Writing reports of modelling and simulation analysis: Our experience in the field of pharmacometrics

Viviana Moroso, **Mats O. Magnusson**[1], **E. Niclas Jonsson**[2]
Pharmetheus AB, Uppsala, Sweden
[1] Author's ORCiD: 0000-0001-6010-9153
[2] Author's ORCiD: 0000-0002-2615-0280

doi: 10.56012/wmqy8556

Correspondence to:
Viviana Moroso
viviana.moroso@pharmetheus.com

Abstract
Pharmacometric analyses generate mathematical models that can describe and simulate the pharmacokinetics and pharmacodynamics of drugs. The role of these modeling and simulation (M&S) analyses is growing both in drug development and regulatory assessment. Reporting M&S analyses can be technically challenging given the large amount of input and output data that need to be summarised and accurately described in regulated reports. Therefore, reproducibility, automation, traceability, and standardisation are considered key aspects of this process. We present here a system that, using a combination of software, meets these challenges and improves the efficiency, accuracy, and reliability of our work.

Introduction
Pharmacometrics, an emerging field in drug development, combines information from biology, physiology, pathology, and pharmacology, into mathematical models that can quantify the interaction between drugs and patients. Modeling and simulation (M&S) of drugs' pharmacokinetics (PK) and pharmacodynamics (PD) are used to inform drug development, support regulatory assessments and trial design, and extrapolate predictions for specific populations. Altogether, this information can contribute to better patient care and support regulatory decisions. M&S analyses for regulatory submission are characterised by short timelines, large input data sets, and extensive output files; all these processes need to be tracked, organised, and interpreted in regulated reports. We describe here a reproducible reporting system developed to meet these challenges: a combination of several software programs (R, RStudio, knitr, and LaTeX) that integrates analyses and partly automates report writing.

A natural need for automation in reporting pharmacometrics results
Like many other areas of regulatory writing, the field of pharmacometrics also benefits from some level of automation and standardisation during report writing. To understand where this need comes from, we should first consider how the role of M&S analysis developed over the years.

The evolving role of modelling and simulation analysis in drug development
The role of M&S in drug development and regulatory assessment has grown in the last few decades. The benefit of using M&S is demonstrated by the integration of this type of analysis in the regulatory guidelines, as well as the creation, and continuous development, of "good practices" documents.[1-7] An extensive overview of these documents, as well as the scientific articles published on recommendations for model building and its documentation, is provided in a white paper from 2016.[8]

Given a closer look, the role of M&S in drug development has rapidly grown beyond the sole internal decision-making within pharmaceutical companies. What we nowadays call "model-informed drug development" (MIDD) is used, among other aims, to support regulatory assessments, trial design, dose selection, and extrapolation to special populations. Moreover, in some cases, the authorities have used M&S studies to approve a variation of indication even in the absence of clinical data (e.g., in paediatric studies).[9,10]

This expanding role of M&S has led to more pharmaceutical companies applying these analyses to complement their submission packages and/or to inform the subsequent phases of drug development. Pharmaceutical companies either perform these analyses in-house, when competences and resources allow for it, or request them from specialised contract research organisations (CROs).

Pharmacometrics, an emerging field in drug development, combines information from biology, physiology, pathology, and pharmacology, into mathematical models that can quantify the interaction between drugs and patients.

Why automation in reporting?
Very often, time is key for M&S analyses. When these analyses need to be performed immediately after clinical data become available, either to inform internal decision or to support regulatory submissions, results are expected within short timelines.

In pharmacometrics, not only the analysis phase but also the phase of documentation/reporting is regulated, and the produced documents need to conform to specific requirements.[8] For example, original data files, data transformations and the associated code, computation and coding of the final model and simulation files all need to be made available. Furthermore, data and results need to be shown in specific types of plots, and the validity of the developed models must be demonstrated using suitable "model diagnostics".

It goes without saying that reporting such type of analyses benefits from a clearly organised, structured, and reproducible system. By "reproducible" we here refer to a system that, if

starting from identical input data analysed with the same methods, should give the same output and lead to an essentially identical report. This implies that the final analysis report can be more in line with the internal organisational standards and less dependent on the single individual.

