text{reaction count for one strata;} \quad (3.9)$$

$$n_{\text{tot},i,j}^*, \quad \text{total report count for one strata.} \quad (3.10)$$

With those counts the expected for each strata can be calculated like follows

$$E_{i,j}^* = \frac{n_{d,i,j}^* n_{r,i,j}^*}{n_{\text{tot},i,j}^*}. \quad (3.11)$$

From this the expected for one sample j can be deduced like

$$E_j^* = \sum_{i=1}^M E_{i,j}^* = \sum_{i=1}^M \frac{n_{d,i,j}^* n_{r,i,j}^*}{n_{\text{tot},i,j}^*}. \quad (3.12)$$

Note that all parameters are marked with * to indicate that they are taken from the bootstrap sample.

In the next step the IC is calculated for each bootstrap sample using the pair of

O_j^* and E_j^* resulting in N bootstrap IC_j^* s

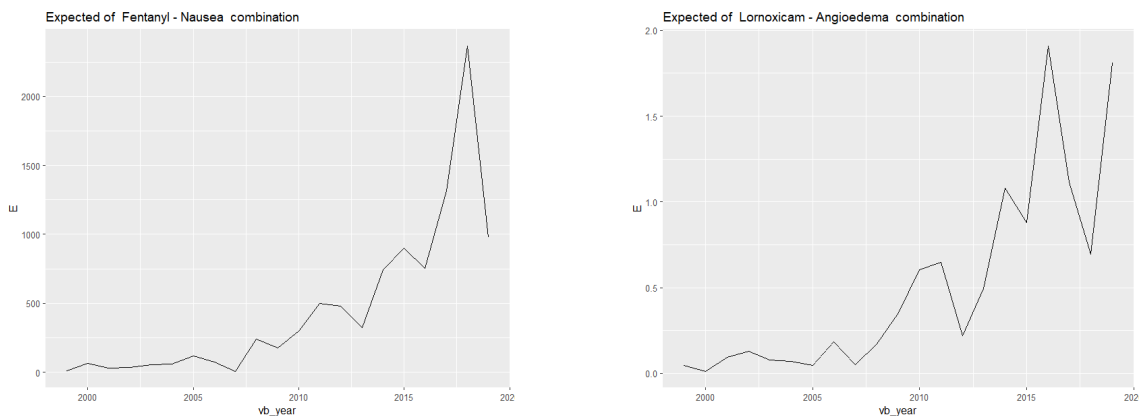
$$IC_j^* = \frac{O_j^* + 0.5}{E_j^* + 0.5}. \quad (3.13)$$

3.5 Structural Uncertainty in the Expected

The previous bootstrap model assumed independence between reports and accounted only for random sampling variability under those assumptions for the expected value and failed to account for structural uncertainty. This section will analyse the type of structural uncertainty present in the data and present a method to approximate and incorporate it into the upcoming models.

The concept of structural uncertainty revolves around the idea that the expected value might be influenced by events unknown to the model and not captured in the data. Such events could include scenarios where doctors become more aware of a particular reaction due to a widely publicised study, leading to increased reporting of the reaction. Similarly, for drugs, such events might involve national health centres encouraging doctors to report cases of specific types of drugs, such as HIV vaccines, to gather more information about them. These are just two examples; numerous other events of this nature could occur. As these events are unknown to the model, they must be treated as random.

To begin exploring the structural uncertainty in the expected, two plots for different combinations are presented in figure 3.1, depicting the expected value per year.



(a) Expected for fentanyl-nausea combination

(b) Expected for lornoxicam-angioedema combination

Figure 3.1: Expected for each year of two combinations

It can be observed that the expected values for both combinations exhibit an upward trend, which is present for all combinations. This phenomenon can be attributed to the fact that the expected E are strongly dependent on the total number of reports, as E is the number of reports that we expect for a given combination. For example, if we have a total of 10,000 reports, with 5 of these belonging to the specified combination, we would expect 10 reports of the combination if the total number of reports were 20,000. This can also be explained in terms of the formula of the expected $E = \frac{n_d n_r}{n_{\text{tot}}}$ where the counts n_d, n_r are proportional to n_{tot} . Hence, if n_{tot} changes by a factor q and the new count is $n_{\text{tot}}^* = q \cdot n_{\text{tot}}$ the other counts would be $n_d^* = q \cdot n_d$ and $n_r^* = q \cdot n_r$. Then the expected would change like follows

$$E^* = \frac{n_d^* n_r^*}{n_{\text{tot}}^*} = \frac{(q \cdot n_d)(q \cdot n_r)}{q \cdot n_{\text{tot}}} = q \cdot \frac{n_d n_r}{n_{\text{tot}}} = q \cdot E. \quad (3.14)$$

To illustrate that further compare the trend of the total counts per year in figure 3.2 to the trend of the expected in figure 3.1.

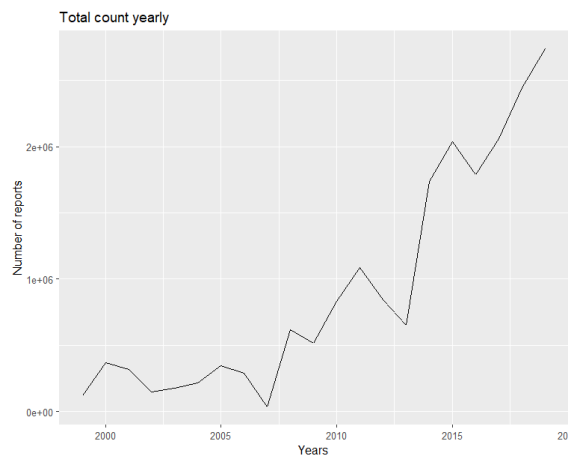


Figure 3.2: Total count of reports per year.

Given that the IC model should be independent of the total count, as effects on the general amount of reporting should not influence the proportional reporting of single drugs or reactions, it is reasonable to condition on the total count n_{tot} . Consequently, when analysing the uncertainty of the expected all effects caused by the n_{tot} should be considered negligible. To accomplish this the expected can be represented as a product of proportion of reports with the drug and the proportion

of reports with the reaction. Let the proportions be defined as

$$p_d = \frac{n_d}{n_{\text{tot}}}, \quad \text{proportion of reports with the drug} \quad (3.15)$$

$$p_r = \frac{n_r}{n_{\text{tot}}}, \quad \text{proportion of reports with the reaction.} \quad (3.16)$$

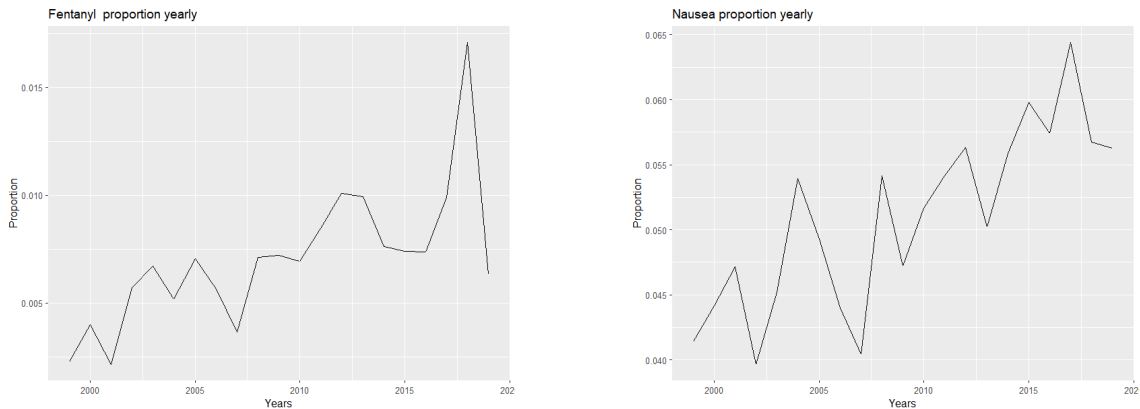
Then the expected can be written as

$$E = \frac{n_d n_r}{n_{\text{tot}}} = p_d \cdot p_r \cdot n_{\text{tot}}. \quad (3.17)$$

This has the advantage that both proportions are independent of the total count and the uncertainty of E can be modelled from the uncertainty in the proportions.

3.5.1 Uncertainty in the proportions

It is now necessary to estimate the structural uncertainty in the proportions. Considering the same two combinations as earlier and plotting the proportions per year in the figures 3.3 and 3.4.

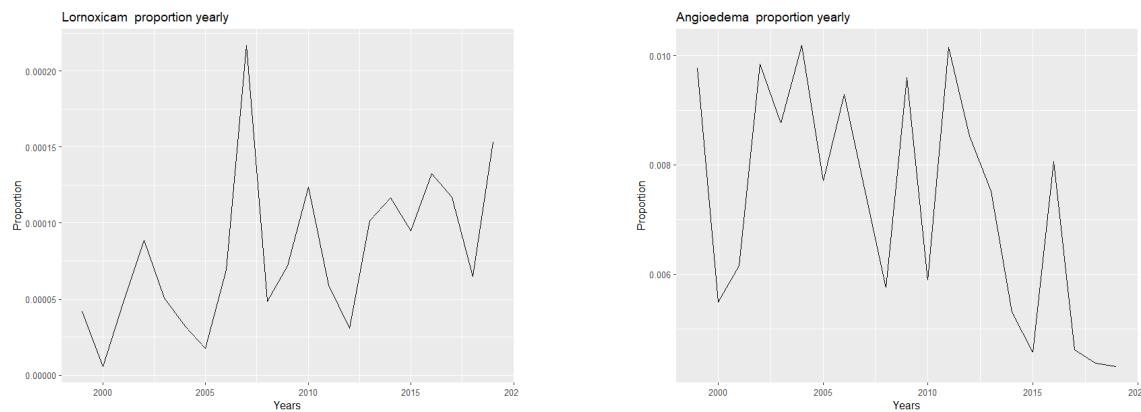


(a) Yearly proportion for the drug fentanyl.

(b) Yearly proportion for the reaction nausea.

Figure 3.3: Yearly proportions for the fentanyl-nausea combination.

In these plots, the trend from the total counts is no longer evident. However, all of them still exhibit some degree of trend. For instance, the nausea reaction appears to have an upward trend (figure 3.3b), while the angioedema reaction appears to have a downward trend (figure 3.4b). It is important to note that these long-term trends should not have an impact on the uncertainty included in the model. Instead, the model should focus on the seemingly random fluctuations in each year.



(a) Yearly proportion for lornoxicam drug.

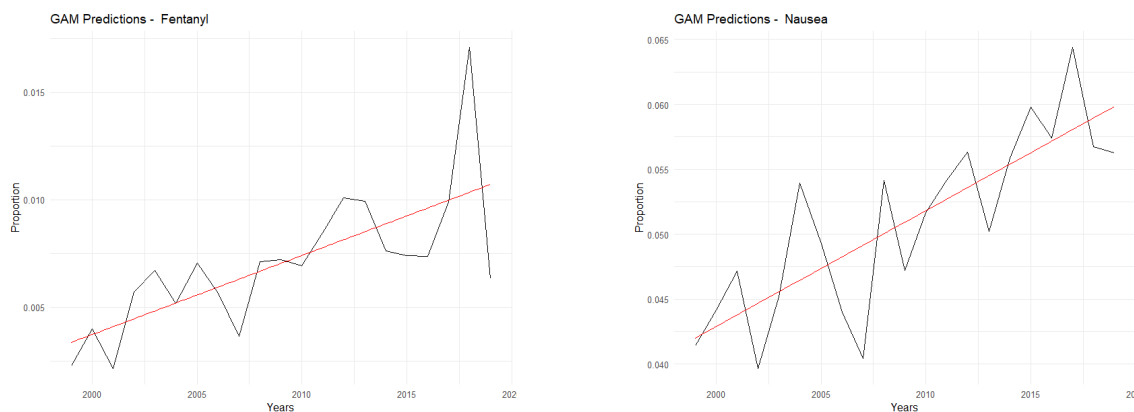
(b) Yearly proportion for angioedema reaction.

