Appropriate statistical model for count data in central statistical monitoring and application on the German Multiple Sclerosis Registry

F. Fneish[1,2], D. Ellenberger[1], N. Frahm[1], A. Stahmann[1], F. Schaarschmidt[2]

1: MS-Forschungs- und Projektentwicklungs-gGmbH, German MS Register, Hannover, Germany

2: Department of Biostatistics, Institute of Cell Biology and Biophysics, Leibniz University Hannover, Hannover, Germany

Background

Central statistical monitoring (CSM) has been a commonly practiced monitoring approach in clinical trials. CSM involves various statistical methods to identify data-related issues that can indicate problems in a trial's conduct. In previous research, we demonstrated the benefit of applying comparisons of centers to the Grand Mean of the data[1]. We previously investigated different statistical models that can implement this type of comparison for binomial, ordinal, and continuous endpoints. In this research, we further investigate whether this comparison can be applied by different subtypes of generalized linear models for count data. We demonstrate this approach on Real-World-Data (RWD) from the German Multiple Sclerosis Registry

Methods

- Bayesian generalized linear model (BayesGLM) is an alternative to classical GLM and can counteract computational problems due to observing only 0s in some centers and can account for overdispersion in count data. We additionally consider a negative binomial model, a common approach to account for overdispersion in count data. In a simulation study, we investigate whether these models can control type I error when comparing centers to the Grand Mean (GM)
- Monte Carlo simulations (1000 data sets) were run for balanced and unbalanced scenarios covering a range of settings that could be found in clinical trials in different centers. Additionally, a random exposure period following the exponential distribution for counts was considered as an offset. The simulations aimed to detect whether both models can control type I error (only unbalanced simulations are shown)
- For a given model with estimated parameters (on the log link) m_i for each center $i=1,\ldots,I$ and possibly unbalanced sample sizes n_i, the Grand Mean $m_.$ can be computed by $m_. = \sum_{i=1}^{I} \frac{n_i}{N} m_i$
- Comparisons of parameters m_i to the Grand Mean $m_.$ can then be written as a set of $k=1,\ldots,K$ linear contrasts, with contrast coefficients $c_k = (c_{k1}, c_{k2}, c_{k3}, \ldots c_{kI})$

$$c_1 = (1 - \frac{n_1}{N}, -\frac{n_2}{N}, \frac{n_3}{N}, \ldots, -\frac{n_I}{N})$$

$$c_2 = (-\frac{n_1}{N}, 1 - \frac{n_2}{N}, \frac{n_3}{N}, \ldots, -\frac{n_I}{N})$$

- The deviations d_k of each center from the Grand Mean can then be written as: $d_k = m_{i=k} - m_. = \sum_{i=1}^{I} c_{ki}\, m_i$
- Simultaneous confidence intervals adjusted for multiple comparisons are computed using the methods of Hothorn et al. (2008)
- The prior assumption on parameters in BayesGLM prevents standard errors from becoming extremely wide. Cauchy distribution with the assumption $10^{-9} < \pi_i < 1\text{-}10^{9}$ was chosen

Results

Simulation of negative binomial GLM, BayesGLM & Classical GLM (quasipoisson)