Tools for report-writing

There are many ways to achieve the main goals of consistency, traceability, and standardisation of reporting. The one that we chose is using a combination of:

- The statistical computing program and modelling software **R** (v4.2.2; R Core Team, 2019)[11] together with RStudio, an integrated development environment (IDE) for R, published by Posit[12]
- The document preparation system for typesetting **LaTeX**[13]
- The R package that enables integration of R code into LaTeX, called **knitr**[14-17]

These 3 tools are used in combination in the process that is described in Table 1.

This type of approach makes the whole process (after data collection) traceable and reproducible, thus complying with the principles of "reproducible research".[18,19]

Importantly, while this system needs to be solid and standardised to comply with regulatory

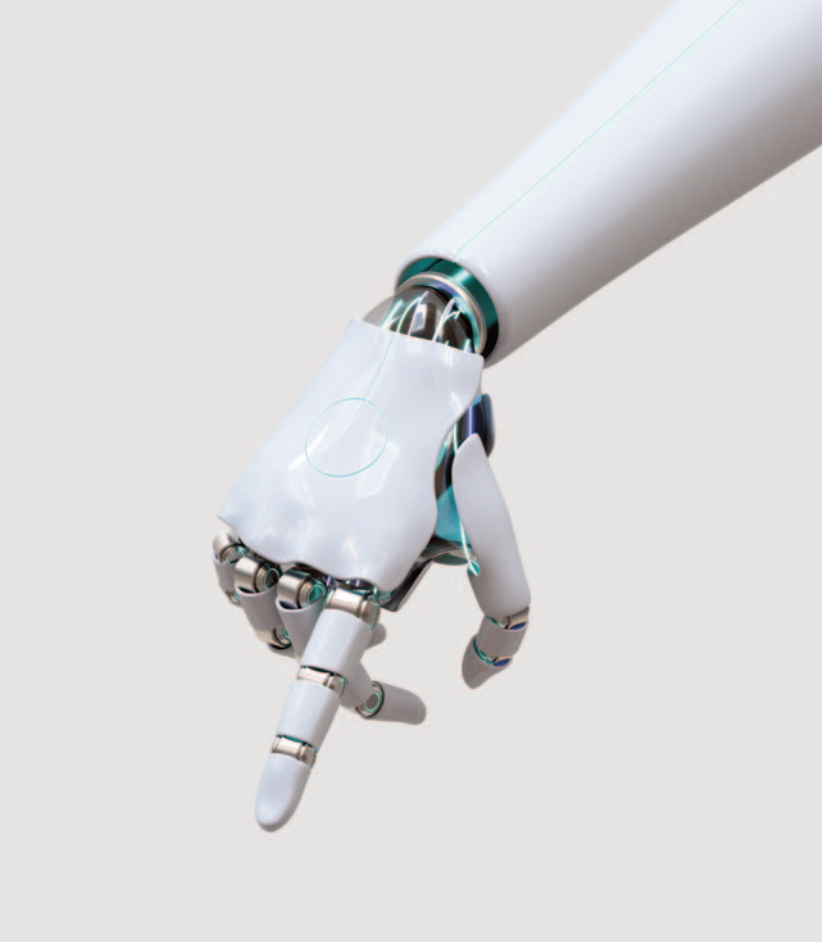


Table 1. Process of pharmacometric analysis and reporting: Typical main steps and examples of the actions that can be performed during each step

Main steps of analysis and reporting	Examples of actions performed during the step
1. Data transformation	Create a variable that groups subjects by type of underlying disease
2. Data exploration	Observe trends in the data, e.g., subjects with a specific disease type eliminate the drug faster
3. Model building	Develop a model that describes the PK profile of a drug while taking into account possible sources of variability, such as type of underlying disease, age, genotype
4. Description of the model results	Results are generated by an external software (e.g., NONMEM®, ICON plc); results are then analysed to demonstrate the validity of the models (e.g., by generating plots, such as visual predictive checks) and to draw conclusions on the endpoint analysed
5. Simulations	If results are further used to perform simulations, this allows making predictions of how a drug is expected to behave, e.g., in specific patient populations
6. Creation of submission-ready reports	Reports are generated, peer reviewed, and quality controlled; when finalised, the report and all supporting documentation are included in an e-submission package that is ready for regulatory submission