Figure 3.4: Yearly proportions for the lornoxicam-angioedema combination.

3.5.2 GAM residual variance

One methodology for estimating this uncertainty is to fit a generalized additive model (GAM) to the yearly proportions and take the sample variance of the residuals as the variance of the yearly proportion. This approach ensures that long-term trends are captured by the GAM, thus preventing them from affecting the variance of the proportion. The fitted GAM assumes the proportion to follow a Gaussian distribution and uses a spline from the thin plate regression spline basis with 4 degrees of freedom to model the relation between the proportion and the years. To illustrate the generalised additive model, Figures 3.5 and 3.6 show the proportions for the same combinations as earlier. The red line represents the prediction of the GAM.

In all six plots in the figures 3.5, 3.6, and 3.7, the GAM model indicated by the red line demonstrates an ability to capture the underlying trend without overfitting the data. While there might be other models capable of modeling these trends GAMs seem to be a good way. This is because they do not over fit the data when the trends are more subtle, compare figure 3.5 and 3.6, and they also are sufficiently flexible to capture diverse non-linear trends as in figure 3.7. For example, these trends could be for a drug that, over the years 1999 to 2008, is released in an increasing number of countries and therefore used and reported with greater frequency. Subsequently, in the year 2009, another drug was developed and gradually superseded the initial drug over the subsequent year, resulting in a decline in reported cases.



(a) Yearly proportion for the drug fentanyl with GAM prediction (in red).

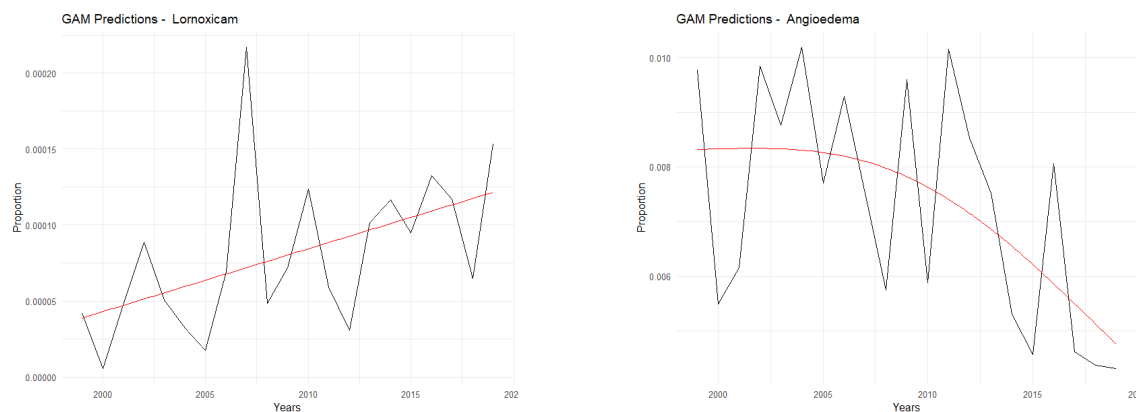
(b) Yearly proportion for the reaction nausea with GAM prediction (in red).

Figure 3.5: Yearly proportions for the fentanyl-nausea combination with GAM prediction.

3.5.3 Special case of new drugs or reactions

A few drugs and reactions fall into a special category because they were introduced a few years after 1999. This happens more frequently for drugs if they only get marketed at a later point but also some reactions are added to the MedDRA dictionary over the years. Therefore, for these drugs and reactions, the counts in the early years are zero and don't have any randomness in them. To estimate the variance of these drugs or reactions, only the years after their introduction into the database should be used. Otherwise, the variance is likely to be underestimated due to the lack of variance in the early years. Two examples are shown in the figure 3.8, where the drug regadenoson and the reaction "product odour abnormal" are only reported after 2008. Note that the GAM model is still fitted to all years, but only for the calculation of the residual variance the residuals are restricted to the years after 2008. The decision not to change the years for the GAM is made for simplicity and because the GAMs are flexible enough to still get a good fit for the remaining years. However, changing the years for the GAM would probably be a slightly more accurate approach.

Another note is that because of this special case, the residual variance or residual standard deviation cannot be obtained using the R-built-in function `sigma()`, but must be calculated directly from the residuals as the uncorrected sample standard deviance. Both approaches should theoretically give the same result, but due to different implementations the results may vary slightly. For consistency, the standard deviation is always calculated directly from the residuals.



(a) Yearly proportion for lornoxicam drug with GAM prediction (in red).

(b) Yearly proportion for angioedema reaction with GAM prediction (in red).

Figure 3.6: Yearly proportions for the lornoxicam-angioedema combination with GAM prediction.

3.5.4 Uncertainty in total proportion

As described in the previous subsection, the variance of the yearly proportion p , which could be for the drug or reaction, is estimated by the residual variance of the GAM. Thus, the GAM yields $\text{Var}(p_i)$, which is the same for all strata i . In the IC model, however, the proportion p is not considered on an annual basis but on a total basis over all years, which we denote as p_{tot} . In this subsection, the variance of p_{tot} is derived from the variance of the annual p_i .

So far, the proportions for drugs and reactions have been treated equally because the IC model is symmetric under interchanging of p_r and p_d . From now on, the two proportions will be treated differently because only the uncertainty of proportions for the reaction will be included in the model and the one for the drug will be considered constant. The decision to include the uncertainty of the reaction and to treat the drug proportion as constant is mathematically arbitrary because of the symmetry of the IC model. However, due to possible different effects for reactions than for drugs, the results could still be different depending on whether the uncertainty of the reaction or the drug is included. The limitation on variance in the reaction enables better traceability of the variance from different parts of the formula and provides a lower limit on the unaccounted variance in the expected and the true effect including also uncertainty in p^d might be even higher. So from now on we will consider the variance in p^r , the proportion of the reaction, and assume p^d to be constant.

Because the expected over all years is the sum of the expected for each year like

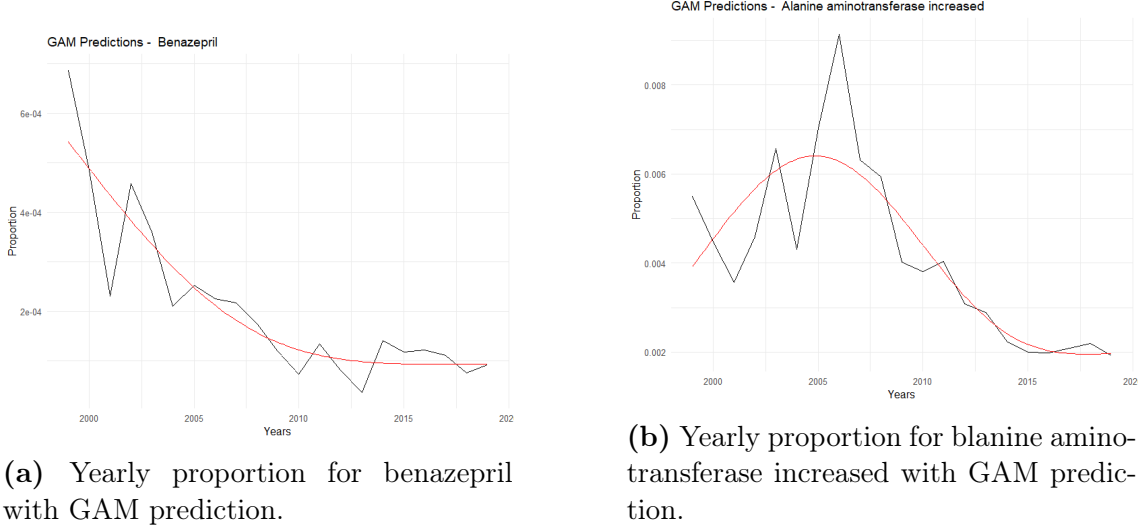


Figure 3.7: Yearly proportions for the benazepril-alanine aminotransferase increased combination with GAM prediction.

this

$$E_{\text{tot}} = \sum_{i=1}^M E_i = \sum_{i=1}^M p_i^r \cdot n_i^d,$$

where the i are the different years and the model considers M different years. Then by using the relation of E with p^r one gets

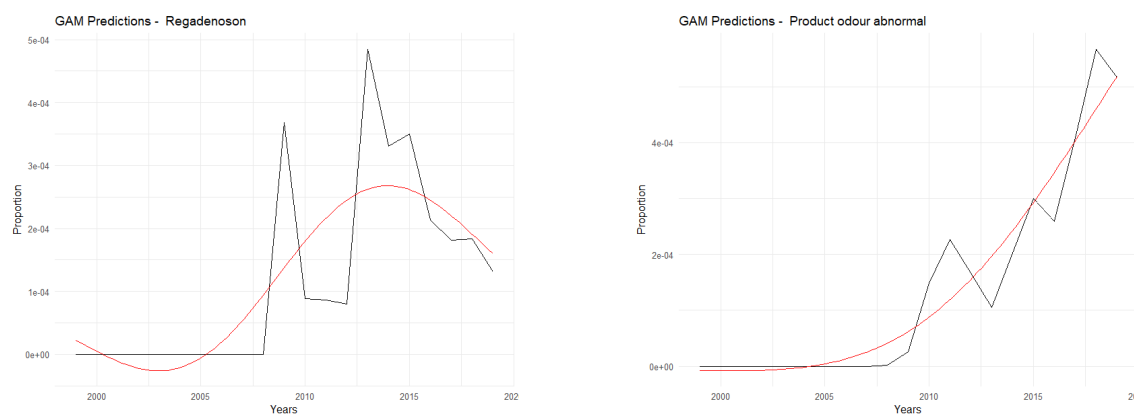
$$p_{\text{tot}}^r = \frac{E_{\text{tot}}}{n_{\text{tot}}^d} = \sum_{i=1}^M p_i^r \cdot \frac{n_i^d}{n_{\text{tot}}^d}.$$

Assume n_i^d and n_{tot}^d are constants and for simplicity the covariance of the p_i^r to be negligible. With those assumptions the following equations hold

$$\text{Var}(p_{\text{tot}}^r) = \text{Var}\left(\sum_{i=1}^M p_i^r \cdot \frac{n_i^d}{n_{\text{tot}}^d}\right) \quad (3.18)$$

$$= \text{Var}(p_1^r) \sum_{i=1}^M \left(\frac{n_i^d}{n_{\text{tot}}^d}\right)^2. \quad (3.19)$$

Because the n_i^d and n_{tot}^d are known from the data, the term in 3.19 is an estimate of the variance of p_{tot}^r that only uses the variance of the yearly proportions that are estimated using the GAM. This estimate is used in the models in the next sections.



(a) Yearly proportion for regadenoson drug with GAM prediction.

