The Journal of Open Source Software

Vigicaen: A vigibase® Pharmacovigilance Database Toolbox.

Charles Dolladille [1] **and Basile Chrétien** [2]

1 University of Caen Normandy, Pharmacology Department, Centre Hospitalier Universitaire de Caen, Caen, France **2** University of Nagoya, Department of biostatistics, Nagoya University Hospital, Nagoya, Japan

DOI: 10.21105/joss.10321

Software

- Review
- Repository
- Archive

Editor: Charlotte Soneson

Reviewers:

- @mhovd
- @pacoramon

Submitted: 25 August 2025
Published: 30 August 2026

Summary

For decades, the World Health Organization (WHO) has been collecting adverse drug reaction reports, called Individual Case Safety Reports (ICSRs), from its member countries, populating more than 40 million reports to date. This pharmacovigilance database is called VigiBase® and is managed by the Uppsala Monitoring Centre in Sweden. (Uppsala Monitoring Centre, 2026) These ICSRs describe the course of patients who experienced an adverse event (a medical condition) after taking a drug. The burning question is whether this adverse event was actually related to the drug intake, e.g., if it is an adverse drug *reaction* (ADR). A pharmacovigilance database analysis aims at uncovering the very first potential signals of association between drugs and ADRs. (Montastruc et al., 2011)

Disproportionality analysis, a method of safety signal detection, represents an essential component of pharmacovigilance. Advanced methodologies are essential when conducting disproportionality analyses, as traditional approaches are susceptible to various biases such as reporting bias and confounding.

The aim of vigicaen is to provide a toolbox for the VigiBase® Extract Case Level database, resolving technical challenges related to the large size of the database, and providing easier and reproducible access to advanced features. The package is built on top of the parquet file format. Functions related to drug and adverse event identification, and descriptive features such as time to onset, dechallenge, and rechallenge outcomes are provided. Command-line side-effect outputs aim at fast resolution of common issues related to drug and adverse event identification. The package is intended for pharmacovigilance practitioners, clinicians, and researchers with or without advanced biostatistical skills. A graphical output can be produced for routine use to support the daily assessment of drug liability.

Statement of need

Disproportionality analysis is a statistical method that produces estimators of how unlikely the number of observed ICSRs reporting on a specific drug and adverse event is to be attributable to chance alone. Together with an uncertainty margin, these estimators are used to raise safety signals on drugs. (Montastruc et al., 2011)

Advanced methodologies are required to address common biases of disproportionality analysis in pharmacovigilance databases. These analyses necessitate expertise in biostatistical software, such as R, which may present substantial challenges in terms of acquiring and maintaining the requisite skills — in addition to a solid understanding of pharmacovigilance principles and reporting systems.

The Uppsala Monitoring Centre grants access to VigiBase® to researchers, either academic

or industrial, under a license contract. The most extensive available version is called Extract Case Level: it contains all the ICSRs, with information such as patient demographics, drug and adverse events related features. However, this version is provided as large text files and requires a substantial processing prior to analysis. Those text files would often exceed the size of the available Random Access Memory, thus requiring advanced knowledge of R computing techniques. Clinicians and pharmacovigilance practitioners typically lack these skills and therefore struggle to use VigiBase® data for their research. As a result, they often rely on partial data with limited statistical modeling options, or might develop homemade biostatistics scripts that are typically used once, often left undocumented, and highly heterogeneous across research teams.

The vigicaen package aims at providing a toolbox for the VigiBase® Extract Case Level database, tackling several technical challenges to run on low-specification computers, and providing easy and reproducible access to advanced features. (Dolladille & Chrétien, 2025) This article will explain the technical choices and computing logic of the package. Examples and use cases are covered in the package vignettes, available on the package website at https://pharmacologie-caen.github.io/vigicaen/. The Uppsala Monitoring Centre, in charge of maintaining VigiBase®, was informed of the package development and kindly allowed its publication, acknowledging the potential benefit of promoting the use of VigiBase®.

State of the field

There are very few packages related to pharmacovigilance in R. Most are focused on interfacing with the Food and Drug Administration Adverse Event Reporting System (FAERS), and most attempt to visualize existing data, run basic disproportionality analyses, or perform web scraping. (Mukhopadhyay & Markatou, 2026) None of them actually allow browsing the entirety of a worldwide database such as VigiBase®, including all types of drugs and adverse events. Also, there is no existing package in the open-source community that prepares pharmacovigilance data in order to build advanced disproportionality metrics, such as machine or deep learning models. Finally, only a few of these packages are available on mainstream platforms such as CRAN. (Embry, 2016, 2018)

Research impact and significance

Our team and collaborators have already published several pharmacovigilance studies using vigicaen. (Alexandre et al., 2021; Chretien et al., 2025; Dolladille et al., 2020; Legallois et al., 2025; Minoc et al., 2025; Nishida et al., 2025) The French Network of Regional Pharmacovigilance Centers is on its way to implementing vigicaen as part of routine practice across the 31 Pharmacovigilance Centers in France. The University of Nagoya has functional routines relying on vigicaen for disproportionality analyses. Vigicaen does not compete with existing open-source tools, but rather addresses an unmet need.