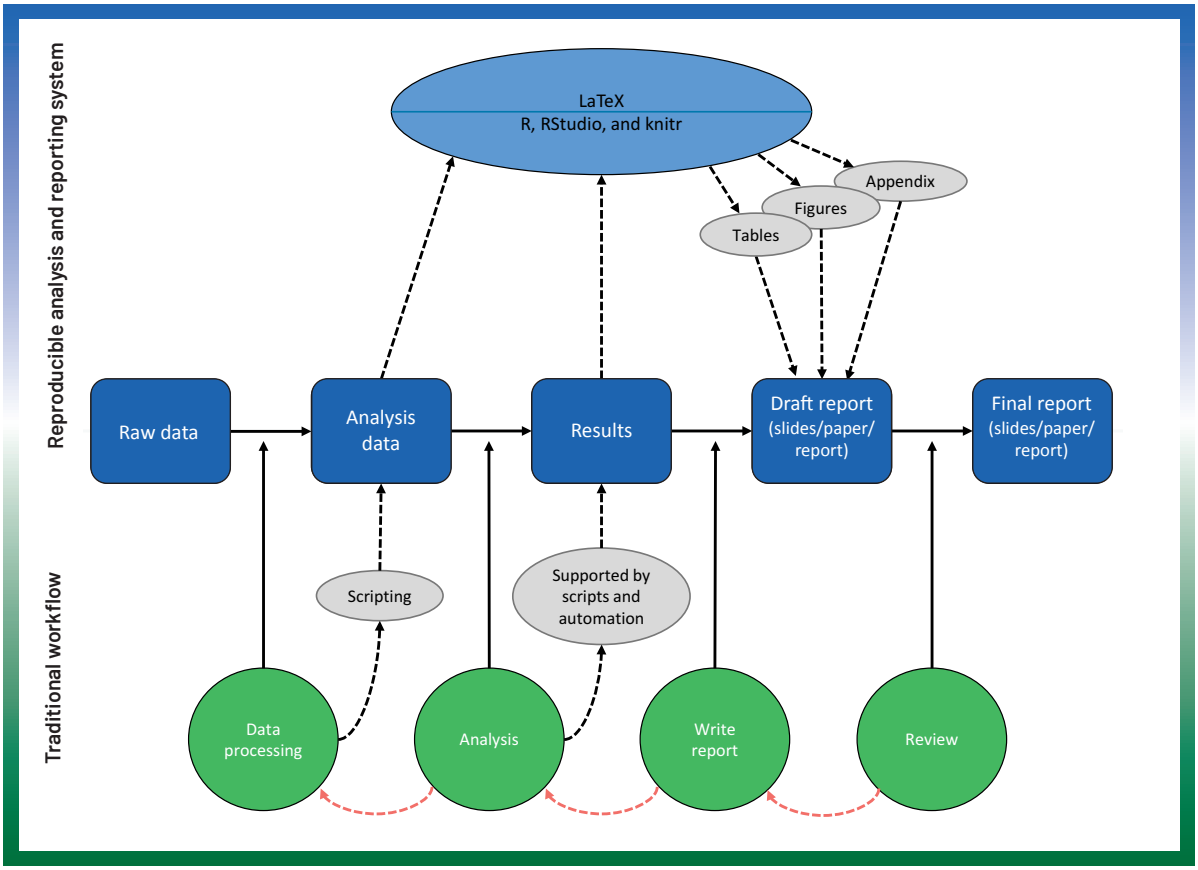


Figure 1. Analysis and reporting workflow

The blue boxes (central part of the figure) represent the typical workflow of pharmacometric analysis, from access to raw data to the phase of reporting the performed analysis. The lower part of the figure (green circles) describes the traditional workflow, in which analysis and reporting consist of consecutive steps and where review feedback (red dashed arrows) needs to be implemented manually for every single step. The upper part of the figure (light-blue ellipse) describes the tools used in the reproducible analysis and reporting system that we describe in this paper. The integration of analysis and report generation creates a seamless chain between raw data and final report. Incorporation of review feedback and correction of data errors is done in one place and then automatically propagated throughout the report.

requirements, it also needs to be sufficiently fit-for-purpose and flexible so as to adapt to the specific type of analysis (e.g., a PK-PD analysis concerning an oral-delivered drug and its active metabolites or an intravenously-infused drug, a PK-PD analysis of a time-to-event endpoint, or an endpoint measured as a continuous variable).

