ChatGPT in pharmacometrics? Potential opportunities and limitations

Michael Cloesmeijer[1], Alexander Janssen[1], Sjoerd Koopman[1], Marjon Cnossen[2], and Ron Mathot[3]

[1]Amsterdam UMC Locatie AMC
[2]Erasmus University Medical Center - Sophia Children's Hospital
[3]Amsterdam UMC - Locatie AMC

April 24, 2023

Abstract

The potential of using chatGPT in pharmacometrics was explored in this study, with a focus on developing a pharmacokinetic (PK) model for standard half-life factor VIII. Our results demonstrated that chatGPT can be utilized to accurately obtain typical PK parameters from literature, generate a population PK model in R, and develop an interactive Shiny application to visualize the results. ChatGPT's language generation capabilities enabled the development of R codes with minimal programming knowledge and helped identify and fix errors in the code. While chatGPT presents several advantages, such as its ability to streamline the development process, its use in pharmacometrics also has limitations and challenges, including the accuracy and reliability of AI-generated data, the lack of transparency and interpretability of AI. Overall, our study demonstrates the potential of using chatGPT in pharmacometrics, but researchers must carefully evaluate its use for their specific needs.

Introduction:

Over the years, the use of artificial intelligence (AI) in medical research has shown great promise in enhancing drug discovery, identifying new treatment targets, and predicting disease outcomes[1]. AI is an umbrella term encompassing several advanced technologies, such as machine learning, natural language processing, and deep learning. These methods facilitate the extraction of patterns and insights from vast amounts of data. A recent exciting development in AI research has been the public release of ChatGPT[2], developed by OpenAI. The model architecture behind ChatGPT (GPT; Generative Pre-trained Transformer[3] has shown to be very capable of achieving strong natural language understanding, while its accessible graphical user interface has resulted in widespread adoption.

Large Language Models (LLMs) such as ChatGPT are trained on an enormous corpus of text in order to generative responses to queries[4]. By devoting considerable human time labeling the quality of generated responses and re-training the model to produce the best responses, ChatGPT has suprised many to produce fluent and accurate responses to human inquiries. Aside from the public interest in the use of ChatGPT, there has also been suggestions of using the model to assist students and researchers by editing text, answering questions, writing code, and finding relevant literature given a query[5–8].

There already exist several publications discussing the potential impact of LLMs on a wide range of different research fields[9–11]. It however remains unknown if tools like ChatGPT can also support researchers from relatively small research fields, potentially resulting from a lower availability of training data. In this work, we investigate if ChatGPT can be used to assist during the development of population pharmacokinetic (PK) models. As an use-case, we use ChatGPT to generate R code for predicting in vivo drug concentrations of standard half-life factor VIII (FVIII) concentrates in patients with haemophilia A[12]. Next, we query

ChatGPT to generate an interactive R shiny application that can be used for the interpretation of the model and the selection of optimal doses. Based on this use-case, we aim to show that researchers unfamiliar with programming in R can nonetheless produce usable code for data analysis.

Methods

Data collection and model development

We used the official implementation of ChatGPT v3.5 (https://chat.openai.com; OpenAI) to send and receive answers to queries. The model was queried to provide typical estimates for the clearance (CL) and volume of distribution (V) for FVIII based on a patient body weight (BW) of 70kg. These estimates were used to define the population PK model and simulate FVIII levels over time.

Examples of queries send to ChatGPT for data collection are as follows:

What are typical CL and V of standard half-life FVIII for a patient of 70kg?

Can you give me references?

To verify the accuracy of the typical PK parameters ChatGPT provided, existing literature was consulted to compare ChatGPT response to reported values from previous studies. We additionally requested ChatGPT to produce the relevant literature along with the typical estimates of the PK parameters

Next, ChatGPT was queried to produce R code describing a population PK model using the previously requested typical PK parameters (figure 1). ChatGPT was iteratively queried to adjust the model to the users preferences and requirements for a FVIII PK model. At each step, produced code was transferred to R (v4.1.1. R Core Team 2023) and the predictions were visualized to inform the next changes to the model.

