

c212: An R Package for the Detection of Safety Signals in Clinical Trials Using Body-Systems (System Organ Classes)

Raymond Carragher^{1, 2, 4} and Chris Robertson^{2, 3}

1 Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK **2** Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK **3** Health Protection Scotland, NHS National Services Scotland, Glasgow, UK **4** Health Data Research (UK), University of Strathclyde, Glasgow, UK

DOI: [10.21105/joss.02706](https://doi.org/10.21105/joss.02706)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Charlotte Soneson](#) ↗

Reviewers:

- [@rrrlw](#)
- [@MelvinSMunsaka](#)

Submitted: 09 September 2020

Published: 04 December 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Safety in clinical trials may be characterised by the incidence or occurrence of adverse events. The statistical analysis of this data is complicated by the large number of adverse events recorded, with low event rates, small effect sizes and low power all contributing to the difficulty in determining a robust safety profile for a treatment during the trial process.

In addition to end of trial analyses, a number of interim analyses may take place at different time points during the trial lifecycle. These offer the additional statistical challenge of testing accumulating data, with possibly differing recruitment rates on trial arms contributing to a lack of balance in the data.

Adverse events are typically defined by medical dictionaries, which provide a common reference terminology for use in and between clinical trials. There are a number of medical dictionaries in current use, all of which provide similar services. One such dictionary is [MedDRA](#) (Medical Dictionary for Regulatory Activities), which was developed by the [ICH](#) (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) and is widely used by regulatory bodies, clinical research organisations (CROs), and pharmaceutical companies. WHO-ART (World Health Organisation Adverse Reaction Terminology) is a similar dictionary maintained by the [Uppsala Monitoring Centre](#) for the [World Health Organisation Collaborating Centre for International Drug Monitoring](#). MedDRA and WHO-ART have a similar hierarchical structure consisting of System Organ Classes (SOC) and various grouping and descriptor terms.

The MedDRA hierarchical structure consists of five levels: System Organ Class (SOC), High Level Group Terms (HLGT), High Level Terms (HLT), Preferred Terms (PT), and Lower Level Terms (LLT). The PT is a single medical description of a symptom or observation while the LLT is how a patient or data recorder would describe a symptom or observation. Each LLT belongs to one PT and, in general, data will be recorded at the LLT level but reported at the PT level (the adverse event). As of 2020 there are 27 SOC and over 80,000 LLTs.

The grouping of adverse events by SOC (or body-system) provides for possible relationships between the adverse events within a SOC. One consequence of this is the possibility that, for treatments which may affect a particular SOC, there may be raised rates for a number of adverse events within that SOC. A number of methods have recently been proposed to address the statistical issues in adverse event analysis by using these groupings of adverse events by body-system or SOC, taking into account the additional information provided by these relationships to increase the power of detecting real adverse event effects. These methods,

which include both error controlling procedures for multiple hypothesis testing ([Benjamini & Hochberg, 1995](#); [Hu et al., 2010](#); [Matthews, 2006](#); [Mehrotra & Adewale, 2012](#); [Yekutieli, 2008](#)), and Bayesian modelling approaches ([Amy Xia et al., 2011](#); [Berry & Berry, 2004](#); [Carragher, 2017b](#)), are implemented in the R package c212 (Table 1).

Table 1: Methods in the c212 package.

Method	Description
Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995)	Control of False Discovery Rate
Group Benjamini-Hochberg procedure (Hu et al., 2010)	Control of False Discovery Rate
Double false discovery rate (Mehrotra & Adewale, 2012)	Control of False Discovery Rate
Subset BH-procedure (Yekutieli, 2008)	Control of False Discovery Rate
Bonferroni correction (Matthews, 2006)	Control of Familywise Error Rate
Berry and Berry model (Berry & Berry, 2004)	Bayesian model for end of trial data
Berry and Berry model without point-mass (Amy Xia et al., 2011)	Bayesian model for end of trial data
Interim analysis model (Carragher, 2017b)	Bayesian model for interim trial data
Interim analysis model without point-mass (Carragher, 2017b)	Bayesian model for interim trial data

Statement of Need

The detection of safety issues in the post-marketing phase of a treatment's life cycle, as opposed to the trial phase, can have a serious effect on the health of patients and also a financial impact both for the companies developing the treatments, and the regulatory bodies responsible for overseeing them.