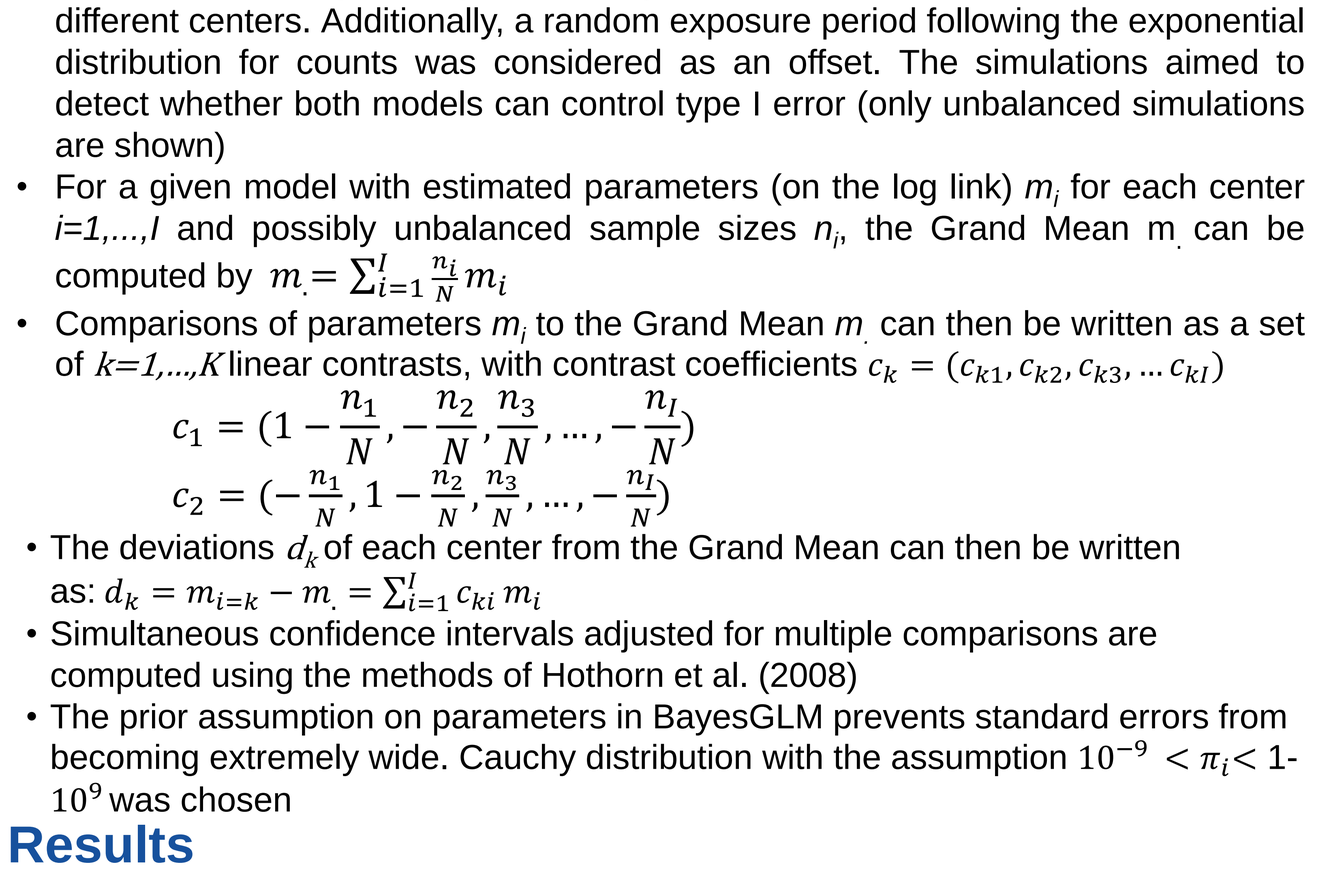

Figure 1: The probability of falsely rejecting the null hypothesis for at least one center as a function of n, mu and phi using the negative Binomial Model, Bayesian GLM and classical GLM. The nominal type I error rate (α = 0.05) is shown as a horizontal line. Overdispersion parameter phi was chosen following quasipoisson for BayesGLM and GLM. Size parameter was chosen as the inverse of Kappa where kappa=mu*phi, mu (mean), GLM (Generalized linear model)

- Each simulation consisted of 10 centers with one deviating center
- The number of patients in the unbalanced center was half of the number of patients in other centers
- With small phi, the negative binomial has a better control of type I error for all simulated scenarios
- With increasing phi, BayesGLM and GLM have similar and better control than the negative binomial model for small sample sizes

Application on the German MS Register data

- Data from 43 centers that recruited patients in the Pharmacovigilance Module of the GMSR
- Contrasts of center means with Grand Mean
- 95% confidence interval allowing statistical inference