Examples of queries send to ChatGPT for model development are as follows:

1. in R, can you develop a 1 compartment population PK model of FVIII, using a typical CL of 2.5 dL/h and a V of 40 dL?
2. Can you use allometric body weight scaling on the PK parameters?
3. Can you add a baseline of FVIII to the simulated FVIII levels?

Allometric scaling is commonly used on PK parameters to account for differences in body size[13]. While patients with haemophilia A also produce endogenous FVIII, therefore a constant baseline value to the simulated FVIII levels is added[12].

Shiny applications development To visualize the simulated FVIII levels over time, ChatGPT was asked to create a Shiny application based on the PK model[14]. The application provided a tool for simulating FVIII levels over time. The application allowed the user to input the patient's body weight, baseline FVIII level, duration of the simulation and desired dosing. Additionally, it included a table output that displayed the simulated FVIII levels at several time points. The application also included a dropdown menu for selecting a therapeutic window, with dashed areas representing treatment windows for major surgery (30-50 IU/dL), intense physical activity (30-15 IU/dL), high risk activities (5-15 IU/dL), and mild activities (3-5 IU/dL). The therapeutic windows for FVIII levels were based on the recommendations by Berntorp et al[15].

The following queries were used in ChatGPT for shiny application development:

1. in R, can you visualize the simulations in a shiny application?
2. In the shiny application I want to adjust the body weight from 40 to 100 kg with steps of 1kg?
3. I also want to choose different doses between 250 and 5000 IU, change the duration of the simulation and choose a baseline FVIII between 0 and 40 IU/dL.
4. Can you add a therapeutic window with dashed lines? The following therapeutic windows should be added: major surgery 30-50 IU/dL, intense physical activity 30-15 IU/dL, higher risk activity 5-15 IU/dL, mild activity 3-5 IU/dL.

5. Can you add a table below the plot that shows the predicted FVIII levels at various time points during the simulation?

Results Developing a population PK model using ChatGPT

Based on the literature review provided by ChatGPT, typical clearance (CL) of FVIII in haemophilia A adults with a body weight of 70kg range between 2.5-3.5 dL/h depending on various factors such as age and health status. ChatGPT also suggested that the volume of distribution for a 70kg patient can range between 40-60 dL. We found that these PK parameters are consistent with previous studies. For example, one study by Björkman et al. found a CL of 2.3 dL/h and a V of 37.1 dL based on adults of 70kg[16]. Another study by McEneny-king and colleagues estimated a CL of 2.4-3.24 dL/h and a V of 35.4-46.2 dL , also based on adults of 70kg[17]. Based on the estimates provided by ChatGPT and the previous literature, we chose 2.5 dL/h and 40 dL as typical PK estimates for CL and V, respectively. Next, ChatGPT was queried to produce a population PK model in R using the previous typical PK estimates. The R code that ChatGPT generated for the base population PK model is displayed in figure 1. Not only does ChatGPT produce functional code, it also gives an explanation of how it works. Initially, the R code sets the initial FVIII level is zero, assuming no dose was given. If a dose was administered, the initial FVIII level (represented by C_0 in the code) as the dose divided by V. We then iteratively asked ChatGPT to add components to the model. ChatGPT understood how to normalize the PK parameters to body weight using allometric scaling and how to include a baseline FVIII level (F0) as a parameter, as shown in figure 2.

Developing a shiny application to visualize FVIII levels

Based on the PK model ChatGPT developed, a shiny application was developed by ChatGPT to simulate FVIII levels profiles based on patient characteristics. The Shiny application allows to adjust the body weight of a patient (from 40 to 100 kg with steps of 1 kg), the baseline FVIII level (from 0 to 40 IU/dL), the duration of the simulation, and the desired dosing (between 250 and 5000 IU). ChatGPT was able to successfully generate a Shiny application that can simulate FVIII levels over time, with realistic predictions[18] (figure 3 and supplement). All queries were successfully implemented in the Shiny application by ChatGPT. Moreover, ChatGPT was capable of implementing multiple requests from a single query, such as questions 3b and 4b.