The R package c212 provides a self-contained set of methods for clinical trial safety investigators, statisticians and researchers, to aid in the early detection of adverse events. It is designed to be easy to use with a simple data input format and interface.

The primary use case for the software is in the statistical analysis of adverse event incidence and occurrence data during clinical trials. A second goal of the package is to provide reference implementations of the methods in Table 1 for use by researchers, both in the area of safety in clinical trials, as well those developing or testing methods for handling error rates when testing multiple hypotheses. Beyond safety in clinical trials, the package will be useful to any project which deals with multiple hypothesis testing, or projects where two groups of comparative data may be modelled by hierarchical Bayesian binomial or Poisson models, with recent extensions of the Bayesian models to observational data being developed ([Carragher et al., 2020](#)).

The c212 package is currently being used both for clinical trial safety analysis ([Fries et al., 2016](#); [Munsaka, 2018](#); [Wang et al., 2018](#)) and as a research tool in the investigation and development of new safety methods ([Tan et al., 2019, 2020](#)). It has also been implemented as part of an [ASA Biopharm Safety Working Group](#) workstream.

Overview

The Bayesian models, under assumptions of conditional independence, are fitted using a Gibbs sampling Markov Chain Monte-Carlo (MCMC) method ([Robert & Casella, 1999](#)). The

posterior distributions of the model parameters are used to assess which adverse events may have increased rates on the treatment arm. In the case of the Berry and Berry model, which is binomial, the `theta` model parameter, representing the increase in the log-odds of an event occurring on the treatment arm, is used for this purpose (Berry & Berry, 2004). For the interim analysis models, which are Poisson based, the increase in the log rate of an event on the treatment arm is used for adverse event assessment. As in the Berry and Berry model this is represented by the parameter `theta` (Carragher, 2017b). Functions for generating summary statistics and highest probability intervals are provided using the services of the `coda` package (Plummer et al., 2006). The main convergence diagnostics available directly within the package are the Gelman-Rubin and Geweke statistics (Gelman et al., 2004), again from the `coda` package. Access to the raw samples is available for further processing should that be required. The error controlling procedures included in the package follow exactly the method definitions in the papers which introduced them (Benjamini & Hochberg, 1995; Hu et al., 2010; Matthews, 2006; Mehrotra & Adewale, 2012; Yekutieli, 2008). The following sections contain examples which cover the main uses of the software. The data sets and functions used are fully documented in the package.

Multiple Hypothesis Testing

The data set `c212.FDR.data` contains the results from a set of multiple hypothesis tests for adverse events grouped by body-system. The Group Benjamini-Hochberg procedure is applied to the data with control of the False Discovery Rate set at the 5% level. The `c212.err.cntrl` function, which provides the interface to the error controlling methods, returns the set of p-values/hypotheses deemed significant.

```
library(c212)
data(c212.FDR.data, package="c212")
head(c212.FDR.data, 2)
B j AE p
1 Bdy-sys_5 7 Adv-Ev_24 1.000000
2 Bdy-sys_6 9 Adv-Ev_34 0.358949

p.sig <- c212.err.cntrl(c212.FDR.data, method = "GBH", alpha = 0.05)
print(p.sig)
```

	B	j	AE	p	p_weighted
1	Bdy-sys_3	6	Adv-Ev_911	0.000079	0.000000
2	Bdy-sys_3	3	Adv-Ev_8	0.000021	0.000000
3	Bdy-sys_3	1	Adv-Ev_6	0.000000	0.000000
4	Bdy-sys_3	5	Adv-Ev_910	0.000039	0.000000
5	Bdy-sys_3	2	Adv-Ev_7	0.000011	0.000000
6	Bdy-sys_3	4	Adv-Ev_9	0.010411	0.000000
7	Bdy-sys_3	7	Adv-Ev_912	0.003554	0.000000
8	Bdy-sys_2	1	Adv-Ev_2	0.005333	0.005333
9	Bdy-sys_2	2	Adv-Ev_3	0.016013	0.016013

