Janick Weberpals

RWD Scientist

Bruchsal, Germany

German citizenship

[redacted]

linkedin.com/in/weberpals

janickweberpals.github.io

scholar.google.com/citations

About Me

I'm a pharmacist and real-world data (RWD) scientist with 10+ 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 30+ peer-reviewed methodological and clinical publications and was recognized with several awards resulting from this work.

Hard Skills

- R (advanced)
- Python, Tensorflow/Keras
- SAS (basics)
- git, GitHub, GitLab, CI/CD
- Apache Arrow, dbplyr, 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

2024– **Associate Director** AstraZeneca, Munich, Germany
R&D Oncology Data Science

- Hands-on development and application of innovative statistical and machine learning methodology and software to support real-world evidence studies in oncology.

2022–2024 **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.
- Lead of a large-scale FDA-funded project aiming to emulate various oncology trials using multiple different US EHR data.
- 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-phase 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.
- 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.

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)*

ci5 - An R package to fit undersmoothed LASSO propensity score models for causal inference (*deployed on GitLab , contributor)*

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*)

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*)

Weberpals J, Wang SV. External Controls to Study Treatment Effects in Rare Diseases: Challenges and Future Directions. **Clin Pharmacol Ther.** (2024).

Weberpals J, [...], Schneeweiss S, Desai, RJ. High-dimensional multiple imputation (HDMI) for partially observed confounders including natural language processing-derived auxiliary covariates. **arXiv** (2024).

Weberpals J, Becker T, Davies J, 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*)

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