(b) Yearly proportion for product odour abnormal reaction with GAM prediction.

Figure 3.8: GAM prediction for a drug and reaction that only appear after 2008.

3.6 Hierarchical Bayesian Model

In this section the structural uncertainty in the expected will be incorporated in the Bayesian model of the IC in an empirical Bayesian approach. The idea is to expand the model described in 1.4.1 to a hierarchical model, from now on called model D, by including a prior distribution for the expected parameter.

In the Bayesian Poisson-Gamma model of the IC from 1.4.1 the expected count E is assumed to be constant. Now the model in this section will treat E as a random variable where the variance of this random variable is estimated from the GAM as described in the previous section. This is where the hierarchical Bayesian model will have some empirical Bayes aspects because the hyperprior is based on estimates from the data. Bayesian statistics makes it quite straightforward to turn a constant parameter into a random variable as one can assume a prior distribution for the parameter. The prior in a Bayesian model depicts the apriori beliefs or knowledge of the parameter. It therefore leaves room to modify the prior distribution of the expected E such that it has the estimated variance. One benefit of including the uncertainty in this way is that it is a natural approach as the current IC (model A) already is a Bayesian model. A more detailed introduction to Bayesian statistics can be found in 2.

3.6.1 The Model

The new model is an extension of the current Poisson-Gamma model and thus has the same likelihood function and prior distribution for μ . One difference however is

that the expected E is represented with the proportion of the reaction

$$E = p_r \cdot n_d. \quad (3.20)$$

With this in mind the hierarchical Bayesian model can be written as follows

$$\text{likelihood:} \quad O|\mu, p_r, n_d \sim \text{Pois}(\mu \cdot p_r \cdot n_d) \quad (3.21)$$

$$\text{prior:} \quad \mu \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \quad (3.22)$$

$$\text{prior:} \quad p_r \sim \text{Beta}(\alpha, \beta) \quad (3.23)$$

From Bayes' theorem it follows that the posterior density for this model is of the form

$$\pi(\mu, p_r|O) = \frac{(\mu \cdot p_r \cdot n_d)^O}{O!} e^{-\mu \cdot p_r \cdot n_d} \frac{\mu^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})} e^{-\mu \frac{1}{2}} \frac{p_r^{\alpha-1} (1-p_r)^{\beta-1}}{B(\alpha, \beta)} \quad (3.24)$$

$$\propto (\mu \cdot p_r)^O e^{-\mu \cdot p_r \cdot n_d} \mu^{-\frac{1}{2}} e^{-\frac{\mu}{2}} p_r^{\alpha-1} (1-p_r)^{\beta-1} \quad (3.25)$$

where $B()$ is the beta function. Because there exists no known density function that is proportional to the function in equation 3.25 and there is no closed form of the posterior density. Note also that the posterior density is a density of the two parameters μ and p_r but we want a density only for μ as this is the parameter that represents the proportion of observed to expected, we are interested in. This density is the marginal posterior density which can be calculated by integrating out the p_r from the posterior density in equation 3.24. This integral cannot be deduced analytically and therefore a numerical MCMC method is used. This will be done by using the Stan program.

The choice of a beta distribution for the prior of the proportion p_r is natural since the support of the beta distribution is $[0, 1]$. Further does the beta distribution with the two parameters allow to specify the mean and variance and the distribution is roughly symmetric around the mean which are reasonable properties for modeling the uncertainty of the proportion. Because deviations of the proportion should be similar on both sides of the mean.

3.6.2 Model Parameters

The parameters μ and p_r are hyper parameters and don't need to be specified in the model. Instead the prior distributions depict the a priori knowledge about the

parameters and indicate how the parameters are distributed. A difference between the two priors is how informative they are for the model. This means how strong they influence the posterior distribution. The difference is that the Gamma prior is in general much less informative than the beta prior in this model which makes sense because on the one hand there is no a priori knowledge about μ and the purpose of the model is to find out knowledge about μ based on the data. On the other hand there is knowledge about p_r because we trust the estimate of p_r from the data. For this reason the parameters α and β of the beta prior are chosen based on the estimates for the mean and variance of p_r . The mean of p_r is estimated as

$$m = \frac{n_r}{n_{\text{tot}}} \quad (3.26)$$

and the estimate for the variance is taken from the GAM as shown in equation 3.19 and denoted with s^2 .

Because the mean and variance of a beta distributed random variable $X \sim \text{Beta}(\alpha, \beta)$ have the form

$$\text{E}(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \quad (3.27)$$

the parameters α and β can be chosen like follows

$$\alpha = \frac{m^2(1 - m) - ms^2}{s^2} \quad \beta = \frac{(m^2(1 - m) - ms^2)(1 - m)}{s^2m} \quad (3.28)$$

to ensure that $\text{E}(X) = m$ and $\text{Var}(X) = s^2$.

This model treats the number of reports with the drug (n_d) as a constant and therefore only includes uncertainty from the reaction counts into the expected.

3.6.3 Modelling using Stan

This model is implemented using the Stan programming language which utilizes efficient MCMC algorithms to model the Bayesian models. This enhanced efficiency allows an analysis of more combinations compared to model B and C. Hence, additionally to the three combinations fentanyl-nausea, lornoxicam-angioedema, and magnesium-"low birth weight baby" a broader analysis of 100 randomly selected combinations is undertaken and can be found in 5.3. These combinations are uniformly drawn from all combinations with at least three reports.

3.7 Parametric Bootstrap

In the previous model D the structural uncertainty was included through a prior. This section introduces a model where the structural uncertainty is included differently, removing the possibility for unintentionally influencing the uncertainty through the data.

The foundation of this model is once again the Poisson-Gamma model (model A) from section 1.4.1. One difference being that the expected E will be represented as $E = p_r \cdot n_d$. That means that the posterior distribution of μ looks like follows

$$\mu|O \sim \text{Gamma}(O + \frac{1}{2}, p_r \cdot n_d + \frac{1}{2}). \quad (3.29)$$

Note that this is for constant p_r and n_d .

To incorporate the uncertainty in the expected, a parametric bootstrap method is used, assuming a beta distribution for the p_r parameter. The estimation of the mean and variance of p_r as well as the deduction of α and β will be analog to section 3.6.2.

The parametric bootstrap model, from now on model E, can be described as a step-wise algorithm. Note that in the following all the bootstrap parameters are indicated with *. The first step is to draw the bootstrap sample of 100.000 $p_{r,j}^*$ with $j = 1, \dots, 100000$ from the distribution $\text{Beta}(\alpha, \beta)$. The bootstrap sample size of 100.000 is chosen to assure a small bootstrap error and is possible because of the low computational cost of this model. The $p_{r,j}^*$ are then plugged into the posterior distribution from equation 3.29 creating 100.000 different gamma distributions

$$\text{Gamma}(O + \frac{1}{2}, p_{r,j}^* \cdot n_d + \frac{1}{2}). \quad (3.30)$$

In the last step a sample μ_j^* is drawn from each of the gamma distributions. The bootstrap sample of the 100.000 μ_j^* is then used to approximate a density for μ that can be analysed and compared with the initial posterior density of μ . This comparison of the bootstrap density with a non bootstrap density is possible because of the large sample size of 100.000. If one would take a bootstrap approximation of model A with that sample size it would be converged to model A.

Chapter 4

Results

In this chapter a comparison of the cumulative and the stratified expected E and the results of the previously introduced models are presented in separate sections.

4.1 E_{strat} and E_{cum}

A preliminary comparison indicates that the values are relatively close to each other. This observation is supported by Figure 4.1, where the data points closely follow the equilibrium line where both expected values are equal.

Upon deeper analysis of the relationship between E_{strat} and E_{cum} , by examining the normalized difference

$$d(E_{\text{strat}}, E_{\text{cum}}) = \frac{E_{\text{strat}} - E_{\text{cum}}}{\frac{E_{\text{strat}} + E_{\text{cum}}}{2}} = \frac{2(E_{\text{strat}} - E_{\text{cum}})}{E_{\text{strat}} + E_{\text{cum}}}, \quad (4.1)$$

the results indicate that E_{strat} tends to be smaller than E_{cum} , with an average normalized distance of -1.100517 taken over all combinations with more than 3 reports. Or following the ratio $\frac{E_{\text{cum}}}{E_{\text{strat}}}$ one can deduce that E_{cum} is in median 81% larger than E_{strat} , which is a rather large difference. Note that this analysis was made on yearly strata and may look different for other stratifications.

4.2 Basic bootstrap analysis

Each model is run four times with 1000 bootstrap samples and the variance of the IC within each bootstrap model together with the resulting percentile confidence interval width is presented in the tables 4.2, 4.4.

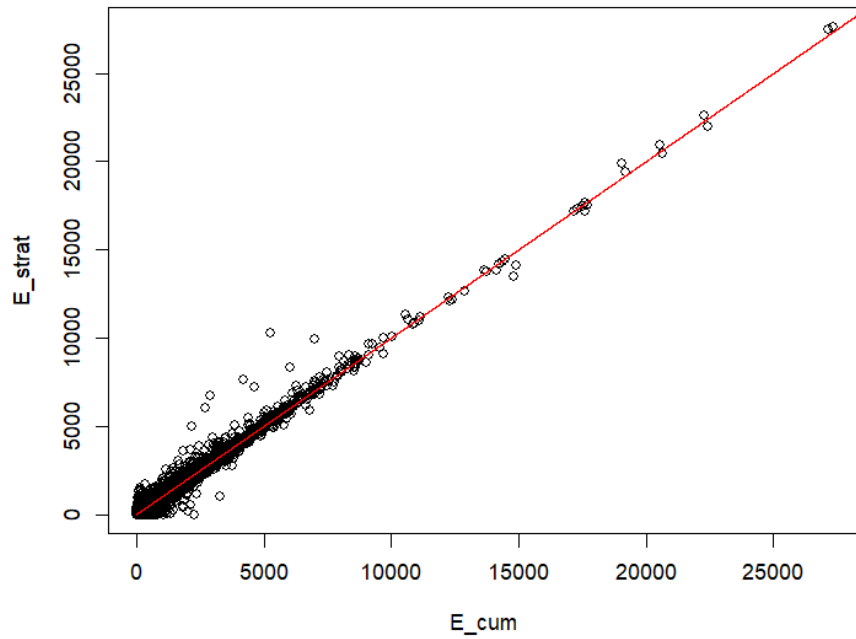


Figure 4.1: Plot of E_{strat} to E_{cum} for all combinations with highlighted equilibrium line (red) where $E_{\text{strat}} = E_{\text{cum}}$.

Starting with the results of the first combination of the drug lornoxicam and the reaction angioedema, see table 4.2. For model B where only the observed O was bootstrapped, the bootstrap estimate of $\text{Var}(IC)$ ranges from 0.1253 to 0.1457, showing precision to the first decimal place. For model C that bootstrapped both the observed O and the expected E , the estimate of $\text{Var}(IC)$ ranges from 0.1273 to 0.1370, also showing precision to the first decimal.

For some context the variance and credible interval width of model A and the counts and estimates from the original data are provided in table 4.1.

Table 4.1: Combination: Lornoxicam - Angioedema

Item	Value
Reports for Combination	16
Reports for Angioedema	113698
Reports for Lornoxicam	1852
IC Value (logarithm)	0.509
IC Value (without log)	1.423
Var(IC) (without log)	0.1128
Cred. int. width (without log)	1.3098

Table 4.2: Variance and bootstrap confidence interval of the IC for the combination lornoxicam angioedema in both models.

