Rodrigo Dal Ben, Ph.D.

+1 587-226-0434 | Calgary, AB, Canada | dalbenwork@gmail.com | linkedin.com/in/rodrigo-dal-ben-phd | github.com/RodDalBen

Authorized to work in Canada, EU, and Brazil - no sponsorship required

SUMMARY

Applied Data Scientist with a Ph.D. and 6+ years of experience in real-world data (RWD) analytics, predictive modeling, and causal inference in healthcare and behavioral science. Proven expertise designing reproducible, FAIR-compliant data pipelines for large-scale longitudinal and observational studies, generating real-world evidence (RWE) for clinical and operational decision-making. Skilled in Bayesian modeling, mixed-effects models, and machine learning applied to patient-reported outcome measures (PROMs), risk stratification, and outcomes research. Experienced with Python, R, and SQL in CI/CD-driven production workflows. Strong record of cross-functional collaboration, peer-reviewed publication, and translating complex data into actionable healthcare insights.

SKILLS

Methods: AI agentic engineering**,** Machine learning, Bayesian modeling, mixed-effects models, causal inference, survival analysis, experimental design, A/B testing, NLP, feature importance analysis, risk stratification, predictive screening

Healthcare & Data Governance: RWD/RWE pipelines, PROMs analysis, FAIR compliance, data harmonization, provenance tracking, IRB-compliant workflows, audit-ready reproducibility, longitudinal cohort studies

Frameworks & Libraries: Python (scikit-learn, pandas, numpy, PyTorch), R (brms, lme4, tidyverse, Shiny), SQL

Tools & Workflow: Git, GitHub/GitLab, CI/CD, Docker, Azure, DataLad, REDCap, LLMs, agentic AI (Edison-FutureHouse, SpecKit, Claude Code)

WORK EXPERIENCE

Senior Research Analyst | Cumming School of Medicine, University of Calgary *2024 to Present*

Calgary, AB | Healthcare RWD/RWE, Longitudinal Pediatric Outcomes Research

- Designed automated RWD harmonization, validation, and provenance pipelines integrating 23 data sources from an 800-family multinational longitudinal pediatric consortium, reducing processing time 280-fold (140 hours to 35 minutes) while maintaining full audit-ready reproducibility via CI/CD workflows.
- Built analysis-ready silver-layer datasets with consistent tidy schema, enabling seamless integration with mixed-effects models, Bayesian models, and ML pipelines for pediatric health outcomes research.
- Led analysis and authorship of 8 RWE manuscripts on pediatric patient-reported outcome measures (PROMs) supporting clinical and policy decision-making: 1 published, 3 under review, 4 in preparation.
- Implemented versioned, FAIR-compliant open science workflows across IRB-approved, EHR-linked longitudinal data sources under GDPR frameworks, replacing weeks of manual non-reproducible work with scalable, audit-ready reproducible processes.
- Leveraged LLM and agentic AI tools to accelerate RWD pipeline development and maintenance, extract structured outputs from unstructured clinical text, and support systematic evidence synthesis, reducing manual overhead across the full data-to-insight workflow and enabling faster iteration across data collection cycles.
- Mentored 3 graduate students and 2 postdoctoral researchers on advanced statistical modeling, reproducibility, versioning, and open science best practices.

Data Scientist | Bee Touch Workforce Mental Health *2023 to 2025*

Remote | Predictive Risk Modeling, Clinical Decision Support, Population Health

- Built predictive, versioned, and reproducible ML pipelines for clinical risk stratification (burnout, absenteeism, presenteeism, acute suicide risk) across 10,000+ individuals in 14 major organizations using RWD mixed-method approaches.

- Identified 300+ cases of acute suicide risk via predictive screening models, triggering immediate personalized clinical intervention - directly informing clinical decision support workflows.
- Identified key drivers of mental health risk (workload, autonomy, management style) via feature importance analysis; communicated findings to senior stakeholders through reproducible data-driven reports, supporting structural interventions with an estimated ROI of $4 per $1 invested.
- Applied experimental design and A/B testing principles to evaluate intervention effectiveness and model performance across cohorts.

Assistant Professor | Ambrose University *2022 to 2025*

Calgary, AB | Research Infrastructure, Open-Source Tools, Applied Statistics

- Directed a cross-functional research program in cognitive development and open science, securing a portion of over $430,000 in research funding.
- Developed and maintained open-source software tools to support reproducible research pipelines aligned with evolving industry and regulatory standards.
- Published 6 peer-reviewed articles in journals including MethodsX and Journal of Cognition and Development; served as peer reviewer for 9 journals including Scientific Reports and MIT Press.

Horizon Postdoctoral Researcher | Concordia University *2020 to 2021*

Montreal, QC | Big-Team Science Infrastructure, Reproducible Workflows

- Built reproducible research infrastructure supporting 800+ researchers across 50+ countries within the ManyBabies consortium; applied FAIR principles, code versioning, and open data standards at scale.
- Engineered reproducible data analysis workflows for complex longitudinal eye-tracking data using mixed-effect logistic regression, generating fine-grained insights into cognitive decision-making.
- Published 4 peer-reviewed articles on open science, cognitive development, and bilingualism (e.g., Open Mind, Behavioral and Brain Sciences).

EDUCATION

Ph.D., Psychology (Cognitive Science & Applied Statistics) *2020*

Federal University of Sao Carlos & University of Tennessee, Knoxville

Professional Certificate: Data Science *2022*

Harvard (edX)

PUBLICATIONS

Full list: scholar.google.com/citations?user=7DufssUAAAAJ

- Marshall, D. A., **Dal Ben, R.**, Currie, G. R., Yeung, R. S. M., Vastert, S. J., Wulffraat, N., Swart, J. F., Benseler, S. (2026). Work productivity and activity impairment in patients under 18 with juvenile idiopathic arthritis (JIA): The international UCAN CAN-DU prospective study. *Arthritis Research & Therapy, 28*(1), 106. https://doi.org/10.1186/s13075-026-03785-y
- **Dal Ben, R.** (2023). SHINE_color: controlling low-level properties of colorful images. *MethodsX, 11*, 102377.
- ManyBabies Consortium, with **Dal Ben, R.** et al. (2023). Validation of an open source, remote web-based, eye-tracking method (WebGazer) for research in early childhood. *Infancy, 29*(3), 31to55