Software design

Key choices were made to build vigicaen:

- Open source design, built on top of state-of-the-art practices to deal with large datasets (e.g., arrow), especially on low-specification computers, using a widespread and consistent syntax R users are familiar with (e.g., tidyverse). In line with this first point, other syntaxes like `data.table`, once at the core of the package, have now been phased out.
- Focus on the most technically challenging issues for beginners in R or biostatistics software in general.
- Consistency in function naming, expected input formats, and outputs, aligning with the tidyverse style guide. (Wickham & RStudio, 2023)

- Provision of help, e.g., messages to users in a command-line interface, to allow external checking of what is produced by the package.
- Absence of model functions implementation, except for basic disproportionality metrics. (Norén et al., 2013) Users will build datasets with vigicaen, then run any model of their choice.

Open-source software practice

The package was developed according to best practices as promoted by R Packages, 2nd edition. (Wickham & Bryan, 2023) It is accompanied by a comprehensive set of unit tests (covering 100% of the code), in-depth documentation for each function and object, and several tutorial vignettes for both newcomers and advanced users. The source code is available on GitHub.com, which is also used to submit issues and propose pull requests. It is available under the open-source CeCILL 2.1 license.

Development history

The first iteration of the package was built in 2020 as local software designed for internal use at Caen University Hospital. Later, it was called pharmacocaen and posted as a private repository on GitHub in 2022, due to intellectual property concerns with the Uppsala Monitoring Centre. After resolution of these property concerns, the package became available as a public repository on GitHub under the name vigicaen in 2024, and was accepted on CRAN in 2025. In the first versions, the package was mainly focused on performing vectorized data management so as to identify a large number of drugs and reactions in a compiled way. Handling edge cases was the main concern for several years. Then, additional features like building datasets from source files and descriptive statistics were added. Contacts were made with members of the Uppsala Monitoring Centre regarding their own work on other topics. These exchanges helped define the exact perimeter of vigicaen, as well as its potential articulation with other open-source software in the future. Also, vigicaen was discussed with end-users from pharmacovigilance centers in France, which led to the development of specific functions like `vigi_routine`.

Processing vigibase® source files.

VigiBase® Extract Case Level files currently exceed 30GB once unpacked, which is too large to be loaded in-memory by mainstream readers like `read.table()` on most computers. Vigicaen relies on `parquet` files, a recent format based on open standards, supported by Arrow. (Apache Software Foundation, 2026b, 2026a; Richardson et al., 2026) Datasets remain out of memory. Various tests of vigicaen on 16GB RAM computers succeeded in processing the source files.

Below, we provide a partial example of the sourcing process, with `tb_vigibase()`. Users may refer to the vignettes for more details.

```
tb_vigibase(path_base, path_sub)
```

```
## -- tb_vigibase() ---------------------------------------------------------
##
## v All expected csv files found in `path_base` and `path_sub`
##
## i Checking for existing tables.
##
## i Creating vigibase tables.
##
## This process must only be done once per database version.
## It can take up to 30minutes.
```

The named list for inputting drug and adverse event names

The `get_*` and `add_*` functions are built with a named list as their first argument. This structure may seem a bit complex, especially for newcomers, but it allows for genuine flexibility when analysis plans increment. As an example, one may create `list(drug_group_1 = c("ipilimumab", "nivolumab"))` to automatically gather all ICSRs reporting one of these two drugs through `get_drecno()` and `add_drug()`.