Model	Iteration	Variance	Confidence Interval Width
Only O bootstrap	1	0.1329	1.366
	2	0.1253	1.3667
	3	0.1457	1.5272
	4	0.1378	1.4244
O and E bootstrap	1	0.1302	1.3364
	2	0.1353	1.4928
	3	0.1370	1.4230
	4	0.1273	1.3607

For the second combination of the drug magnesium and the reaction "low birth weight baby" the results are shown in tables 4.4. For model B, the bootstrap estimate of $\text{Var}(IC)$ ranges from 0.2600 to 0.3070. For the model where both the observed O and the expected E were bootstrapped, the estimate of $\text{Var}(IC)$ ranges from 0.1634 to 0.2700.

For this combination the variance and credible interval of model A together with the counts of the original data can be found in table 4.3.

For both combinations in both models the change in confidence interval width is rather small with a maximum difference of 0.5597 in the second combination.

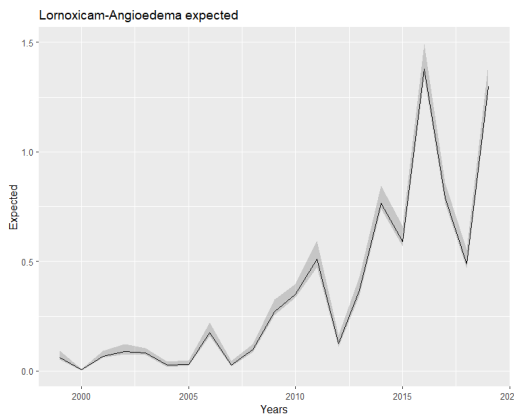
Table 4.3: Combination: Magnesium - Low Birth Weight Baby

Item	Value
Reports for Combination	5
Reports for Low Birth Weight Baby	1633
Reports for Magnesium	1046
IC Value (logarithm)	5.870
IC Value (without log)	58.487
Var(IC) (without log)	15.5859
Cred. int. width (without log)	15.2383

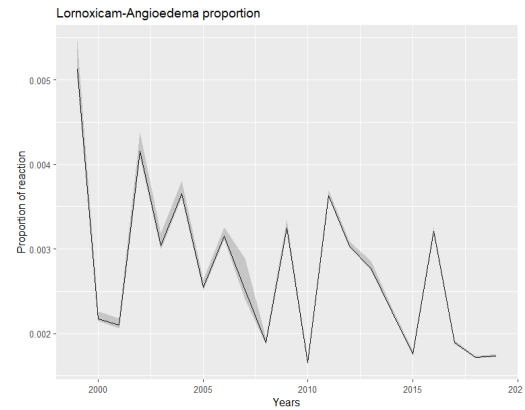
Table 4.4: Variance and bootstrap confidence interval of the IC for the combination magnesium low birth weight baby in both models.

Model	Iteration	Variance	Confidence Interval Width
Only O bootstrap	1	0.2815	2.1468
	2	0.2600	1.9475
	3	0.2686	2.0506
	4	0.3070	2.0506
O and E bootstrap	1	0.1655	1.5871
	2	0.1634	1.7582
	3	0.2700	2.0519
	4	0.2699	2.0642

To illustrate the amount of uncertainty that is present in the expected the plots in figure 4.2 and 4.3 show the expected and proportion of the reaction over the years together with the 95% bootstrap confidence interval indicated by the grey shadow. Note that these intervals are made from one iteration of the bootstrap with 1000 samples, but the interval width is quite stable when increasing the bootstrap sample size. The proportion is included in the plot because it does not exhibit such a strong trend as the expected.

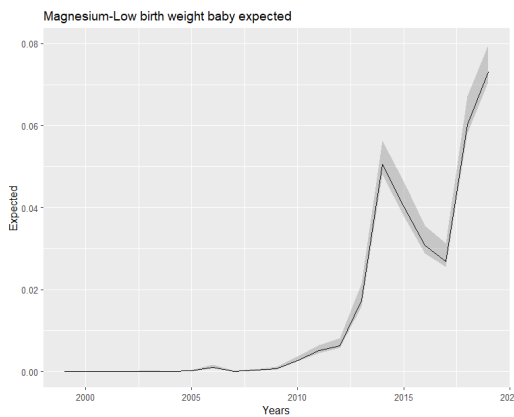


(a) lornoxicam-angioedema expected with confidence interval.

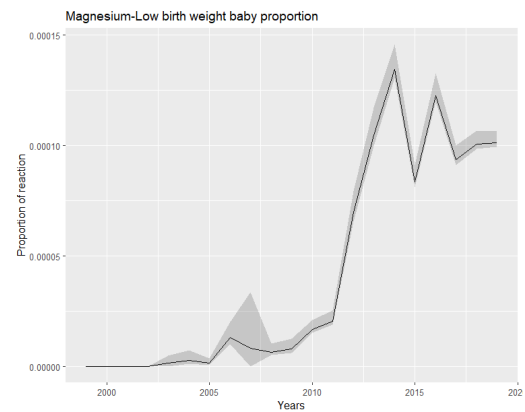


(b) angioedema proportion with confidence interval.

Figure 4.2: Expected and reaction proportion together with bootstrap confidence intervals.



(a) magnesium-low birth weight baby expected with confidence interval.



(b) low birth weight baby proportion with confidence interval.

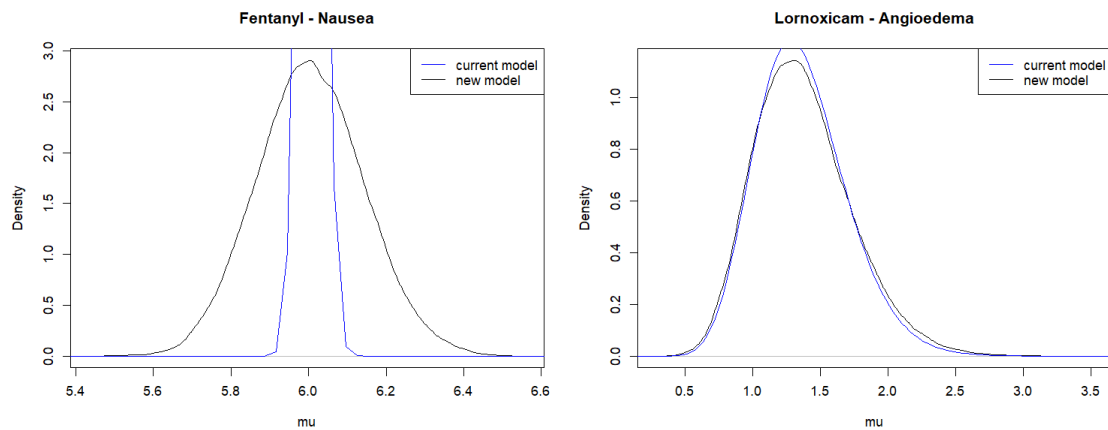
Figure 4.3: Expected and reaction proportion together with bootstrap confidence intervals.

4.3 Hierarchical Bayesian accounting for structural uncertainty

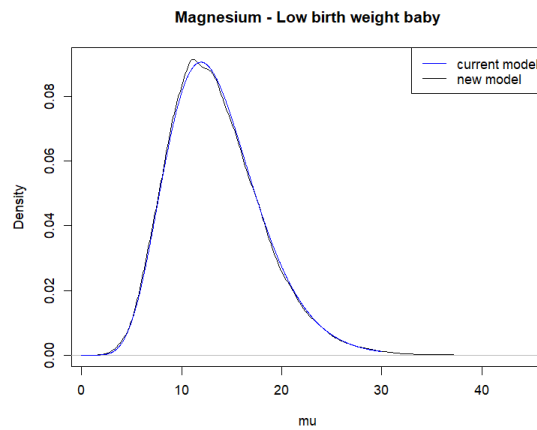
The three combinations fentanyl-nausea, lornoxicam-angioedema, and magnesium-"low birth weight baby" that have partly been studied in previous models are first considered in more detail before a broader analysis of 100 randomly selected combinations is undertaken. The Figures 4.4a, 4.4b, and 4.4c show the posterior density of μ of model D in black and for comparison the posterior density of model A is also plotted in blue. The table 4.6 gives the corresponding values. Additionally the counts and important estimates of model A can be found in tables 4.1, 4.3, and 4.5 for the three combinations.

Table 4.5: Combination: Fentanyl - Nausea

Item	Value
Reports for Combination	48139
Reports for Nausea	1046211
Reports for Fentanyl	139081
IC Value (logarithm)	2.587
IC Value (without log)	6.009
Var(IC) (without log)	0.0008
Cred. int. width (without log)	0.1074



(a) Posterior density of μ for the combination of fentanyl and nausea. (b) Posterior density of μ for the combination of lornoxicam and angioedema.



(c) Posterior density of μ for the combination of magnesium and low birth weight baby.

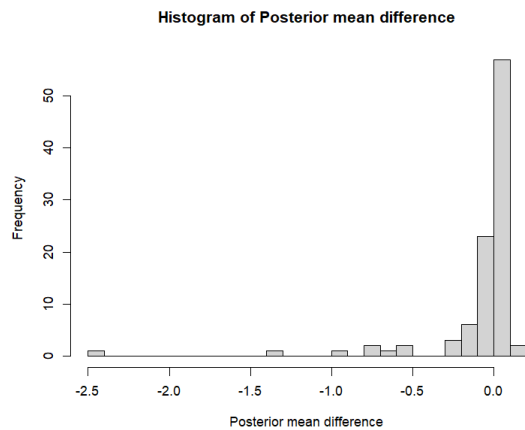
Figure 4.4: Posterior densities of μ for model D in black and Poisson-Gamma model in blue for various drug-condition combinations.

Table 4.6: Comparing main properties of the current IC model (Gamma-Poisson) with model D across different combinations.

Case	Statistic	model A	model D
fentanyl-nausea	Mean	6.01	6.01
	Variance	0.0008	0.0192
	0.025 Quantile	5.96	5.74
lornoxiam-angioedema	Mean	1.36	1.37
	Variance	0.1128	0.1293
	0.025 Quantile	0.79	0.77
magnesium-low birth weight baby	Mean	9.26	9.2
	Variance	15.586	15.416
	0.025 Quantile	3.21	3.15

Arriving at the results from an analysis of 100 randomly selected combinations.

The histogram in Figure 4.5 shows the difference in the mean of the posterior densities. It reveals that the mean does vary between model A and D.

**Figure 4.5:** Histogram of the difference of posterior mean between model D and model A.

The Figure 4.6 compares the posterior variance of model D with model A by plotting $\text{Var}(\mu_D) - \text{Var}(\mu_A)$ against the prior standard deviation. This difference is notably more often positive than negative, indicating an increase in variance on average.

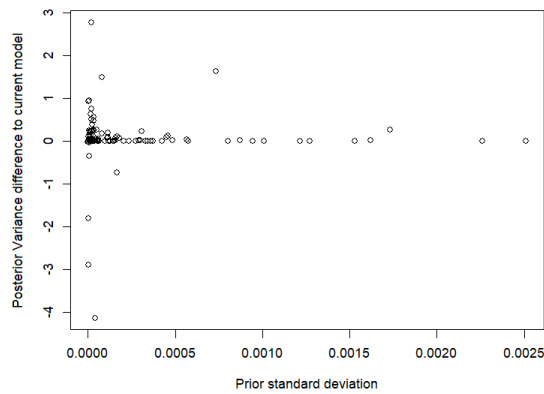


Figure 4.6: Difference of posterior variance of μ in model D compared to model A with standard deviation of beta prior on the x axis.