What it cannot do

While ChatGPT was successful to generate R code for a single dose simulation, our experience showed that it struggled to provide an appropriate code for simulating multiple doses of FVIII and the results it generated were unrealistic. Therefore, caution must be exercised when using R codes generated by ChatGPT in pharmacometrics. We also asked ChatGPT to produce NONMEM code for the same model. Although the produced code did resemble a NONMEM control stream, the produced file contained multiple errors and redunancies, and failed to run.

Discussion

We show that ChatGPT is not only capable of accurately obtaining typical PK parameters from literature, but also has the ability to generate functional R code for predicting drug concentration using a population PK model as well to develop an interactive Shiny application to visualize model predictions.

ChatGPT generated an one-compartment population PK model in R and updated the code based on user specifications. By using ChatGPT to develop a Shiny application in R, users inexperienced with R shiny can easily produce web applications for interpreting their models. Both applications show how ChatGPT can be used without extensive coding or programming knowledge. This can significantly reduce development time and effort, while potentially improveing user experience of such applications. Another advantage of using ChatGPT for programming is its ability to assist developers in identifying and fixing errors in their code. ChatGPT can suggest possible solutions for errors and other coding mistakes, which helps inexperienced users to debug their code[19,20]. This feature can help streamline the development process and improve the overall quality of R code.

There are also some limitations and challenges to the use of ChatGPT for applications related to pharmacometrics. For example, the accuracy and reliability of AI-generated data may be affected by biases and knowledge gaps in the training data or the complexity of the query, for example when asking to produce code for more complex biological systems[21,22]. It can be especially difficult for inexperienced users to detect errors in the responses provided by ChatGPT, potentially with errors getting into proposed model code, affecting downstream results. This may lead to inaccurate or misleading results. Additionally, the lack of transparency and interpretability of AI algorithms may raise ethical concerns and limit their widespread adoption[23,24]. A more practical limitation of ChatGPT v3.5 that was used In this study is that response size is limited, potentially cutting off longer blocks of code. It can still present a challenge for users who require complete and accurate code snippets. Although ChatGPT might be especially interesting for inexperienced programmers, it might still be necessary to carefully review and edit code generated by ChatGPT to ensure its correctness and completeness. It is possible that future versions of ChatGPT may address some of the limitations observed in version 3.5, including the truncation of longer blocks of code. Another limitation to consider is that ChatGPT did not seem to be able to generate NONMEM control streams very well, which is unfortunate as NONMEM is the gold standard in pharmacometrics research, and the identification of errors in these streams can greatly aid students learning to use it[25]. This may be due to the limited availability of publicly available control streams, making it difficult for ChatGPT to learn from and generate accurate and reliable code for NONMEM models.

In conclusion, the integration of ChatGPT in pharmacometrics has the potential to streamline the development process and improve the user experience for pharmacometrics researchers. We deem it unlikely that ChatGPT will replace pharmacometricians in its current state. ChatGPT does have great value with respect to aiding researchers in finding and explaining information, generating and helping to debug code, and the education of new generations of pharmacometricians. As ChatGPT continues to evolve and improve, it has the potential to become an even more valuable tool in the field of pharmacometrics. As ChatGPT continues to evolve, it is likely that other pharmacometricians will find new and innovative ways to integrate it into their workflows and further enhance its capabilities in the field of pharmacometrics.

Acknowledgements

The SYMPHONY consortium aims to orchestrate personalized treatment in patients with bleeding disorders, and is a unique collaboration between patients, healthcare professionals and translational & fundamental researchers specialized in inherited bleeding disorders, as well as experts from multiple disciplines. It aims to identify best treatment choice for each individual based on bleeding phenotype. In order to achieve this goal, work packages have been organized according to three themes e.g. Diagnostics (workpackage 3&4); Treatment (workpackages 5–9) and Fundamental Research (workpackages 10–12). This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the NWA-ORC Call grant agreement NWA.1160.18.038. Principal investigator: Dr M.H. Cnossen; project coordinator: Dr S.H. Reitsma.