Setting up such a system and defining all its technical details, as well as creating user-friendly instructions for each step, requires the collaboration of a multidisciplinary team (pharmacometricians, system developers, data programmers, medical writers, quality control reviewers, etc.). Despite its technical complexity, once set up, this system is rather straightforward to use. The process makes use of R and its literate programming capabilities:[20] according to this principle, the system not only delivers a user-friendly PDF document, but is also more robust and easily maintained. With regards to the latter, when pitfalls are identified by users and when new methodologies or software updates are released, the system can be updated and refined. In this way, new versions of the system can be released, where standard code is adapted and dependencies across programmes are revised.

> The reporting system represents the point in which scientific analysis, automation/ scripting, and medical writing meet.

Brief description of the system

When a CRO performs pharmacometric analysis for a pharmaceutical company, a typical project starts with discussions with the client about the objectives, project planning, and definition, to reach an agreement on the analysis plan. When data from clinical studies become available, large data files, possibly also in different file formats, are delivered by the client to the CRO. These data are explored and transformed to create data files that can be read and used as input by a modelling software (such as NONMEM®, ICON plc)[21] (Figure 1). Pharmacometricians then analyse the data, develop models that appropriately describe the data, and possibly perform simulations in accordance with the purpose of the analysis (Table 1).

The hands-on process starts with specific input files and generates large amounts of output files and output data, in different formats, that should be summarised and interpreted. Therefore, already during the analysis, modelers need to gradually put all this information

Case study

This case study exemplifies the advantages of shifting the time spent on report writing to the early analysis phase. A pharmaceutical company requested our CRO to perform pharmacometric analyses on data from a phase III study of a drug used for cardiovascular diseases. Our company performed much of the work during the preparation phase: planning the analysis in detail, creating data files of dummy data, generating a dummy report with simulations to prepare for several alternative study outcomes, and performing scientific review and QC of the analysis and the report. As soon as the clinical study was completed and final study data were made available, scientists could spend time on actual data analysis rather than on extensive writing and editing of the report. This resulted in a 7-week turnaround time from final data access to regulatory submission of the M&S report.

together in a clear, structured, and understandable way (Figure 1). By the end of the analysis, this bundle of information needs to be organised in a report that should not only be consistent with the CRO's and clients' standards, but also conform with the content and quality requirements imposed by the regulatory authorities.

To support this process, the reporting system represents the point at which scientific analysis, automation/scripting, and medical writing meet. From a user perspective, the reporting system appears as consisting of 3 main "blocks" for each report section:

- **Instructions** on which specific information to include, how to include it, and which output files to append, allowing to deliver a "structured content";
- Section-specific verified **scripts** and functions that generate standard figures and tables in accordance with regulatory requirements (of note, R scripting uses R-packages validated to comply with good clinical practices);
- **Standard text** to help describe and clarify methods and processes typically used; in addition, optional standard text is provided to describe the most common alternative analyses.

The full product of this reporting system is a few hundred-pages-long PDF document ready for regulatory submission. However, the system can also generate shorter reports if leaner documents are better fit for the specific purpose. For such alternative cases, the system we developed allows tailoring the length and the subsections of the report to the client's requests. This can be done before report writing begins by selecting the specific document template and the type of analysis. This way, reports of different sizes or with specific subsections, as well as slide decks, can easily be produced using the same, flexible reporting system.

Advantages

The advantages of this reporting system are related, on the one hand, to the more "technical" aspects of the process and, on the other hand, to the characteristics of the final document.

The most evident technical advantage is probably the fact that, despite using several different file formats as input, the product of this system is a PDF document. This is often the format that clients prefer for final reports. Besides, a PDF is convenient since it does not allow accidental modifications and can easily be signed with official e-signing software. Additional advantages of this system are the type of software involved (LaTeX+ R + RStudio + knitr), which are open source, and thus available to everyone at no cost. Furthermore, these software programs are not specific to pharmacometrics, and thus can profit also by developments in other fields. In addition, these programs can handle large and complex technical documents. Finally, RStudio offers an environment that integrates code for statistical analysis and regular text for document preparation (Figure 2). In simple words, when RStudio receives the command to compile a PDF, it will automatically:

- Execute the R code and replace it with the appropriate LaTeX code;
- Typeset the LaTeX document into a PDF;
- Update the bibliography numbering and references list (according to the information in a file named bibtex);
- Update the glossary and correctly include all abbreviations according to the company standards.

> In pharmacometrics, not only the analysis phase but also the phase of documentation/reporting is regulated, and the produced documents need to conform to specific requirements.