The histograms in Figure 4.7 and 4.8 also compare model D with model A. In 4.7 the difference $\text{Var}(\mu_D) - \text{Var}(\mu_A)$ in posterior variance is plotted. A noteworthy result here is that 10 combinations (10%) have a lower posterior variance in model D compared to model A. In 4.8 the difference is then normalized by dividing it by the mean variance as follows

$$\frac{\text{Var}(\mu_D) - \text{Var}(\mu_A)}{\frac{\text{Var}(\mu_D) + \text{Var}(\mu_A)}{2}}. \quad (4.2)$$

This normalizing is done to account for the large differences in the scale of μ between different combinations. The main takeaway from the histogram 4.8 is that 9 combinations (9%) experience an increase in variance of more than 100%, which equals a normalized difference of more than $\frac{2}{3}$.

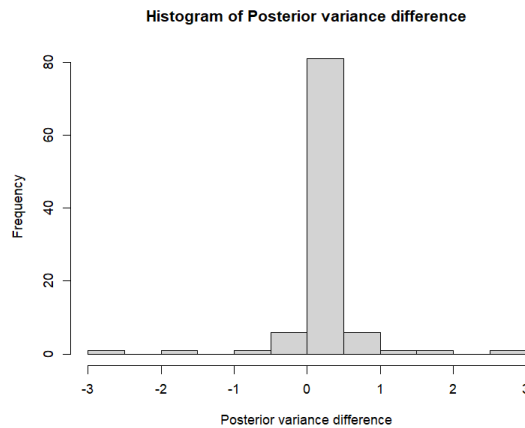


Figure 4.7: Histogram of the difference of posterior variance between model D and model A.

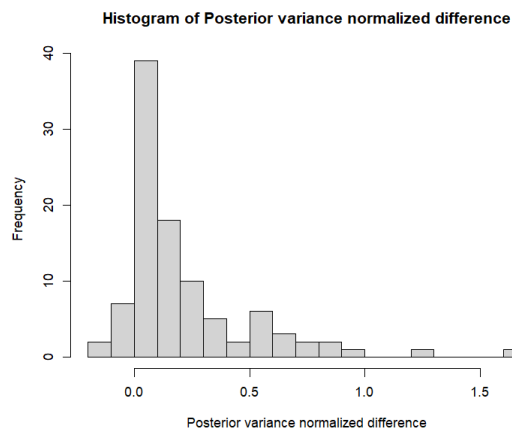
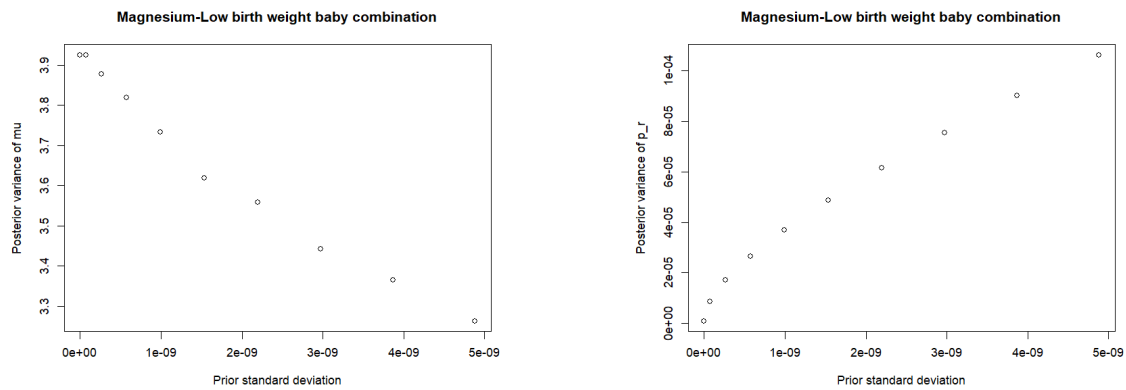


Figure 4.8: Histogram of the difference of posterior variance between model D and model A, normalized by the average of the variance.

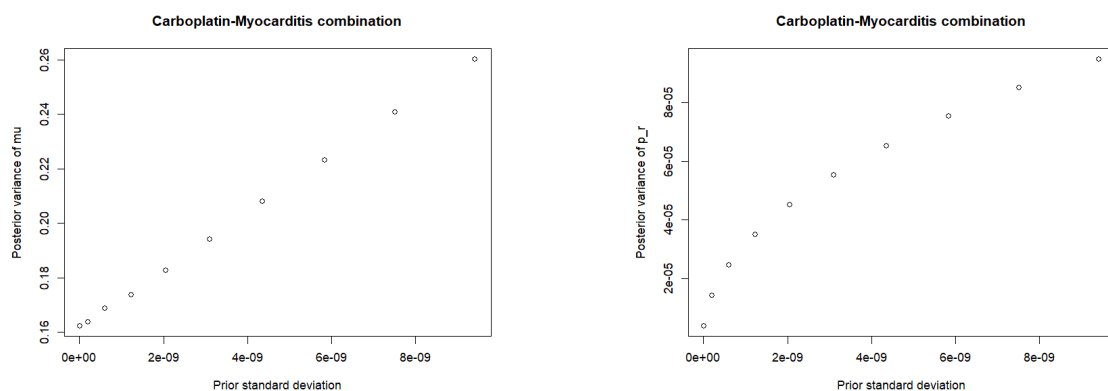
The Figures 4.9 and 4.10 show the posterior variance of μ and p_r against the prior standard deviation. Note that the values of the prior standard deviation are adjusted around the estimated standard deviation for the combination. The noteworthy insight is that the posterior variance of μ has a downwards trend for the magnesium-"low birth weight baby" combination.



(a) Posterior variance of μ to prior standard deviation for one combination.

(b) Posterior variance of p_r to prior standard deviation for one combination.

Figure 4.9: Change in the posterior variance for the magnesium-low birth weight baby combination.



(a) Posterior variance of μ to prior standard deviation for one combination.

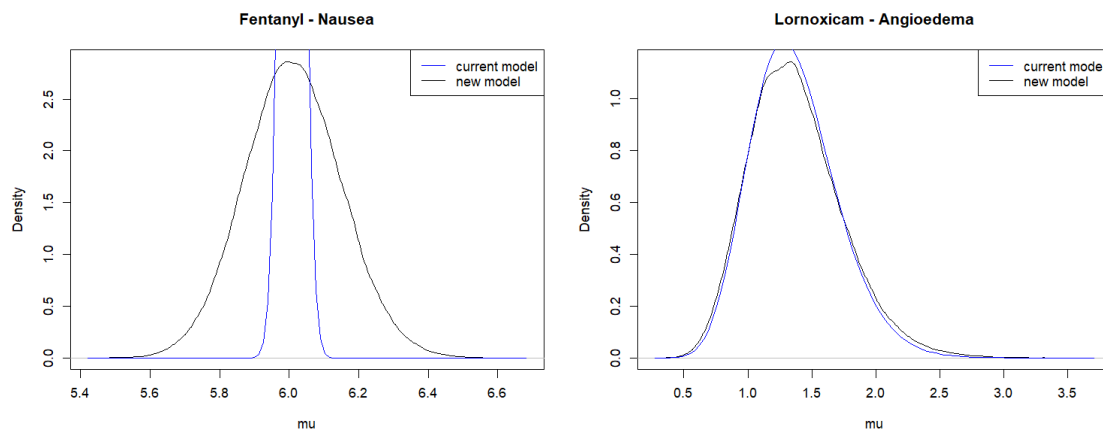
(b) Posterior variance of p_r to prior standard deviation for one combination.

Figure 4.10: Change in the posterior variance for the Carboplatin-Myocarditis combination.

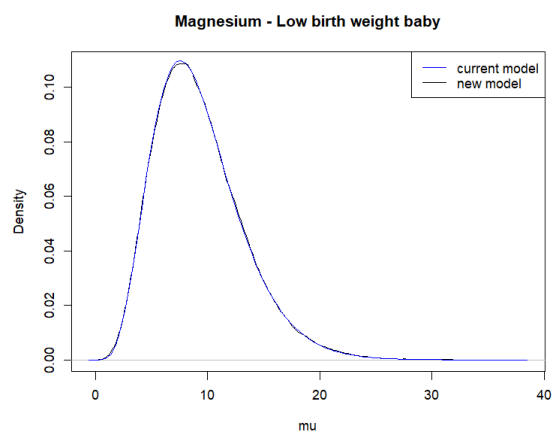
4.4 Parametric bootstrap accounting for structural uncertainty

The results of this model are presented in the same way as for the hierarchical Bayesian model. First discussing the same three combinations fentanyl-nausea, lornoxicam-angioedema, and magnesium-"low birth weight baby" before analysing a sample of 1998 randomly selected combinations.

The Figures 4.4a, 4.4b, and 4.4c show the posterior density of μ of model E in black and for comparison the posterior density of model A is also plotted in blue. The table 4.6 gives the corresponding values.



(a) Posterior density of μ for the combination of fentanyl and nausea. (b) Posterior density of μ for the combination of lornoxicam and angioedema.



(c) Posterior density of μ for the combination of magnesium and low birth weight baby.

Figure 4.11: Posterior densities of μ for model E in black and model A in blue for various drug-condition combinations.

Table 4.7: Comparing descriptive statistics of the model A with model E across different combinations.

Combination	Statistic	Model A	Model E
fentanyl-nausea	Mean	6.01	6.01
	Variance	0.0008	0.0193
	0.025 Quantile	5.96	5.75
lornoxiam-angioedema	Mean	1.36	1.38
	Variance	0.1128	0.1318
	0.025 Quantile	0.79	0.77
magnesium-low birth weight baby	Mean	9.26	9.26
	Variance	15.586	15.606
	0.025 Quantile	3.21	3.25

Arriving at the results obtained from 1998 randomly drawn combinations.

The histogram in Figure 4.12 shows the difference in the mean of the posterior densities. It reveals that the mean does vary between model A and E.

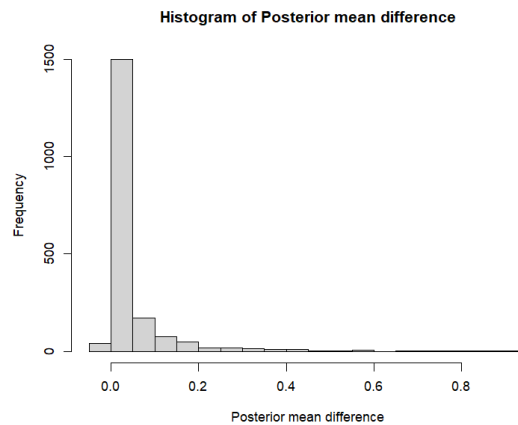
**Figure 4.12:** Histogram of the difference of posterior mean between parametric bootstrap and model A.

Figure 4.13 contrasts the posterior variance of model E with that of model A, plotting $\text{Var}(\mu_E) - \text{Var}(\mu_A)$ against the prior standard deviation. This comparison reveals that the difference is more frequently positive than negative, indicating a tendency toward increased variance. Furthermore, the difference doesn't seem to be driven by the prior standard deviation as no such trend can be observed.

