



# Janick Weberpals

## RWE Data Scientist

- 30 April 1989, Bruchsal, Germany
- German citizenship
- Boston, MA, United States
- janick.weberpals@gmail.com
- +1 857 381-7865
- linkedin.com/in/weberpals
- janickweberpals.github.io
- scholar.google.com/citations

## About Me

I'm a pharmacist and real-world data (RWD) scientist with 8+ years of experience in the design and analysis of large real-world evidence studies in both industry and academia. I'm passionate about integrating various data modalities (EHR, claims, genomics, pathology, imaging) using machine learning and advanced analytics to solve problems in healthcare and causal inference. I have gained deep clinical expertise in the field of oncology and other disease areas, (co-) authored 25+ peer-reviewed methodological and clinical publications and was recognized with several awards resulting from this work.

## Hard Skills

- R (advanced) tidyverse, devtools
- Python (basics), Tensorflow/Keras
- SAS (basics)
- git, GitHub, GitLab, CI/CD
- dbplyr, Apache Arrow, SQL (basics)
- HPC, SLURM, Unix, AWS
- Quarto, Markdown,  $\LaTeX$
- Medical coding standards & common data models (ICD, NDC, ATC, LOINC, CPT-4, HCPCS, OMOP, ...)

## Education

- 2018 – 2020 **Postdoctoral Fellowship** Data Science, Roche Innovation Center, Germany  
*Deep learning on electronic health record (EHR) data*
- 2015 – 2018 **Ph.D. Epidemiology** Medical Faculty, Heidelberg University, Germany  
*Graduated with Summa cum laude honors*
- 2015 – 2018 **Board certification** Bavarian Chamber of Pharmacists, Munich, Germany  
*Specialized Pharmacist in Drug Information*
- 2010 – 2015 **Pharmacy** College of Pharmacy, Marburg University, Germany  
*Registered Pharmacist (RPh)*

## Professional Experience

- 2022– **Instructor in Medicine** Harvard Medical School, Boston, MA, USA  
*Division of Pharmacoepidemiology & Pharmacoeconomics*
  - Leading an oncology-focused research group leveraging routinely collected healthcare data (EHR, claims, imaging) to deliver high-quality real-world evidence (RWE) studies.
  - Attracted > \$1 Mio. in research funding as (co-) principal investigator from the US FDA to develop a systematic approach to missing data in causal inference and RWE studies. Project resulted in implementable *smdi* R package as deliverable.
  - Established transparent and reproducible programming workflows using git and automated CI/CD pipelines via GitLab repositories and taught R programming courses.
- 2020–2022 **Data Scientist** Hoffmann-La Roche, Basel, Switzerland  
*PHC Data Analytics & Imaging (2021-22), RWD Collaborations (2020-21)*
  - Scientific lead and contributor to several work streams aiming to link, integrate and evaluate multimodal data (EHR, claims, imaging, NLP) in collaboration with Flatiron Health and other data partners to expedite and support late-stage oncology drug development using RWD. Parts of this work were recognized by the Roche/Genentech senior management with the product development innovation breakthrough award.
  - Was awarded "Exceptional Performance" in 2021.
  - Planned, executed and published RWE studies in collaboration with Foundation Medicine and academic key opinion leaders to support the personalized healthcare integrated evidence generation program with focus on genetic testing in cancer populations.
  - Led a strategic cross-functional data partnership due diligence effort which saved the organization ~ \$300k in avoidable spendings.
  - Scientific lead of collaboration with second largest HMO in Israel which contextualized COVID-19 infection and vaccination patterns among multiple sclerosis patients with ocrelizumab exposure.
  - Served as RWD expert in a multidisciplinary study management team for *Wayfind-R*, an international cancer registry enrolling patients who underwent next generation sequencing (NGS) testing (NCT04529122) enabling large-scale tumor agnostic studies.
- 2018-2020 **Postdoctoral Fellow Data Science** Roche Diagnostics, Munich, Germany  
*Pharma Research and Early Development (pRED) Data Science*
  - Developed a deep learning-based propensity score algorithm in Flatiron Health EHR data to improve validity of RWE studies.
  - Contributed to the development of a pan-tumor prognostic score for overall survival as a decision-support tool for trial eligibility and patient enrichment in early-phase oncology trials.
  - Contributed to RWD efforts (e.g., development of external control arms) to contextualize and support business-critical go/no-go decision-making for clinical phase I molecule development.
  - Attracted ~ 135k EUR of internal research funding to advance AI in RWD & drug development and grow the team with M.Sc. & Ph.D. students in collaboration with national universities.