Another advantage is that, as already pointed out earlier, this system allows for reproducibility of reporting and traceability of data sources, data transformation, and analysis. Additionally, instead of having to type all the content manually and having to create plots and figures from scratch for each new project, the use of validated R-scripts is more efficient and much less error-prone. With coded content, possible errors can be efficiently corrected by changing a value (or code, filename, or directory) only once in the master document: code dependencies generate a cascade of changes that will automatically propagate the corrected item in the rest of the report. These are all factors that are known to reduce the overall time and money spent on performing, writing, and QC reviewing these analyses.[22] Other authors have previously emphasised the effects of using automation tools to accelerate document writing, in some cases also quantifying them in terms of time saved.[23,24]

Another factor contributing to the efficiency of this system is that reports can be prepared even before clinical trial data are accessible and before final models are generated. In a sort of "preparation phase", the main analyses can be performed (e.g., using dummy data or previously published data from a similar study), and the report can be shaped (and already partly QC reviewed in the relevant sections). Then, when final study data become available, one can focus most of the effort on the outcome-related aspects that need to be interpreted and reported. This preparation phase allows shifting the time spent on report writing from the final and time-critical stages to an earlier and more convenient phase of the analysis. The case study described in the box gives an example of the possible time gain.

Another advantage of this system is the possibility to deliver "structured content", i.e., a document in which information is placed in the appropriate section. This specific aspect has been defined by some authors as a labour-intensive and "the most time-consuming, tedious task" for a medical writer.[23,25] Besides the use of standard text and instructions, another tool contributing to this is given by the use of technology (e.g., scripts) that retrieves information from different parts of the document (or even separate documents) and combines it in the appropriate

Input

```
\section{Demonstration of  \LaTeX and \knitr}

\LaTeX is a markup language that allows authors to focus on the content of their
document while the system takes care of formatting and layout. With simple commands
, text formatting such as \textbf{bold font} and \textit{italics} can be applied.
\LaTeX syntax can also be seamlessly combined with program code, such as the
statistical program R. This can be achieved through inline equations, such as
writing "10 + 10 = \Sexpr{10 + 10}" or more complex code chunks as illustrated
below.

<<Rcode,results='asis',fig.env="figure",fig.cap="This is a figure based on the Iris
data set",fig.pos='H'>>=

ggplot(data=iris,mapping=aes(x=Petal.Length,y=Petal.Width))+
  geom_point(aes(color=Species,size=Sepal.Width))+
  xlab("Petal lenght") + ylab("Petal width") +
  labs(size = "Sepal width")+
  theme(legend.position="top")

@
```

Output

1 Demonstration of LaTeX and knitr

LaTeX is a markup language that allows authors to focus on the content of their document while the system takes care of formatting and layout. With simple commands, text formatting such as **bold font** and *italics* can be applied. LaTeX syntax can also be seamlessly combined with program code, such as the statistical program R. This can be achieved through inline equations, such as writing "10 + 10 = 20" or more complex code chunks as illustrated below.

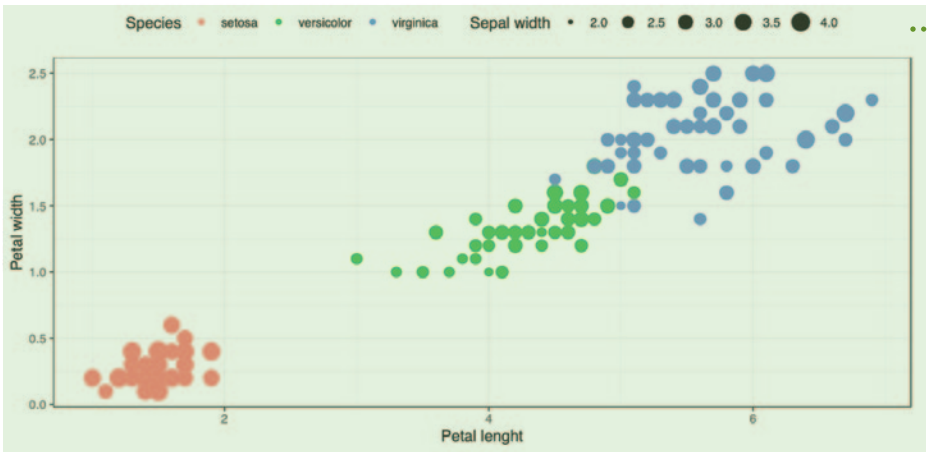


Figure 1: This is a figure based on the Iris data set

Figure 2. Example showing how the tools in the reproducible reporting system (R, RStudio, LaTeX, and knitr) allow the integration of the analysis and reporting processes.
On the upper panel (input), an example of LaTeX syntax (in blue) and R code (in green); on the bottom panel, the output PDF document that includes the respective text generated by LaTeX syntax (in blue) and the figure generated by the R code (in green).

sections, with only minor manual adaptations needed.