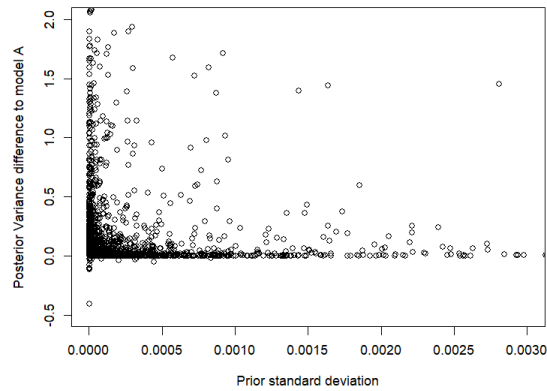


Figure 4.13: Difference of posterior variance of μ in model E compared to model A with standard deviation of beta prior.

The histograms in figure 4.14 and 4.15 also compare model E with model A. In figure 4.14 the difference $\text{Var}(\mu_E) - \text{Var}(\mu_A)$ in posterior variance is plotted. In figure 4.15 this difference is then normalized by dividing it by the mean variance as follows

$$\frac{\text{Var}(\mu_E) - \text{Var}(\mu_A)}{\frac{\text{Var}(\mu_E) + \text{Var}(\mu_A)}{2}}. \quad (4.3)$$

This normalizing is done to account for the large differences in the scale of μ between different combinations.

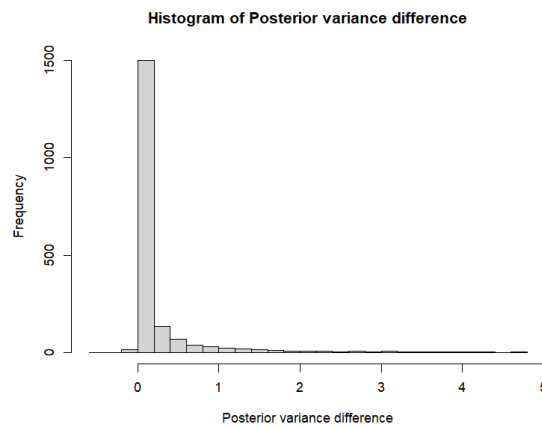


Figure 4.14: Histogram of the difference of posterior variance between model E and model A.

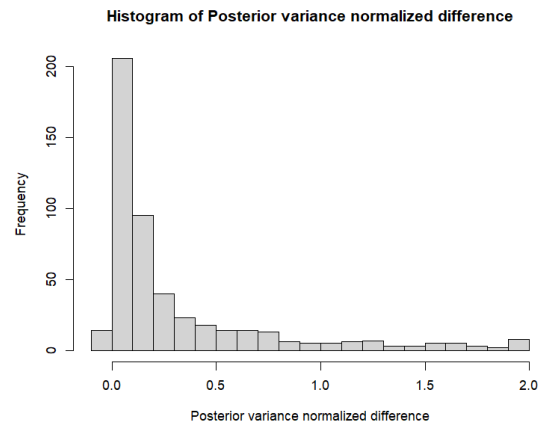


Figure 4.15: Histogram of the difference of posterior variance between parametric bootstrap and model A, normalized by the average of the variance.

The plots in figure 4.16 give further intel about the 292 combinations by plotting their observed and expected count against each other. It should be noted that although both counts having some extreme outliers most of the counts are below 100. Another result, best seen in figure 4.16b, is that the observed to expected ratio varies a lot between the combinations from far below to far above one.

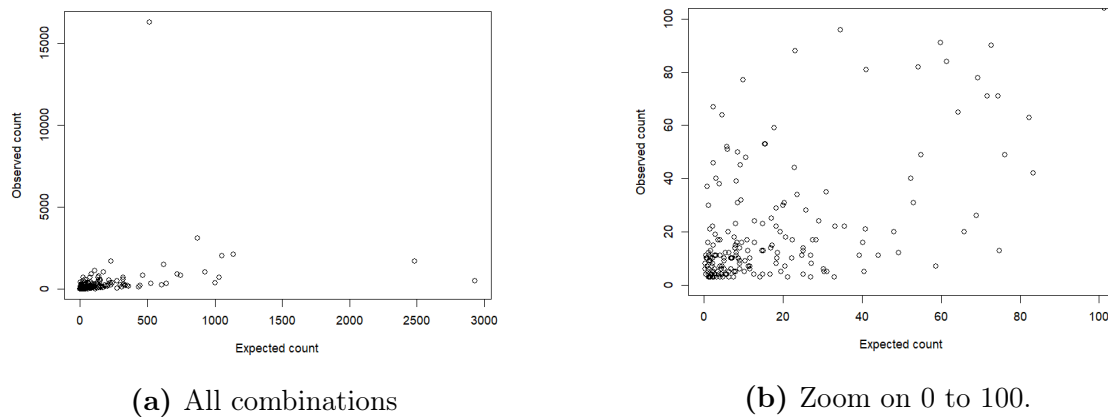


Figure 4.16: Plot of observed count against expected count for all combinations with more than 100% increase in variance compared to model A.

The tables 4.8 and 4.9 give the important counts and estimates for all 14 combinations where based on the IC025 the change in variance between model A and E results in them no longer being flagged as a signal of disproportionate reporting (because $IC025 < 1$). Remarkable is the result for the combination infiximab-blood potassium where the IC025 dropped from 3.33 in model A to 0.84 in model E.

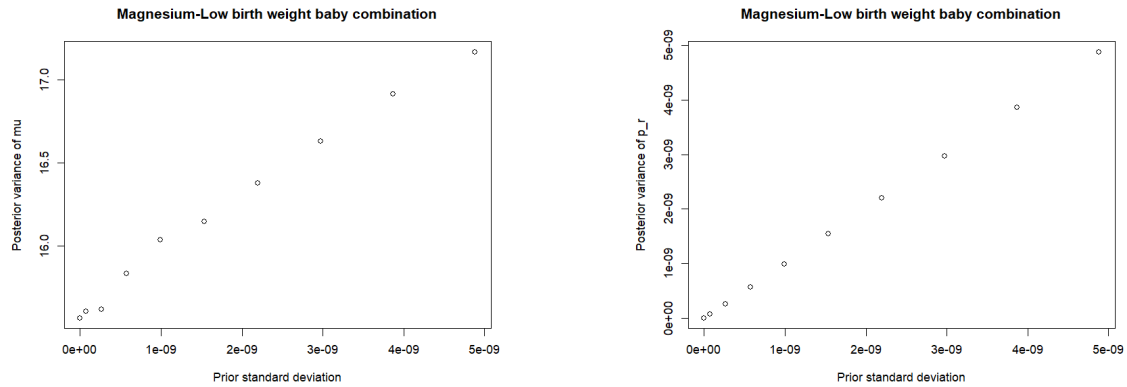
Table 4.8: Important counts for all 14 changing combinations.

Drug	Reaction	n_drug	n_reaction	O	E
Olanzapine	Drug abuser	58125	5718	29	18.29
Bromfenac	Pharyngitis	2837	19675	11	3.07
Ibrutinib	Encephalitis	29522	5115	16	8.31
Infliximab	Blood potassium	144542	212	13	1.68
Ciclosporin	Cardiomegaly	44288	8353	31	20.36
Denosumab	Feeling abnormal	111372	150891	1028	925.18
Rofecoxib	Myositis	62143	5826	30	19.93
Lamotrigine	Oropharyngeal swelling	48314	1018	11	2.70
HIB vaccine	Bradycardia	69905	47723	233	183.66
Metoclopramide	Social problem	47011	2109	12	5.45
Sacubitril;Valsartan	Visual impairment	39050	96042	252	206.47
Pirfenidone	Headache	19978	677784	806	745.47
Varicella zoster vaccine	Listless	148429	7515	84	61.40
Dimenhydrinate	Coma	4819	29793	15	7.90

Table 4.9: Important estimates for all 14 changing combinations.

Drug	Reaction	Prior sd.	Model A		Model E	
			IC	IC025	IC	IC025
Olanzapine	Drug abuser	8.06e-04	1.56	1.05	5.63	0.39
Bromfenac	Pharyngitis	1.98e-03	3.21	1.63	9.29	0.59
Ibrutinib	Encephalitis	3.19e-04	1.87	1.08	2.27	0.76
Infliximab	Blood potassium	3.37e-05	6.17	3.33	18.43	0.84
Ciclosporin	Cardiomegaly	8.53e-04	1.50	1.02	1.59	0.87
Denosumab	Feeling abnormal	7.30e-03	1.11	1.04	1.12	0.87
Rofecoxib	Myositis	5.52e-04	1.49	1.01	1.55	0.89
Lamotrigine	Oropharyngeal swelling	1.12e-04	3.58	1.82	6.27	0.95
HIB vaccine	Bradycardia	3.99e-03	1.26	1.11	1.29	0.97
Metoclopramide	Social problem	8.26e-05	2.09	1.10	2.22	0.98
Sacubitril;Valsartan	Visual impairment	6.18e-03	1.21	1.07	1.23	0.99
Pirfenidone	Headache	3.77e-02	1.08	1.00	1.09	0.93
Varicella zoster vaccine	Listless	3.36e-04	1.36	1.08	1.40	0.94
Dimenhydrinate	Coma	2.94e-03	1.84	1.04	1.94	0.93

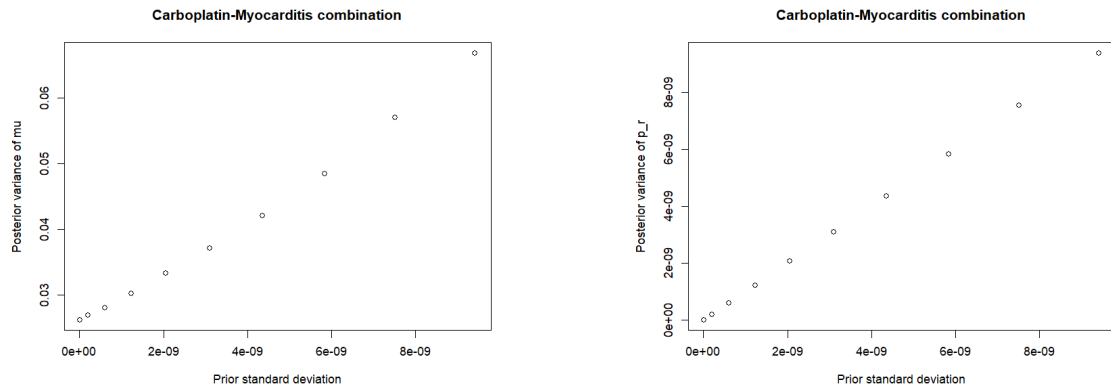
The Figures 4.17 and 4.18 show the posterior variance of μ and p_r against the prior standard deviation. Note that the values of the prior standard deviation are adjusted around the estimated standard deviation for the combination. Remarkable is that the posterior variance of μ has an upwards trend for the magnesium-"low birth weight baby combination" unlike for model D.



(a) Posterior variance of μ to prior standard deviation for one combination.

(b) Posterior variance of p_r to prior standard deviation for one combination.

Figure 4.17: Change in the posterior variance for the magnesium-low birth weight baby combination.



(a) Posterior variance of μ to prior standard deviation for one combination.

(b) Posterior variance of p_r to prior standard deviation for one combination.