Berry and Berry End of Trial Analysis

The data set `c212.trial.data` contains sample end of trial adverse event incidence counts. The data is modelled using the Berry and Berry model as follows:

```
library(c212)
data(c212.trial.data, package="c212")
head(c212.trial.data, 2)
```

```
B j AE Group Count Total
1 Bdy-sys_2 1 Adv-Ev_2 1 20 450
2 Bdy-sys_2 4 Adv-Ev_5 2 21 450
```

```
mod.BB <- c212.BB(c212.trial.data)
```

mod.BB contains the raw samples generated from the model fitting procedure. To perform a convergence check:

```
conv.BB = c212.convergence.diag(mod.BB)
c212.print.convergence.summary(conv.BB)
```

In order to assess which adverse events may be associated with treatment the function `c212.ptheta` is used. This calculates the posterior probability of an increase in log-odds of an event occurring on the treatment arm. A threshold may be used to view the adverse events which exceed some defined level, for example: 0.90:

```
theta.pos_BB <- c212.ptheta(mode.BB)

theta_pos_BB[theta_pos_BB$ptheta > 0.9,]
B AE ptheta
2 Bdy-sys_2 Adv-Ev_2 0.9410417
6 Bdy-sys_3 Adv-Ev_6 1.0000000
7 Bdy-sys_3 Adv-Ev_7 1.0000000
8 Bdy-sys_3 Adv-Ev_8 1.0000000
9 Bdy-sys_3 Adv-Ev_9 0.9873500
10 Bdy-sys_3 Adv-Ev_910 0.9999333
11 Bdy-sys_3 Adv-Ev_911 0.9999917
12 Bdy-sys_3 Adv-Ev_912 0.9964917
```

The high posterior probabilities may indicate a possible association with treatment for these adverse events.

Interim Analysis

Apart from the function used to fit the model, the procedure for fitting and accessing interim analysis data is exactly the same as for the Berry and Berry model.

```
library(c212)
data(c212.trial.interval.data1)
head(c212.trial.interval.data1, 2)
Interval I_index      B      AE Group Count Exposure
0.0-180.0      0 Bdy-sys_1 Adv-Ev_1      1      87 160133.7
0.0-180.0      0 Bdy-sys_1 Adv-Ev_1      2     103 163224.6
mod.BB.interim <- c212.BB.interim(c212.trial.interval.data1)
```

Software Details and Availability

The `c212` package was initially released to CRAN in 2017 and has been through a number of release cycles. Before each release a full set of unit and functional tests are performed on the package development system, including memory checks with `valgrind` (Seward & Nethercote,

2005) and Google address sanitizer (Serebryany et al., 2012). The package documentation also contains tests and examples based on data included in the package.

The c212 package is most easily downloaded and installed directly from CRAN (Carragher, 2017a) or, alternatively, from the corresponding GitHub repository (Carragher, 2020).

The authors are interested in extending the software to include new methods, particularly in the area of safety analysis, and would welcome collaborations in this area. Any support issues or questions can be addressed directly to the corresponding author, through the associated CRAN maintainer email address, or through the Github repository.

Performance

The Bayesian models are expensive to fit in terms of both computation and memory. The main issue is the number of parameters in the model. For the Berry and Berry model with N adverse events, and B SOCs there are $(2 \times N + 5 \times B + 6)$ parameters in total. With C parallel MCMC chains and I iterations of the MCMC sampler, this will require space for $C \times I \times (2 \times N + 5 \times B + 6)$ double precision numbers to store the samples. For larger datasets, the calculation of the convergence diagnostics and summary statistics may be time consuming as the samples are passed to the coda package.

Example: A trial with 23 SOCs and 497 AEs, 5 parallel chains and 40,000 iterations after burn-in, will require space for 133,800,000 doubles, equating to approximately 1GB of memory storage for the samples alone (assuming 8 bytes per double).

It is recommended that at least twice the storage required for the samples be available for fitting any of the Bayesian models.

Acknowledgements

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) UK (award reference 1521741) and Frontier Science (Scotland) Ltd.

References

- Amy Xia, H., Ma, H., & Carlin, B. P. (2011). Bayesian hierarchical modeling for detecting safety signals in clinical trials [Journal Article]. *Journal of Biopharmaceutical Statistics*, 21(5), 1006–1029. <https://doi.org/10.1080/10543406.2010.520181>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing [Journal Article]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berry, S. M., & Berry, D. A. (2004). Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model [Journal Article]. *Biometrics*, 60(2), 418–426. <https://doi.org/10.1111/j.0006-341X.2004.00186.x>
- Carragher, R. (2017a). c212: Methods for detecting safety signals in clinical trials using body-systems (system organ classes). In *CRAN repository*. CRAN. <https://CRAN.R-project.org/package=c212>
- Carragher, R. (2017b). *Detecting safety signals in randomised controlled trials* [PhD thesis]. University of Strathclyde.