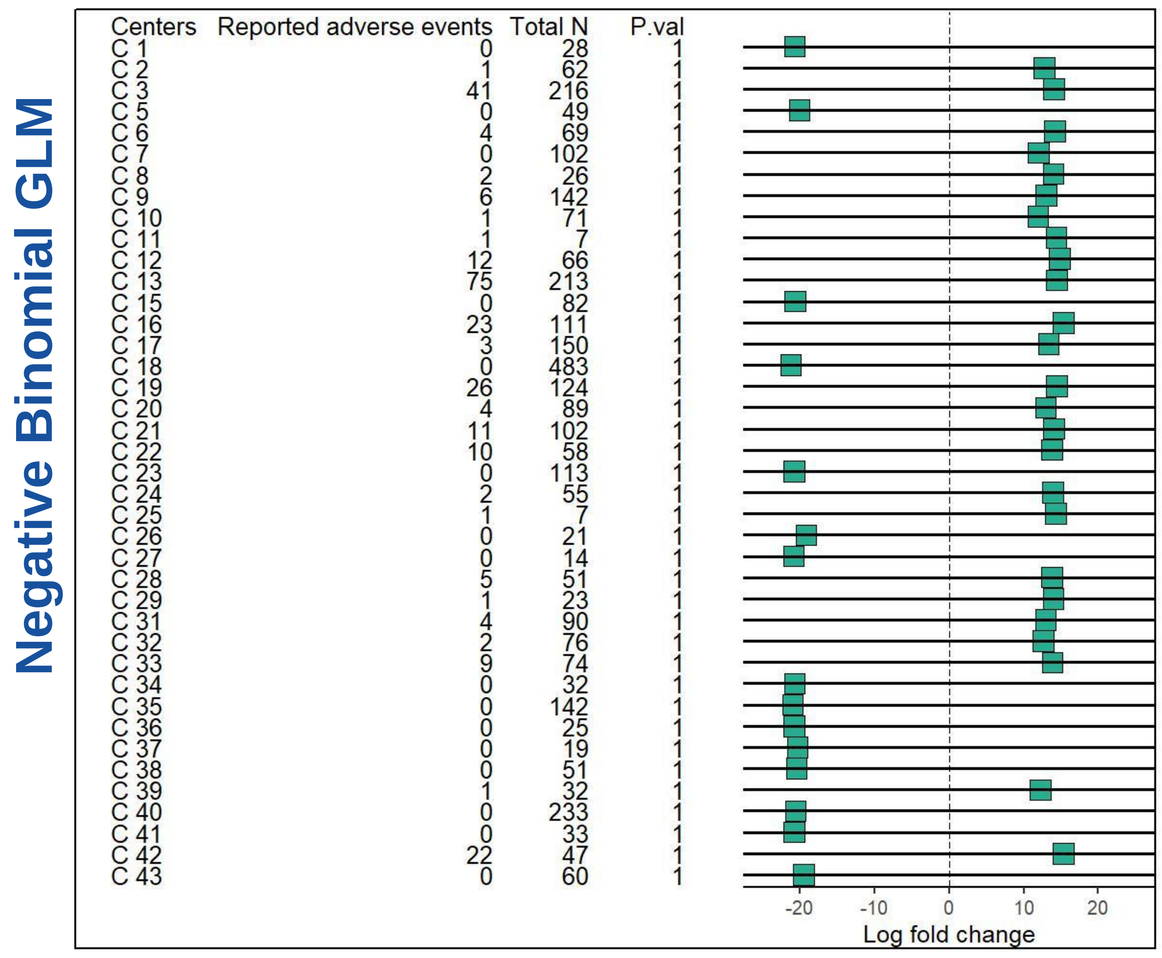

Figure 2: Application of mean comparisons of Adverse Events rates to GM using the Negbin. Simultaneous confidence intervals for contrasts of center means with GM for the GMSR. Extremely wide simultaneous confidence limits and failure to detect any deviation are due to observing 0 AEs in some centers. GM (Grand Mean)