Figure 4.18: Change in the posterior variance for the Carboplatin-Myocarditis combination.

Chapter 5

Discussion

5.1 E_{strat} and E_{cum}

These results indicate that E_{cum} and E_{strat} cannot be interchanged. Not only do they differ on the level of individual combinations, but they also exhibit a systematic difference when examined on average. To ensure the comparability of results between different models, the same E needs to be used consistently across all of them. The motivation to chose the stratified version is to enable the models to account for structural uncertainty between the strata.

5.2 Basic bootstrap analysis

Comparing the results in table 4.2 the difference in the variance seems to be small and thus is overshadowed by the noise from the resampling. In both models the variance is around 0.13 and the small differences are more likely due to the precision of the bootstrap method than to any real difference in variance. Considering also the width of the confidence intervals, it can be concluded that the effect of the included uncertainty in the expected has only a small effect on the variance of the IC that can not be measured with this model.

Looking at the results for the second, even rarer, combination of magnesium and "low birth weight baby" in table 4.4 we see a similar picture. The variance is not largely different between the models. If anything, the variance in the model with uncertainty in the expected value is lower than in the other model. However, this should again be interpreted as due to the precision of the bootstrap method.

Comparing the variance of the bootstrap models with the variance of model A, it is noticeable that for the very rare combination of magnesium and "low birth weight

baby", the variance is much higher in model A. This shows that for rare combinations model A already estimates a high variance and therefore adding the variance of the expected does not make a visible difference. For the other combination, the variance of model A and the bootstrap models are almost the same. This suggests that the variance of model A is strongly dependent on the size of the expected. This could mean that the model is more likely to overestimate its precision for common combinations with a large expected.

One can also observe that the precision of the bootstrap estimates for the variance is quite low relative to the size of the estimate. This could be improved by using more bootstrap samples for each run of a model (increasing the number from 1000 to 10 000 would increase the precision by one decimal). However, this would be computationally very expensive and, because the effect on the confidence interval is so small, would not produce a vastly different result for these combinations. Instead, the use of different, less computationally intensive models like model D or E could facilitate the measurement of variance differences.

The figures 4.2 and 4.3 show that the bootstrap sample has a rather small variance in the expected for each year. Even for the proportion that has more random variability between years, the bootstrap confidence intervals indicated by the grey shading are very narrow compared to how the proportion fluctuates between different years. This is likely due to the expected count being based on thousands of reports, which drives down the width of the confidence intervals, despite there being large fluctuations between years, which should be included in the intervals as well. This opens up the question whether the bootstrap model included the right uncertainty in the expected that can be found in the data. Therefore, the following models D and E were constructed to include the structural uncertainty in the expected, while also being less computationally expensive to allow greater accuracy of the variance estimators.

For this model, the results can be summarised as the IC model is robust to the inclusion of the uncertainty from this non-parametric bootstrap in the expected value. This is reassuring as otherwise the larger uncertainty found in the data would potentially cause the IC model to be even more perturbed.

5.3 Hierarchical Bayesian accounting for structural uncertainty

First, note that the change of the posterior mean of μ between model D and model A as seen in figure 4.5 is not expected, but rather a side effect of the hierarchical structure of the model as the inclusion of the prior for p_r impacts the mean and variance of the posterior. In such a model it is hard to analyse an effect on the variance without also affecting the mean. Note that therefore, the 0.025 quantile is not a good indicator of the change in variance because the quantile changes depending on the mean and variance. Instead, the focus of the analysis lies more on the posterior variance, although in practice the 0.025 quantile is used to evaluate the combinations making it also interesting.

Looking at the three figures 4.4a, 4.4b and 4.4c together with the table 4.6 we can see that the effect of the uncertainty from the prior has very different effects on the posterior distribution for different combinations. For example, for the combination of fentanyl and nausea, the posterior variance increases from 0.0008 in model A to 0.0192 in model D. This is an increase by the factor 24 and also has an impact on the quantile, reducing it by 0.12. This can also be seen in figure 4.4a where the two densities are noticeably different. For the combination of lornoxicam and angioedema, the change in variance is not quite as large, as it only increased by 0.0165 from 0.1128 to 0.1293, which is an increase by factor 1.15. Therefore, the two densities in figure 4.4b are more similar than for the fentanyl-nausea combination, which is also reflected by the change of only 0.03 in the quantile. For the combination of magnesium and "low birth weight baby", the variance even decreased slightly in model D. This is an interesting observation to which we will return later.

Overall, the analysis of these three combinations suggests that the effects of including uncertainty in the expected can be very different and can depend on several parameters such as O and n_d . To better understand possible trends in these effects, the sample of 100 combinations is analysed.

Looking at the difference in posterior variances between the models in 4.6 we see that the median of the difference is with 0.107 below what would be considered of practical importance in decision making. In almost all the combinations that do differ, the variance tends to be larger in model D. This is what one would expect as this model incorporates more uncertainty. However, for 10 combinations model D actually has a lower variance. Note that this happens despite the convergence of the Stan model, which is suggested by both diagnostic metrics \hat{R} and n_{eff} . Because

$\hat{R} = 1$ and n_{eff} is well above 1000. The decrease in variance effect can be seen in the two histograms 4.7, and 4.8. Especially the one with the normalised difference shows that for 10% of the combinations the variance is reduced. 90% have an increased posterior variance, which also tends to increase more.

Let us take a closer look at a combination representative of the 10% of combinations where the variance decreased after including more uncertainty in the model. For more information on these cases, see figure 4.9, where the posterior variance of both parameters μ and p_r is plotted for different prior standard deviations of the beta prior. An example of a combination with decreasing variance is found in figure 4.9, and it can be seen that as the standard deviation of the prior, or in other words the uncertainty added to the model, increases, the variance of the posterior density of μ decreases. This shows that the model is behaving strangely for certain combinations. Because what you would expect is the increasing trend as you can see in figure 4.10 for a combination with increasing variance. Both figures also show in the second plot that the posterior density of the parameter p_r is increasing, as one would expect.

A possible explanation for this decreasing variance is that the beta prior in this hierarchical Bayesian model is informed by the data when building the posterior distribution of p_r . This causes the mean and variance of the posterior distribution of p_r to differ from the parameters specified for the beta prior. The degree to which they differ depends on many parameters such as O, n_d, α and β . The exact way in which the different parameters affect this is complicated and hard to deduce because it is analytically intractable. A couple of results that go in this direction can be found in the appendix. However, no clear conclusions could be drawn from them. The detailed analysis of the interaction of the multiple parameters would go beyond the scope of this thesis and is left as an open question for further research.

5.4 Parametric bootstrap accounting for structural uncertainty

The results of this model are presented using similar plots and tables as for the hierarchical bootstrap model. In general, one can say that the results of model E align quite well with those of model D. Therefore, the discussion will focus on pointing out the similarities and differences. For a more detailed analysis of some of the plots, please refer to the discussion of model D in section 5.3.

It can be observed that this model also undergoes a change in the posterior

mean compared to model A, as illustrated in figure 4.12. Consequently, the primary indicator for the analysis of variance must be the posterior variance itself. The IC025 quantile however is still the preferred metric to judge if a combination should be flagged as a signal of disproportionate reporting as this is how it is done in practice.

The results for the three combinations presented in figures 4.11a, 4.11b, and 4.11c are nearly identical to those of model D. One notable difference, however, is that the variance exhibited an increase for all combinations. For the magnesium-"low birth weight baby" combination the difference means that now the variance of model E is larger than of model A. In the case of the fentanyl-nausea and lornoxicam-angioedema combinations, this difference is minimal and does not have an impact in practice.

The most striking results from this model are from the histogram in figure 4.15 which reveals that 14% of the 1998 combinations increase their variance compared to model A by more than 100%. This is a drastic increase that as seen in figure 4.16 affects very different combinations from rare to frequently reported ones. Also the observed to expected ratio varies indicating IC values below and above one. This suggests that for some combinations, with an IC025 above 1 for model A, the increase in variance can cause the IC025 to drop below 1 for model E. 14 of these combinations exist in this sample and are listed in the tables 4.8 and 4.9. For 9 of them the difference in IC025 is less than 0.2. For the other 5 combinations the difference is larger and would probably change the interpretation in practice. The most extreme case is the combination of infliximab-blood potassium which with an IC025 value of 3.33 in model A was interpreted as a "stronger" signal of disproportionate reporting. In model E it has a IC025 value of 0.84 which is interpreted as no signal.

Note that for model E, there exist combinations that have a lower variance compared to model A. However, the magnesium-"low birth weight baby" combination is not a unique case, as while in model D, 10% of the combinations exhibited a lower variance compared to model A, in model E, only 0.8% of the combinations fell into that category.

A significant distinction between this model and model D is that no combination exhibits the peculiar phenomenon where the variance in the posterior becomes lower for increasing prior variance. As an illustration, consider the figures 4.17 and 4.18 where the difference becomes visible in figure 4.17a. This reinforces the suspicion that the cause for the strange behavior is the influence of the data on the beta prior because this effect is not present in model E.

Chapter 6

Conclusion

On the way to answering how the uncertainty in the expected changes the precision of the IC model this thesis introduced three different models: the bootstrap model (model B and C), hierarchical Bayesian model (model D), and the parametric bootstrap model (model E). All of them have their limitations but because of their differences together allow for some conclusion that can answer the question of the impact of the uncertainty in the expected and the robustness of the IC model with respect to this uncertainty.

The first bootstrap model suffered from inaccuracy due to its extreme computational expense. In fact the accuracy was too low to draw any conclusion about the change of the variance of the IC. But through the analysis of this model a form of structural uncertainty in the expected was discovered even though the model failed to include this structural uncertainty. This led to the idea of using GAMs to estimate the structural uncertainty as described in section 3.5. This estimate of the uncertainty of E was used in the other two models. The hierarchical Bayesian model which used MCMC methods from the Stan package was computationally more efficient than the bootstrap model and reached better accuracy and allowed the analysis of more combinations. This model gave the first useful results about the impact on the variance of the IC. But it also had the unwanted features that the estimated uncertainty of E is informed by the data which complicated the interpretation of the model for certain combinations. The parametric bootstrap model disallowed the influence of the data on the uncertainty of E while also reducing the computational complexity compared to model D. This means that the model E reached good accuracy and allowed analysis for a large sample of combinations.

Before drawing the conclusions from the results a few limitations of the models should be addressed. The results have shown that many parameters influence the

variance of the IC and therefore also the change in the variance when including uncertainty in the expected. This makes the interpretation of the results complicated because the same change in variance of the IC can have different causes depending on the combination. For example could a large change in variance for one combination be caused by a large estimate of the uncertainty in the expected. For another combination this estimate is small but it experiences the same large change in the variance because other parameters like the observed O or n_d amplify the uncertainty in E stronger. Another difficulty of interpreting the results is that the quantile of the IC models is not a good measure for the influence of the added uncertainty because they reflect changes in the mean and variance of the model. The effects of the mean and variance cancel each other out because both tend to increase and an increase in the mean makes the 0.025 quantile go up and an increase in the variance makes it go down. This means that the quantile is not a good indicator of the change of the variance of the model. But because in practice the 0.025 quantile is the estimate used to screen the combinations it is still used to compare the models.