## Soft Skills

- Agile & hands-on
- Creative & innovative
- Organized
- Collaborative
- Versatile
- Exchange & sharing
- Pragmatic

## Software

**smdi** - An R package to perform routine structural missing data investigations (*deployed on CRAN* [↗](#), *main author*)

**autoencoderPS** - An autoencoder-based propensity score in Python/Tensorflow (*deployed on GitHub* [↗](#), *main author*)

**FlatironKitchen** [↗](#) - An R package for end-to-end analyses in the Flatiron Health database (*proprietary Roche internal package*, contributor)

## Patents

PROPENSITY SCORE BASED ASSESSMENT OF PATIENT DATA. International Patent Application Number PCT/EP2020/ 064 134, filed 20 May 2020, Patent Pending. Publication WO2020234388A1 [↗](#)

## Languages

- English ● ● ● ● ●
- German ● ● ● ● ●

## Memberships



International Society for Pharmacoepidemiology (ISPE)



German Pharmaceutical Society (DPhG)



German Society for Epidemiology (DGEPi)

## Professional Experience *(continued)*

- 2015-2018 **Doctoral Researcher** German Cancer Research Center, Heidelberg, Germany  
*Division of Clinical Epidemiology and Aging Research*  
◦ Managed, QC'ed and analyzed large oncological database linkages with a focus on drug repurposing [↗](#).  
◦ Partnered with (inter)national cancer registries which resulted in 14 peer-reviewed publications impacting *public cancer survival surveillance* [↗](#).
- 2014-2015 **Research Scholar** University of Florida, Gainesville, FL, USA  
*Department of Pharmaceutical Outcomes and Policy*  
Contributed to a multidisciplinary project to develop an EHR-based *predictive risk model* [↗](#) to prevent adverse events among hospitalized patients. The model was implemented in select US hospitals.

## Talks & Outreach *(selected, full list: [janickweberpals.github.io/talks](https://janickweberpals.github.io/talks) [↗](#))*

*Introducing smdi: An R package to perform structural missing data investigations for real-world evidence studies.* **R/Pharma Conference** [↗](#). Virtual presentation, 2023.

*Issues and Solutions When Estimating Treatment Effects Using US Electronic Health Record Data.* (invited panelist [↗](#)). **International Society for Health Economics and Outcomes Research (ISPOR) Annual Meeting**, Boston, MA, 2023.

*Deep Learning on Electronic Health Records for Research in Pharmacoepidemiology: Examples from The Field of Oncology.* Invited speaker at the **FDA Sentinel Innovation And Methods Seminar Series** [↗](#) Virtual presentation, 2022.

## Publications *(selected, full list: [janickweberpals.github.io/publications](https://janickweberpals.github.io/publications) [↗](#))*

Weberpals J, Wang SV. The FAIRification of research in real-world evidence: A practical introduction to reproducible analytic workflows using Git and R. **Pre-print on Authorea** [↗](#) (2023).

Sondhi A<sup>1</sup>, Weberpals J<sup>1</sup>, Yerram P, Jiang C, Taylor MD, Samant M, Cherng S. A systematic approach towards missing lab data in electronic health records: A case study in non-small cell lung cancer and multiple myeloma. **CPT Pharmacometrics Syst Pharmacol** [↗](#) (2023). <sup>1</sup>*contributed equally*

Weberpals J, Becker T, Schmich F, Ruettinger D, Theis FJ, Bauer-Mehren A. Deep learning-based propensity scores for confounding control in comparative effectiveness research: a large-scale real-world data study. **Epidemiology** [↗](#) (2021).

Loureiro H, Becker T, Bauer-Mehren A, Ahmidi N, Weberpals J\*. Artificial Intelligence for Prognostic Scores in Oncology: a benchmarking study. **Frontiers in Artificial Intelligence** [↗](#) (2021). \**supervised project as senior author*

## Awards *(selected, full list: [janickweberpals.github.io/awards](https://janickweberpals.github.io/awards) [↗](#))*

- 2018 Stephan-Weiland Award 1st prize German Society for Epidemiology  
2018 Advancement Award for best Ph.D. thesis in Epidemiology GMDs  
2017 Poster Award Helmholtz International Graduate School  
2016 & 2017 Travel Scholarships International Society for Pharmacoepidemiology

## Other Activities

- 2019-2021 Adjunct lecturer (Medical Faculty, University of Heidelberg, Germany), Conceptualization and teaching of an introductory course on design and biases of pharmacoepidemiological studies for Biostatistics/Medical Biometry M.Sc. students
- 2020 - Core member, Real World Evidence Task Force - Statistical Methods, International Society for Pharmacoepidemiology