Beneficiaries of the SYMPHONY consortium: Erasmus University Medical Center-Sophia Children's Hospital, project leadership and coordination; Sanquin Diagnostics; Sanquin Research; Amsterdam University Medical Centers; University Medical Center Groningen; University Medical Center Utrecht; Leiden University Medical Center; Radboud University Medical Center; Netherlands Society of Hemophilia Patients (NVHP); Netherlands Society for Thrombosis and Hemostasis (NVTH); Bayer B.V., CSL Behring B.V., Swedish Orphan Biovitrum (Belgium) BVBA/SPRL.

Autorship contributions: M.E.C. performed the analysis. M.E.C., S.F.K. and A.J. wrote the manuscript. All authors contributed substantially to the critical revision of the manuscript and approved the final draft.

Conflict of interest statement

M.H.C.'s institution has received investigator-initiated research and travel grants as well as speaker fees over the years from the Netherlands Organisation for Scientific Research (NWO) and Netherlands National research Agenda (NWA), the Netherlands Organization for Health Research and Development (ZonMw), the

Dutch Innovatiefonds Zorgverzekeraars, Baxter/Baxalta/Shire/ Takeda, Pfizer, Bayer Schering Pharma, CSL Behring, Sobi Biogen, Novo Nordisk, Novartis and Nordic Pharma, and for serving as a steering board member for Roche, Bayer and Novartis for which fees go to the Erasmus MC as an institution. R.A.A.M. has received grants from governmental and societal research institutes such as NWO, ZonMW, Dutch Kidney Foundation and Innovation Fund and unrestricted investigator research grants from Baxter/ Baxalta/ Shire/Takeda, Bayer, CSL Behring, Sobi and CelltrionHC. He has served as advisor for Bayer, CSL Behring, Merck Sharp & Dohme, Baxter/ Baxalta/ Shire/Takeda. All grants and fees paid to the institution. Other authors have no conflict of interest to declare for this paper

Funding information

M.E.C., A.J. and S.F.K. are funded by the SYMPHONY consortium.

Data availability statement

Not applicable

References

1. Bohr A, Memarzadeh K. *The Rise of Artificial Intelligence in Healthcare Applications* .; 2020. doi:10.1016/B978-0-12-818438-7.00002-2

2. Gordijn B, Have H ten. ChatGPT: evolution or revolution? *Med Heal Care Philos* . 2023;26(1):1-2. doi:10.1007/s11019-023-10136-0

3. Radford A, Narasimhan K, Salimans T, Sutskever I, others. Improving language understanding by generative pre-training. Published online 2018.

4. Floridi L, Chiriatti M. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds Mach* . 2020;30(4):681-694. doi:10.1007/s11023-020-09548-1

5. Taecharungroj V. "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *Big Data Cogn Comput* . 2023;7(1):35. doi:10.3390/bdcc7010035

6. Liévin V, Hother CE, Winther O. Can large language models reason about medical questions? Published online 2022:1-33. doi:https://doi.org/10.48550/arXiv.2207.08143

7. Kashefi A, Mukerji T. ChatGPT for Programming Numerical Methods. Published online 2023:1-51. http://arxiv.org/abs/2303.12093

8. Temsah O, Khan SA, Chaiah Y, et al. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. *Cureus* . 2023;15(4). doi:10.7759/cureus.37281

9. Dahmen J, Kayaalp ME, Ollivier M, et al. Artificial intelligence bot ChatGPT in medical research: the potential game changer as a double-edged sword. *Knee Surgery, Sport Traumatol Arthrosc* . 2023;31(4):1187-1189. doi:10.1007/s00167-023-07355-6

10. Zhu JJ, Jiang J, Yang M, Ren ZJ. ChatGPT and Environmental Research. *Environ Sci Technol* . Published online 2023:1-4. doi:10.1021/acs.est.3c01818

11. Sallam M. The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations. *medRxiv* . Published online 2023:2023.02.19.23286155. http://medrxiv.org/content/early/2023/02/21/2023.02.19.23286155.abstract

12. Franchini M, Mannucci P. Past, present and future of hemophilia: a narrative review. *Orphanet J Rare Dis* . 2012;7(1):24. doi:10.1186/1750-1172-7-24