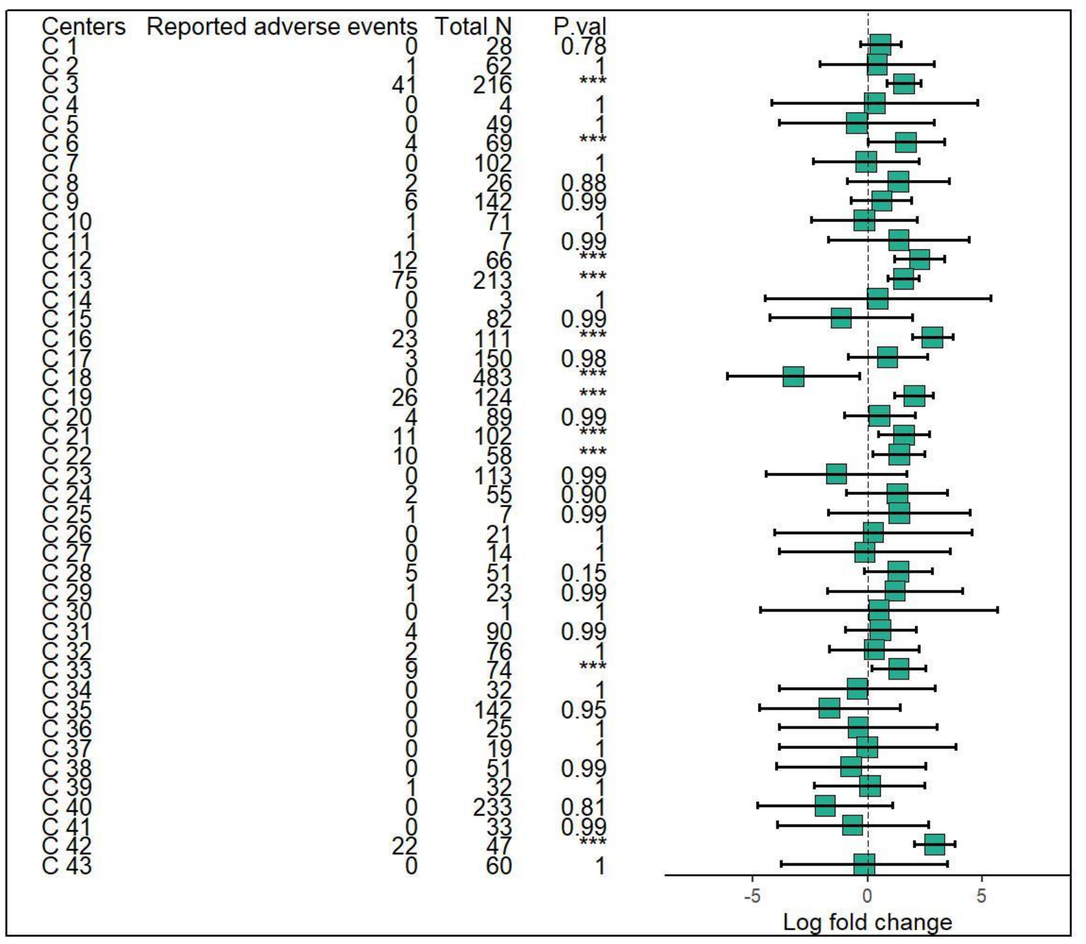

Figure 3: Application of mean comparisons of Adverse Events rates to GM using the BayesGLM. Simultaneous confidence intervals (credible intervals) for contrasts of center means with GM for the GMSR. GM (Grand Mean)

- Negative binomial model (also GLM) computationally fail to perform the comparison due to inflated standard errors as a result of 0 presence in the data

Conclusions

- Utilize the negative binomial model for the comparisons to GM for count data in the absence of 0 counts and overdispersion, and BayesGLM in the presence of 0 counts and/or presence of overdispersion
- All considered methods tend to show inflated type-I-error when sample sizes are as small as 10, 20 or 40 per center, and counted events are rare and have clear overdispersion

References

1. Fneish, F., Ellenberger, D., Frahm, N. et al. Application of Statistical Methods for Central Statistical Monitoring and Implementations on the German Multiple Sclerosis Registry. Ther Innov Regul Sci (2023). https://doi.org/10.1007/s43441-023-00550-0
2. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008. https://doi.org/10.1214/08-AOAS191.full
3. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biom J* 2008; 50: 346–363. https://doi.org/10.1002/bimj.200810425
4. Ohle, LM., Ellenberger, D., Flachenecker, P. et al. Chances and challenges of a long-term data repository in multiple sclerosis: 20th birthday of the German MS registry. Sci Rep 11, 13340 (2021). https://doi.org/10.1038/s41598-021-92722-x