From a more linguistic point of view, an additional advantage is that by using guiding text and standard sentences for alternative scenarios, the formatting, tone, and language style are more consistent across scientists; this, additionally, supports professionals that are less focused on the linguistic aspects of reporting. Consequently, less time needs to be spent by a medical writer rephrasing entire paragraphs to adjust language style and formatting.

Limitations

The main drawbacks of this reporting system relate to the technical complexity of the tools that are employed. The less IT-skilled professionals may find the interface rather unfamiliar and somewhat "archaic". An example of this complexity is linked to the use of LaTeX instead of Microsoft (MS) Word as software for document preparation. Although most users may be largely familiar with MS Word, this software does not allow the programmatic integration of text, plots, tables, and abbreviations generated with standard code, yet this can be done using LaTeX, in combination with R and knitr. Of note, when developing this system, we also considered using R Markdown (RStudio, PBC) instead of LaTeX, where R Markdown is more user-friendly; however, R Markdown may not produce PDF documents meeting all the requirements of regulatory agencies, and thus LaTeX remained the preferred software. A second example of technical complexity is the need to understand most of the R code that generates tables and figures so as to be able to adapt it when certain functionalities need to be modified (e.g., a standard plot shows the subjects' median drug concentrations over time, but the client requests showing the mean values instead). A final example of complexity is related to the phase of report revision and finalisation: the document generated by the system is in PDF format instead of MS Word. Once again, users may be more familiar with the review functions in MS Word, implying that adding comments and revisions in a PDF document may require some training. All these technical aspects, together with the need to learn and adapt to the company-specific standards of analysis, lead to a steep learning curve for those using this system.

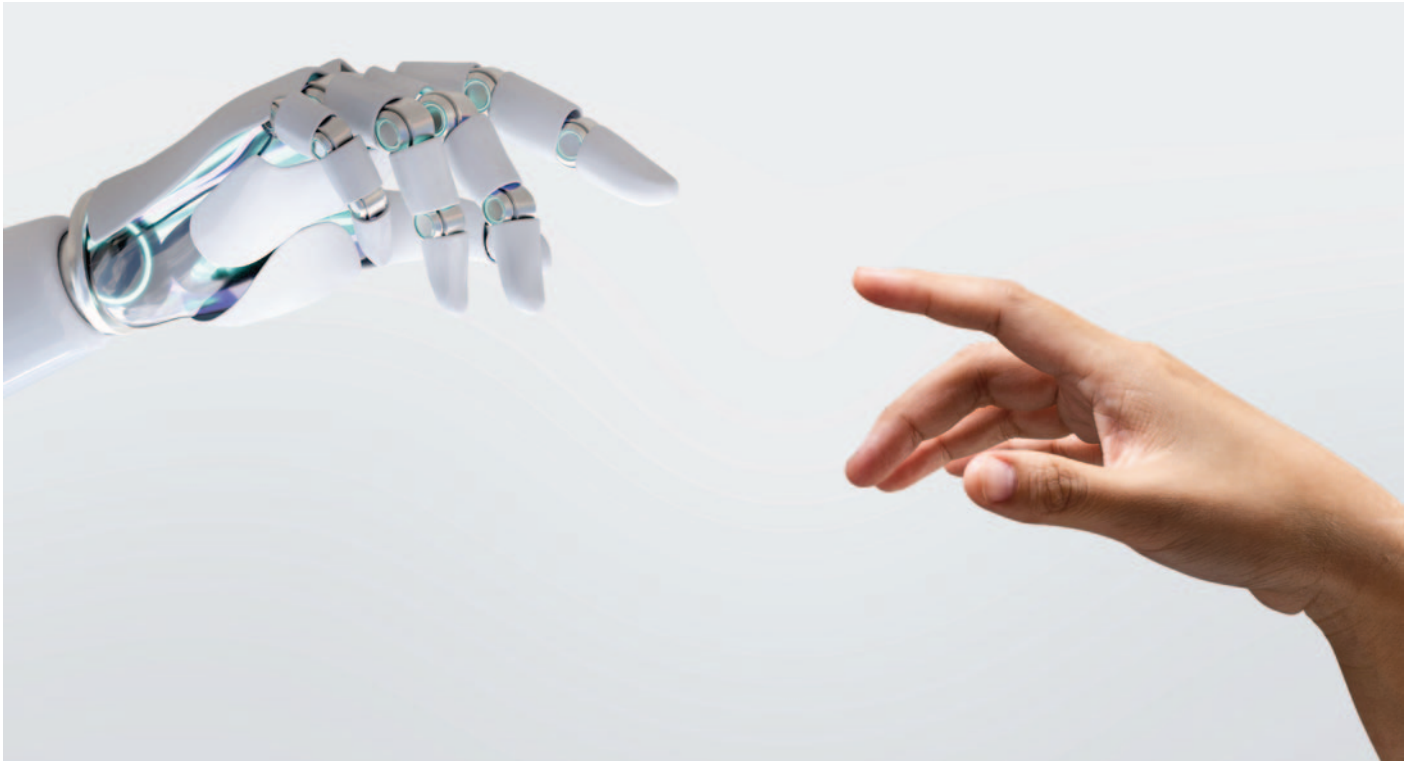
Another technical disadvantage is that all the tools listed (R, RStudio, LaTeX, and knitr) need to be correctly integrated into the existing IT environments in use at the company. Furthermore, any process adaptations or program updates need to be compatible with the rest of the system.