The general conclusion from the results is that across the different models the increase in variance for the IC is not drastic and for the majority of combinations will not make a difference to the screening. However, for individual combinations, the inclusion of the uncertainty of the expected makes a more substantial difference, such that the outcome of the screening could be altered. To answer the research question, it can be concluded that the current IC model is quite robust to uncertainty in the expected. Of the three models discussed, the parametric bootstrap model in particular could be an improvement on the current model because it incorporates a structural uncertainty in the expected value that is present in the data. Furthermore, it seems possible to use this model on a larger scale for all combinations because of its rather low computational requirements. However, further research on this model is needed to better understand how the uncertainty affects different combinations.

In the following I list a few aspects of the models that could be of interest for further research.

- Include uncertainty in n_d together with the uncertainty in n_r or replacing the uncertainty in n_r by the one in n_d .
- Closer analysis of the fit of the GAM on combination level.
 - How the GAM fit can be improved to catch all trends by adapting the model to each combination.

- A focused analysis of combinations whose IC is in the critical region around 0, where changes in variance are more likely to affect the outcome of the screening.
- Analysis of interactions/relationships between different parameters in the model D and E and their effect on the change in variance (n_d , O , size of E , and size of n_r).

Bibliography

- [Car+17] Bob Carpenter et al. “*Stan* : A Probabilistic Programming Language”. In: *J. Stat. Soft.* 76.1 (2017). ISSN: 1548-7660. DOI: [10.18637/jss.v076.i01](https://doi.org/10.18637/jss.v076.i01). URL: <http://www.jstatsoft.org/v76/i01/>.
- [Cou10] Council for International Organizations of Medical Sciences, ed. *Practical aspects of signal detection in pharmacovigilance: report of CIOMS Working Group VIII*. OCLC: ocn676726220. Geneva: CIOMS, 2010. 143 pp. ISBN: 978-92-9036-082-7.
- [DuM99] William DuMouchel. “Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System”. In: *The American Statistician* 53.3 (1999). Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 177–190. ISSN: 0003-1305. DOI: [10.2307/2686093](https://doi.org/10.2307/2686093). URL: <https://www.jstor.org/stable/2686093>.
- [Efr79] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *The Annals of Statistics* 7.1 (Jan. 1979). Publisher: Institute of Mathematical Statistics, pp. 1–26. ISSN: 0090-5364, 2168-8966. DOI: [10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552). URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-7/issue-1/Bootstrap-Methods-Another-Look-at-the-Jackknife/10.1214/aos/1176344552.full>.
- [Has70] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: [10.1093/biomet/57.1.97](https://doi.org/10.1093/biomet/57.1.97). eprint: <http://biomet.oxfordjournals.org/cgi/reprint/57/1/97.pdf>. URL: <http://biomet.oxfordjournals.org/cgi/content/abstract/57/1/97>.
- [HG14] Matthew D. Homan and Andrew Gelman. “The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo”. In: *J. Mach. Learn. Res.* 15.1 (Jan. 1, 2014), pp. 1593–1623. ISSN: 1532-4435.

- [HN12] Johan Hopstadius and G. Niklas Noren. “Robust discovery of local patterns: subsets and stratification in adverse drug reaction surveillance”. In: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*. IHI '12. New York, NY, USA: Association for Computing Machinery, Jan. 28, 2012, pp. 265–274. ISBN: 978-1-4503-0781-9. DOI: [10.1145/2110363.2110395](https://doi.org/10.1145/2110363.2110395). URL: <https://doi.org/10.1145/2110363.2110395>.
- [Lag+20] Olof Lagerlund et al. “WHODrug: A Global, Validated and Updated Dictionary for Medicinal Information”. In: *The Innov Regul Sci* 54.5 (Sept. 1, 2020), pp. 1116–1122. ISSN: 2168-4804. DOI: [10.1007/s43441-020-00130-6](https://doi.org/10.1007/s43441-020-00130-6). URL: <https://doi.org/10.1007/s43441-020-00130-6>.
- [Met+53] Nicholas Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: 21.6 (June 1953), pp. 1087–1092. DOI: [10.1063/1.1699114](https://doi.org/10.1063/1.1699114).
- [Mon+11] Jean-Louis Montastruc et al. “Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database”. In: *Brit J Clinical Pharma* 72.6 (Dec. 2011), pp. 905–908. ISSN: 0306-5251, 1365-2125. DOI: [10.1111/j.1365-2125.2011.04037.x](https://doi.org/10.1111/j.1365-2125.2011.04037.x). URL: <https://bpspubs.onlinelibrary.wiley.com/doi/10.1111/j.1365-2125.2011.04037.x>.
- [Nea11] Radford M. Neal. *MCMC using Hamiltonian dynamics*. May 10, 2011. DOI: [10.1201/b10905](https://doi.org/10.1201/b10905). arXiv: [1206.1901](https://arxiv.org/abs/1206.1901)[physics,stat]. URL: <http://arxiv.org/abs/1206.1901>.
- [NHB13] G. Niklas Noren, Johan Hopstadius, and Andrew Bate. “Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery”. In: *Stat Methods Med Res* 22.1 (Feb. 2013), pp. 57–69. ISSN: 1477-0334. DOI: [10.1177/0962280211403604](https://doi.org/10.1177/0962280211403604).
- [Nor03] Niklas Noren. “A Monte Carlo method for Bayesian dependency derivation”. MA thesis. Chalmers University of Technology, 2003.
- [Sea+16] Suzie Seabroke et al. “Performance of Stratified and Subgrouped Disproportionality Analyses in Spontaneous Databases”. In: *Drug Saf* 39.4 (Apr. 1, 2016), pp. 355–364. ISSN: 1179-1942. DOI: [10.1007/s40264-015-0388-3](https://doi.org/10.1007/s40264-015-0388-3). URL: <https://doi.org/10.1007/s40264-015-0388-3>.

-
- [WHO] WHO. *What is Pharmacovigilance?* URL: <https://www.who.int/teams/regulation-prequalification/regulation-and-safety/pharmacovigilance> (visited on 06/08/2024).
- [Woo17] Simon N. Wood. *Generalized Additive Models: An Introduction with R, Second Edition*. 2nd ed. New York: Chapman and Hall/CRC, May 3, 2017. 496 pp. ISBN: 978-1-315-37027-9. DOI: [10.1201/9781315370279](https://doi.org/10.1201/9781315370279).
- [ZM19] Silvelyn Zwanzig and Behrang Mahjani. *Computer Intensive Methods in Statistics*. Dec. 2, 2019. ISBN: 978-0-367-19423-9. DOI: [10.1201/9780429202322](https://doi.org/10.1201/9780429202322).

Appendix A

Appendix

In this appendix a few further results are presented that indicate the difficulty of analysing the posterior variance in the models D and E because of the influence of multiple parameters. What means that the interpretation of the following plots is difficult and does not allow for clear conclusions on the relation of the parameters.

The figures [A.1](#) and [A.2](#) are for the 100 combinations in model D. In Figure [A.1](#) the logged posterior variance of the parameter μ is plotted against the logged standard deviation of the beta prior for all of the 100 combinations. This is done to show if a higher prior variance results in a higher posterior variance. But an according trend is not visible in the plot. Hence, suggesting that other parameters influence the posterior variance of μ as well.

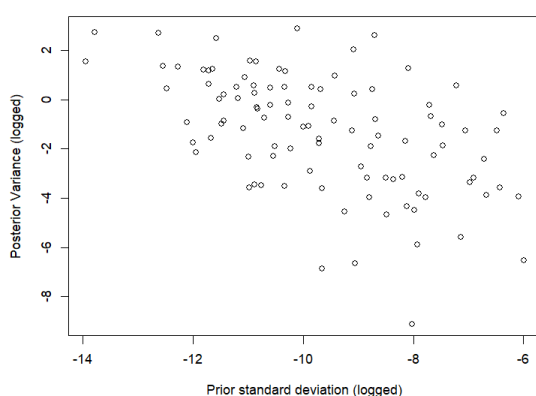


Figure A.1: Posterior variance of μ to standard deviation of the beta prior.

The Figure [A.2](#) plots the posterior variance of μ against the observed count O on the x-axis. This should give an insight into possible relationships between the variance of μ and the frequency of the combination. The observed O is taken

as an indicator of the frequency of the combination instead of E because it is not influenced by the model. But the plot suggests that no relationship between the frequency and variance of μ exists.

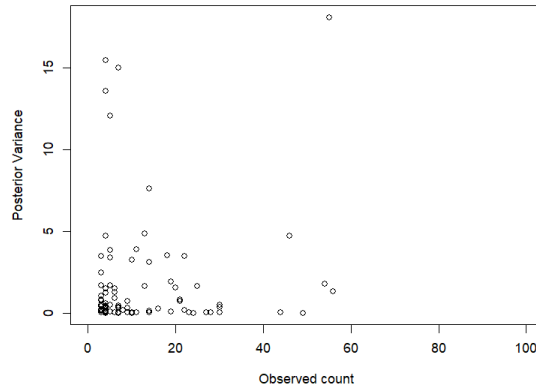


Figure A.2: Posterior variance of μ to observed count O .

Both figures A.3 and A.4 are for the sample of 1998 combinations in model E. They reveal the same patterns as the two plots for model D and lead to the same conclusions. Figure A.3 illustrates the logarithm of the posterior variance of the parameter μ plotted against the logarithm of the standard deviation of the beta prior for all combinations. This comparison aims to determine whether a higher prior variance corresponds to a higher posterior variance. However, no recognisable trend is evident in the plot.

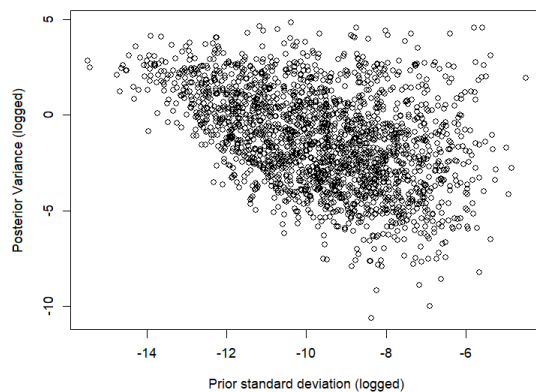


Figure A.3: Posterior variance of μ to standard deviation of the beta prior.

The Figure A.4 plots the posterior variance of μ against the observed count O on the x-axis. This should give an insight into possible relationships between

the variance of μ and the frequency of the combination. The observed O is taken as an indicator of the frequency of the combination instead of E because it is not influenced by the model. But the plot suggests that no relationship between the frequency and variance of μ exists.

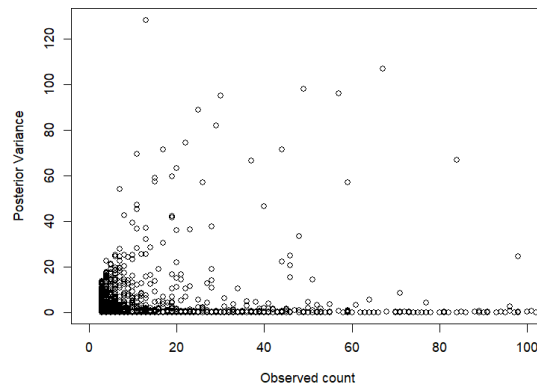


Figure A.4: Posterior variance of μ to observed count O .

Looking at Figure A.1 and 4.6, there is no tendency for the posterior variance to increase as the prior variance increases, unlike one might expect. Instead we see that the majority of combinations have a small posterior variance.