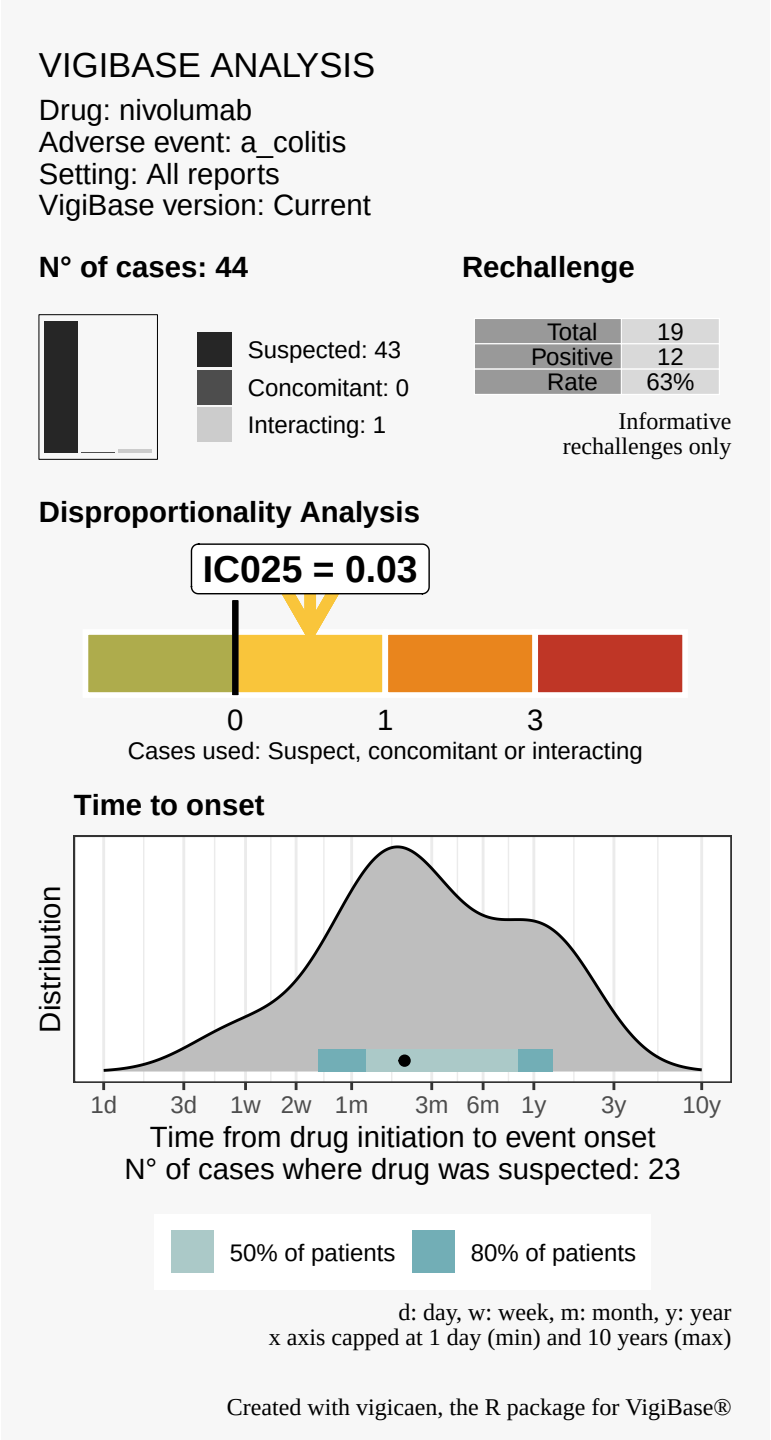
Descriptive features

Descriptive features often play an important role in pharmacovigilance studies. They can be as important as producing statistical estimates to assess the liability of a given drug. Vigicaen provides a series of functions to compute such descriptive features, spanning time to onset to drug rechallenge outcome.

Routine use

For a routine pharmacovigilance practitioner, key information on a drug-adverse event pair may be needed out-of-the-box, without further need for manipulating the underlying tables. To address typical needs (disproportionality estimates, descriptive features), `vigi_routine()` creates a graphical output for a given pair. It is intended as a daily practice tool to support routine assessment of liability. The plot can be exported to an external file.

```
vigi_routine(
  demo_,
  drug_,
  adr_,
  link_,
  d_code = ex_$d_drecno[4],
  a_code = ex_$a_llt[2],
  vigibase_version = "Current"
)
```



AI usage disclosure

GitHub Copilot and other AI assistants were episodically used during the software development. The main goals were to draft pull requests from existing issues, to assist with code syntax, and to improve the English writing in the documentation. It was especially useful for drafting variants of existing checkers or avoiding typographical errors when transforming larger sections due to architectural changes. All AI-written code was human-checked by one of the package authors before being accepted. Generative AI was used to check spelling and improve the

syntax of this manuscript.

Conclusion

Easier, reproducible research in pharmacovigilance databases is key to appropriate safety signal detection. Vigicaen proposes a set of tools based on popular open standards to facilitate pharmacovigilance analysis in R.

Acknowledgements

The information presented in this study does not represent the opinion of the Uppsala Monitoring Centre or the World Health Organization. We thank the research team at the Uppsala Monitoring Centre (Uppsala, Sweden) who provided case-level data from VigiBase®.