13. Knibbe CAJ, Zuideveld KP, Aarts LPHJ, Kuks PFM, Danhof M. Allometric relationships between the pharmacokinetics of propofol in rats, children and adults. *Br J Clin Pharmacol* . 2005;59(6):705-711. doi:10.1111/j.1365-2125.2005.02239.x

14. Wojciechowski J, Hopkins AM, Upton RN. Interactive pharmacometric applications using R and the Shiny package. *CPT Pharmacometrics Syst Pharmacol* . 2015;4(3):146-159. doi:10.1002/psp4.21

15. Berntorp E, Hermans C, Solms A, Poulsen L, Mancuso ME. Optimising prophylaxis in haemophilia A: The ups and downs of treatment.*Blood Rev* . 2021;50. doi:10.1016/j.blre.2021.100852

16. Björkman S, Oh M, Spotts G, et al. Population pharmacokinetics of recombinant factor VIII: the relationships of pharmacokinetics to age and body weight. *Blood* . 2012;119(2):612-618. doi:10.1182/blood-2011-07-360594

17. McEneny-King A, Chelle P, Foster G, Keepanasseril A, Iorio A, Edginton AN. Development and evaluation of a generic population pharmacokinetic model for standard half-life factor VIII for use in dose individualization. *J Pharmacokinet Pharmacodyn* . 2019;46(5):411-426. doi:10.1007/s10928-019-09634-7

18. Bukkems LH, Preijers T, Van Spengler MWF, Leebeek FWG, Cnossen MH, Mathôt RAA. Comparison of the Pharmacokinetic Properties of Extended Half-Life and Recombinant Factor VIII Concentrates by in Silico Simulations. *Thromb Haemost* . 2021;121(6):731-740. doi:10.1055/s-0040-1721484

19. Ahmad A, Waseem M, Liang P, Fehmideh M, Aktar MS, Mikkonen T. Towards Human-Bot Collaborative Software Architecting with ChatGPT. Published online 2023:1-7. http://arxiv.org/abs/2302.14600

20. Biswas S. Role of ChatGPT in Computer Programming.*Mesopotamian J Comput Sci* . Published online 2023:8-16. doi:10.58496/mjcsc/2023/002

21. Borji A. A Categorical Archive of ChatGPT Failures. Published online 2023:1-41. http://arxiv.org/abs/2302.03494

22. Mann DL. Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine: A JACC: Basic to Translational Science Interview With ChatGPT. *JACC Basic to Transl Sci* . 2023;8(2):221-223. doi:10.1016/j.jacbts.2023.01.001

23. Editorials N. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* . 2023;613:612.

24. Biswas S. ChatGPT and the Future of Medical Writing.*Radiology* . 2023;307(2):3-5. doi:10.1148/radiol.223312

25. Keizer RJ, Karlsson MO, Hooker A. Modeling and simulation workbench for NONMEM: Tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst Pharmacol* . 2013;2(6):1-9. doi:10.1038/psp.2013.24

Tables

Not applicable

Figure legends

Figure 1. R code for simulated Factor VIII levels over time in a one-compartment pharmacokinetic model . The model and R code for the one-compartment population pharmacokinetic model were generated by ChatGPT.

Figure 2. Prompts to adds components to the population pharmacokinetics model generated by ChatGPT.

Figure 3. Generated Shiny application for simulating FVIII level profiles. The figure in the top right shows the simulated FVIII level profile over 72 hours for a patient with a body weight of 70kg, receiving a dose of 3000 IU and with a baseline FVIII level of 0 IU/dL. The red dotted lines represent the currently

selected therapeutic window. The table below the plot shows the predicted FVIII level at various time points during the simulation, including the peak level and the end of the simulation period.

Appendences

Final R code of the population PK model for factor VIII and Shiny application

M

in R, can you develop a 1 compartment population PK model of FVIII, using a typical CL of 2.5 dL/h and a V of 40 dL?