- Carragher, R. (2020). c212: Methods for detecting safety signals in clinical trials using body-systems (system organ classes). In *GitHub repository*. GitHub. <https://github.com/rcarragh/c212>
- Carragher, R., Mueller, T., Bennie, M., & Robertson, C. (2020). A Bayesian hierarchical approach for multiple outcomes in routinely collected healthcare data. *Statistics in Medicine*, 39, 2639–2654. <https://doi.org/10.1002/sim.8563>
- Fries, M., Kracht, K. K., Li, J., Munsaka, M. S., Sanchez-Kam, M., Singh, K. P., Wang, W., Whalen, E., & Zhou, K. (2016). Safety monitoring methodology in the premarketing setting. *JSM Proceedings, Biopharmaceutical Section*, 2247–2269. <https://ww2.amstat.org/MembersOnly/proceedings/2016/data/assets/pdf/389675.pdf>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* [Book]. Chapman & Hall/CRC.
- Hu, J. X., Zhao, H., & Zhou, H. H. (2010). False discovery rate control with groups [Journal Article]. *Journal of the American Statistical Association*, 105(491), 1215–1227. <https://doi.org/10.1198/jasa.2010.tm09329>
- Matthews, J. N. S. (2006). *Introduction to randomized controlled clinical trials, second edition* [Book]. Chapman & Hall/CRC. <https://doi.org/doi:10.1201/9781420011302.fmatt>
- Mehrotra, D. V., & Adewale, A. J. (2012). Flagging clinical adverse experiences: Reducing false discoveries without materially compromising power for detecting true signals [Journal Article]. *Statistics in Medicine*, 31(18), 1918–1930. <https://doi.org/10.1002/sim.5310>
- Munsaka, M. S. (2018). A question-based approach to the analysis of safety data. In K. E. Peace, D.-G. Chen, & S. Menon (Eds.), *Biopharmaceutical applied statistics symposium : Volume 2 biostatistical analysis of clinical trials* (pp. 193–216). Springer Singapore. https://doi.org/10.1007/978-981-10-7826-2_11
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7–11. <http://CRAN.R-project.org/doc/Rnews/>
- Robert, C. P., & Casella, G. (1999). *Monte Carlo statistical methods* [Book]. Springer New York.
- Serebryany, K., Bruening, D., Potapenko, A., & Vyukov, D. (2012). AddressSanitizer: A fast address sanity checker. *Proceedings of the 2012 USENIX Conference on Annual Technical Conference*, 28.
- Seward, J., & Nethercote, N. (2005). Using valgrind to detect undefined value errors with bit-precision. *Proceedings of the Annual Conference on USENIX Annual Technical Conference*, 2.
- Tan, X., Chen, B. E., Sun, J., Patel, T., & Ibrahim, J. G. (2020). A hierarchical testing approach for detecting safety signals in clinical trials. *Statistics in Medicine*, 39(10), 1541–1557. <https://doi.org/10.1002/sim.8495>
- Tan, X., Liu, G. F., Zeng, D., Wang, W., Diao, G., Heyse, J. F., & Ibrahim, J. G. (2019). Controlling false discovery proportion in identification of drug-related adverse events from multiple system organ classes. *Statistics in Medicine*, 38(22), 4378–4389. <https://doi.org/10.1002/sim.8304>
- Wang, W., Whalen, E., Munsaka, M., Li, J., Fries, M., Kracht, K., Sanchez-Kam, M., Singh, K., & Zhou, K. (2018). On quantitative methods for clinical safety monitoring in drug development. *Statistics in Biopharmaceutical Research*, 10(2), 85–97. <https://doi.org/10.1080/19466315.2017.1409134>
- Yekutieli, D. (2008). False discovery rate control for non-positively regression dependent test statistics [Journal Article]. *Journal of Statistical Planning and Inference*, 138(2), 405–415. <https://doi.org/10.1016/j.jspi.2007.06.006>