Regarding the final document delivered by this system, the main drawback is probably the fact that the resulting report is more template-oriented than project-oriented. In practical terms, this implies that sometimes particular client requests or project-specific needs may require additional effort to implement.

Finally, a challenging aspect is that the whole system needs to be accurately installed, so as to protect business confidentiality, information security, and access to confidential regulatory documents.[22]

Conclusions

In conclusion, despite not being free from challenges, the reporting system that we developed has increased the efficiency, accuracy, and reliability of our work. Moreover, in line with the principles driving our analysis within pharmacometrics, this reporting system contributes to the reproducibility, automation, traceability, and standardisation of our deliverables.



Acknowledgements

The authors would like to acknowledge all (current and previous) Pharmetheus AB colleagues who contributed to the development of this reproducible reporting system. The authors thank Raquel Billiones, Editor-in-Chief of *Medical Writing*, for reviewing this article.

Disclaimers

The opinions expressed in this article are the authors' own and not necessarily shared by their employer or EMWA.

Disclosures and conflicts of interest

The authors are current employees at Pharmetheus AB, where the reproducible reporting system discussed in this article was developed. The authors have no connection or affiliation with the companies that created or developed any of the software used in this reporting system.

References

1. Incorporating economic evaluation. Developing NICE guidelines: the manual. Guidance. NICE. 2014 [cited 2023 Mar 15]. Available from: https://www.nice.org.uk/process/pmg20/chapter/incorporating-economic-evaluation
2. Research Center for Drug Evaluation. Exposure-Response Relationships – Study Design, Data Analysis, and Regulatory Applications. FDA, 2020 [cited 2023 Mar 15]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/exposure-response-relationships-study-design-data-analysis-and-regulatory-applications
3. Research Center for Drug Evaluation. Population Pharmacokinetics. FDA, 2022 [cited 2023 Mar 15]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/population-pharmacokinetics
4. EMA. Reporting the results of population pharmacokinetic analyses – Scientific guideline. EMA, 2018 [cited 2023 Mar 15]. Available from: https://www.ema.europa.eu/en/reporting-results-population-pharmacokinetic-analyses-scientific-guideline
5. Dykstra K, Mehrotra N, Tornøe CW, et al. Reporting guidelines for population pharmacokinetic analyses. J Pharmacokinet Pharmacodyn 2015;42(3):301–14. doi:10.1007/s10928-015-9417-1
6. Salmonson T, Rane A. Clinical pharmacokinetics in the drug regulatory process. Clin Pharmacokinet. 1990;18(3):177–83. doi:10.2165/00003088-199018030-00001
7. Manolis E, Rohou S, Hemmings R, et al. The role of modeling and simulation in development and registration of medicinal products: output from the efpia/ema modeling and simulation workshop. CPT Pharmacometrics Syst Pharmacol. 2013;2(2):31. doi:10.1038/psp.2013.7
8. Marshall SF, Burghaus R, Cosson V, et al. EFPIA MID3 Workgroup. Good practices in model-Informed drug discovery and development: Practice, application, and documentation. CPT Pharmacometrics Syst Pharmacol. 2016;5(3):93–122. doi:10.1002/psp4.12049
9. Cablivi EM. EMA. 2018 [cited 2023 Mar 15]. Available from: https://www.ema.europa.eu/en/medicines/human/EPAR/cablivi
10. Bergstrand M, Hansson E, Delaey B, et al. Caplacizumab Model-based dosing recommendations in pediatric patients with acquired thrombotic thrombocytopenic purpura. J Clin Pharmacol. 2022; 62(3):409–21. doi:10.1002/jcph.1991
11. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria. R Foundation for Statistical Computing, 2019 [cited 2023 Mar 28]. Available from: https://www.r-project.org/
12. RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC. Boston, Mass., 2020. Available from: http://www.rstudio.com/
13. Lamport L. LATEX: a document preparation system. Reading, (Mass): Addison-Wesley Pub. Co; 1986.
14. Xie Y. knitr. 2023 [cited 2023 Mar 15]. Available from: https://github.com/yihui/knitr