Certainly! Here's an example of how to create a one-compartment population PK model for FVIII using R and the deSolve package:

R Copy code

```r
library(deSolve)

# Define PK parameters
CL <- 2.5  # Clearance (dL/h)
V <- 40    # Volume of distribution (dL)

# Define the one-compartment PK model
onecomp.model <- function(t, state, parameters) {
  # Define PK parameters
  CL <- parameters[1]
  V <- parameters[2]

  # Define differential equations
  dCdt <- (-CL/V * state[1])

  # Return list of differential equations
  list(dCdt)
}

# Define initial state
C_0 <- 0  # Factor VIII concentration at t=0 (IU/dL)
state_0 <- c(C = C_0)

# Define time points for simulation
times <- seq(from = 0, to = 24, by = 0.1)

# Simulate the PK model
out <- ode(y = state_0, times = times, func = onecomp.model, parms = c(CL, V

# Plot the results
plot(out[, "time"], out[, "C"], type = "l", xlab = "Time (h)", ylab = "Facto
```

In this example, we define the typical PK parameters of FVIII, namely a clearance (CL) of 2.5 dL/h and a volume of distribution (V) of 40 dL. We then define a one-compartment PK model function that takes in the time (t), state (C, the factor VIII concentration), and parameters (CL and V) as inputs, and returns the differential equation for the rate of change of factor VIII concentration over time. We set the initial factor VIII concentration (C_0) to 0, define a range of time points for simulation (0 to 24 hours in this case), and simulate the PK model using the `ode` function from the deSolve package. Finally, we plot the simulated factor VIII concentration over time.

M

Can you add a baseline of FVIII to the simulated FVIII levels?

Yes, we can add a baseline value of FVIII to the simulated concentrations. Here is the updated code:

R Copy code

```
library(deSolve)

# Define baseline FVIII value
F0 <- 1  # Baseline FVIII concentration (IU/dL)

# Simulate the PK model
out <- ode(y = state_0, times = times, func = onecomp.model,
           parms = c(CL, V, F0), rtol = 1e-10, atol = 1e-10)

# Add the baseline value to the simulated concentrations
out[, "C"] <- out[, "C"] + F0
```

M

Can you use allometric body weight scaling on the PK parameters?

Sure, here's an example of how to use allometric body weight scaling on the PK parameters:

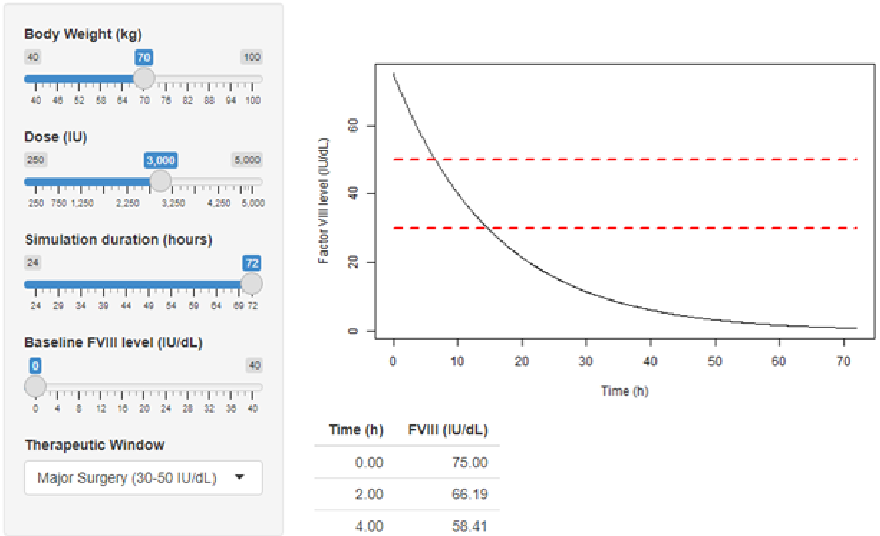
R Copy code

```
library(deSolve)

# Define body weight in kg
BW <- 70

# Define PK parameters using allometric scaling
CL <- 2.5 * (BW/70)^0.75  # Clearance (dL/h)
V <- 40 * (BW/70)         # Volume of distribution (dL)
```

Factor VIII levels over time



Time (h)	FVIII (IU/dL)
0.00	75.00
2.00	66.19
4.00	58.41
12.00	35.43
24.00	16.73
48.00	3.73
72.00	0.83