References

Alexandre, J., Salem, J.-E., Moslehi, J., Sassier, M., Ropert, C., Cautela, J., Thuny, F., Ederhy, S., Cohen, A., Damaj, G., Vilque, J.-P., Plane, A.-F., Legallois, D., Champ-Rigot, L., Milliez, P., Funck-Brentano, C., & Dolladille, C. (2021). Identification of anticancer drugs associated with atrial fibrillation: analysis of the WHO pharmacovigilance database. *European Heart Journal. Cardiovascular Pharmacotherapy*, *7*(4), 312–320. https://doi.org/10.1093/ehjcvp/pvaa037

Apache Software Foundation. (2026a). *Apache arrow*. https://arrow.apache.org/

Apache Software Foundation. (2026b). *Apache Parquet*. https://parquet.apache.org/

Chretien, B., Dolladille, C., Nishida, K., Aleksic, B., Alexandre, J., & L'orphelin, J.-M. (2025). Analysis of anticancer drug associated adverse reactions in depressive patients from vigibase. *Scientific Reports*, *15*(1), 45751. https://doi.org/10.1038/s41598-025-28563-9

Dolladille, C., & Chrétien, B. (2025). *vigicaen: 'VigiBase' Pharmacovigilance Database Toolbox*. https://github.com/pharmacologie-caen/vigicaen

Dolladille, C., Ederhy, S., Sassier, M., Cautela, J., Thuny, F., Cohen, A. A., Fedrizzi, S., Chrétien, B., Da-Silva, A., Plane, A.-F., Legallois, D., Milliez, P. U., Lelong-Boulouard, V., & Alexandre, J. (2020). Immune Checkpoint Inhibitor Rechallenge After Immune-Related Adverse Events in Patients With Cancer. *JAMA Oncology*. https://doi.org/10.1001/jamaoncol.2020.0726

Embry, I. (2016). *vaersNDvax: Non-domestic vaccine adverse event reporting system (VAERS) vaccine data for present*. https://doi.org/10.32614/CRAN.package.vaersNDvax

Embry, I. (2018). *Vaersvax: US vaccine adverse event reporting system (VAERS) vaccine data for present*. https://doi.org/10.32614/CRAN.package.vaersvax

Legallois, D., Da Silva, A., Alexandre, J., Milliez, P., Sabatier, R., Blanchart, K., Plane, A.-F., Font, J., Chrétien, B., & Dolladille, C. (2025). Identification of anticancer drugs associated to cancer therapy-related cardiac dysfunction: a VigiBase® disproportionality analysis. *European Heart Journal. Cardiovascular Pharmacotherapy*, *11*(5), 459–468. https://doi.org/10.1093/ehjcvp/pvaf027

Minoc, E.-M., Villain, C., Chrétien, B., Benbrika, S., Heraudeau, M., Lafont, C., Béchade, C., Lobbedez, T., Lelong-Boulouard, V., & Dolladille, C. (2025). Association between antidepressant drugs and falls in older adults: A mediation analysis in the World Health Organization's pharmacovigilance database. *Therapies*, *80*, 561–571. https://doi.org/10.1016/j.therap.2025.01.004

Montastruc, J.-L., Sommet, A., Bagheri, H., & Lapeyre-Mestre, M. (2011). Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *British Journal of Clinical Pharmacology*, *72*(6), 905–908. https://doi.org/10.1111/j.1365-2125.2011.04037.x

Mukhopadhyay, R., & Markatou, M. (2026). SurVigilance: An application for accessing global pharmacovigilance data. *SoftwareX*, *34*, 102546. https://doi.org/10.1016/j.softx.2026.102546

Nishida, K., Chrétien, B., Dolladille, C., Ebina, T., Aleksic, B., Cabé, N., Savey, V., Onoue, T., & Yatsuya, H. (2025). Psychiatric and psychological adverse effects associated with dulaglutide, semaglutide, and liraglutide: A vigibase study. *Clinical Nutrition*, *51*, 252–265. https://doi.org/10.1016/j.clnu.2025.06.011

Norén, G. N., Hopstadius, J., & Bate, A. (2013). Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Statistical Methods in Medical Research*, *22*(1), 57–69. https://doi.org/10.1177/0962280211403604

Richardson, N., Cook, I., Crane, N., Dunnington, D., François, R., Keane, J., Mecum, B., Moldovan-Grünfeld, D., Ooms, J., Wujciak-Jens, J., & Apache Arrow. (2026). *Arrow: Integration to 'apache' 'arrow'*. https://doi.org/10.32614/CRAN.package.arrow

Uppsala Monitoring Centre. (2026). *About VigiBase*. https://who-umc.org/vigibase/

Wickham, H., & Bryan, J. (2023). *R Packages: Organize, Test, Document, and Share Your Code* (2e ed.). O'Reilly Media. ISBN: 978-1098134464

Wickham, H., & RStudio. (2023). *Tidyverse: Easily install and load the 'tidyverse'*. https://doi.org/10.32614/CRAN.package.tidyverse