15. Xie Y. Dynamic documents with R and Knitr. 2nd ed. Boca Raton: CRC Press/Taylor & Francis; 2015.
16. Xie Y. knitr: A Comprehensive Tool for Reproducible Research in R. In: Implementing reproducible research Boca Raton: CRC Press, Taylor & Francis Group, 2014. (The R series).
17. knitr: A General-Purpose Package for Dynamic Report Generation in R version 1.42 from CRAN [cited 2023 Mar 15]. Available from: https://rdrr.io/cran/knitr/
18. Alston JM, Rick JA. A beginner's guide to conducting reproducible research. Bull Ecol Soc Am. 2021;102(2): e01801. https://doi.org/10.1002/bes2.1801
19. Barba LA. Terminologies for Reproducible Research. arXiv, 2018 [cited 2023 Mar 10]. Available from: http://arxiv.org/abs/1802.03311
20. Knuth DE. Literate Programming. Comput J 1984; 27(2): 97–111. doi: 10.1093/comjnl/27.2.97
21. Beal SL, Sheiner LB, Boeckmann AJ, et al. NONMEM 7.5 Users Guides. (1989–2020). Gaithersburg, MD: ICON plc, 2020 [cited 2023 Mar 28]. Available from: https://nonmem.iconplc.com/nonmem750
22. Zhou Y. Drug development and medical writing in the digital world. Med Writ 2019; 28(4):18–21.
23. Virdee N. How AI Tech Is Changing Regulatory Writing. Certara. 2017 [cited 2023 Mar 15]. Available from: https://www.certara.com/blog/how-ai-tech-is-changing-regulatory-writing/
24. Meloni M. Utilising innovative tools to accelerate regulatory document writing - Abstracts from the EMWA Spring Conference Poster Session. Med Writ. 2022;31:7. doi: 10.56012/uqis8887
25. Parisis N. Medical writing in the era of artificial intelligence. Med Writ. 2019;28(4):4–9.

Author information

Viviana Moroso, PhD, is an Associate Quality Control and Medical Writer at Pharmetheus (since 2020) actively working in client projects, with continuous improvement of our internal processes, and providing medical writing services. She is a member of EMWA since 2010 and has previously worked as a freelance medical writer/editor.
Google Scholar profile Viviana Moroso: https://scholar.google.com/citations?user=PF4f_7UAAAAJ&hl=en

Mats Magnusson, PhD, co-founded Pharmetheus in 2012, and is now a Principal MIDD Consultant, Team Leader & Deputy COO at Pharmetheus. He has held various research and teaching positions at Uppsala University and has worked as Pharmacometrician at Pfizer R&D and Senior Consultant at Exprimo NV.

Google Scholar profile Mats Magnusson: https://scholar.google.cz/citations?user=jD0nbpUAAAAJ&hl=en

Niclas Jonsson, PhD, co-founded Pharmetheus in 2012 and is now a Principal MIDD Consultant leading Pharmetheus' pharmacometrics research. He has held several research and teaching positions at Uppsala University and has worked as Global Head of Modelling and Simulation at Hoffmann-La Roche and as Senior Consultant at Exprimo NV.
Google Scholar profile Niclas Jonsson: https://scholar.google.se/citations?user=n7rLbqYAAAAJ&hl=en