Clinical Data Science in R

Methods in Clinical AI and Translational Trauma Research

AUTHOR
R. Heller

Preface

R. Heller · Date: 2026-11-08 · Last changed: 2026-11-09

A clinical team receives a fracture alert. A research group measures hundreds of spectra from a handful of animals. An emergency department plans its next shift. This book follows the decision that each of these situations is meant to support. Readers should find three things in every chapter: a calculation they can retrace, a way to try to refute their own claim, and a reasoned answer to the question of who may do what with the result.

Who this book is for

The book is written for medical students, doctoral researchers, clinical research teams and people who review models. Newcomers start from the clinical question and learn the R they need along the way (4 Reconstructing a research project in R). Part I assumes basic medical knowledge but no programming experience. Readers who want to check a specific result can enter by data type, method or failure pattern; the navigator in 28 Case atlas, glossary and transfer leads from the starting question to the right chapter.

How the book is organised

The 28 chapters are grouped into seven parts. The chapter number doubles as a fixed identifier used by the course and by teaching teams: chapter 9 is B09. Every chapter follows the same pattern:

1. **Case and decision.** One person or role, one question and an action with consequences.
2. **Core reading.** A labelled figure or table, the key assumption and a question for discussion. The core reading is clearly marked, runs to about 800–1,200 words (8–12 minutes) and is the required reading in the course.
3. **Lab or decision exercise.** In computational chapters, the complete route in R from inputs to result; in legal and communication chapters, reading of primary sources, a role assignment and a reasoned decision.
4. **Counter-check.** A parameter, an assumption or the target population is changed. Predict the direction first, then compare with the result.
5. **Consequences for patients and operations.** Ethics, regulation, data protection and security, ordered by their relevance to the case.

6. **Look up and transfer.** A reviewer question, a self-test with explanation, annotated primary sources, the link to the course and further reading in this book and its companion volumes.



Data status

All computational examples use synthetic data generated in the chapter with a fixed seed, or explicitly identified published figures. Every number in the text belongs to one of four categories:

Label	Meaning
synthetic	generated in the book; describes no real population
published	taken from the cited primary source
assumption	value set for teaching purposes that would need local verification in practice
open	a question the chapter raises but does not answer

No number in this book describes a commercial product, a real hospital or unpublished project results.

Reading paths in the course

The course Clinical Data Science in R uses selected chapters of this book as required reading. Readers working through the book on their own do not need the course.

Course stage	Required reading per workshop	Supplementary
I · Entry	W1: 1 Clinical decisions and the translational chain, 3 How data arise: missingness and denominators · W2: 7 Reading an AI study critically, 8 Language models in clinic and science	23 Data protection along the entire data flow, 25 Security risks of learning and generative systems
II · Application	W3: 6 Accuracy, calibration and clinical utility, 9 A fracture alert in the emergency department · W4: 10 Hyperspectral imaging, 19 From one centre or species to the next	11 Measuring gait from video, 12 Spatial biology and image analysis
III · Advanced	W5: 2 Describe, predict, explain, 13 Longitudinal biomarkers · W6: 17 Forecasting patient flow and 15 Mechanistic models and pharmacometrics or 18 Mass casualty incidents, operating capacity and simulation	14 Cluster, gradient or artefact, 16 Bayes, small samples and the 3Rs, 27 Model passport, study report and decision
IV · Translation	W7: 22 Regulating research software and clinical AI, 23 Data protection along the entire data flow · W8: 27 Model passport, study report and decision, 28 Case atlas, glossary and transfer	20 Pilot operation, impact and monitoring, 21 Ethics at the bedside and in study design, 24 Information security and recovery, 25 Security risks of learning and generative systems

For a quick look-up, the first question decides the route. “How reliable is the alert for this patient?” leads from 9 A fracture alert in the emergency department via 6 Accuracy, calibration and clinical utility to 21 Ethics at the bedside and in study design. “May our test set contain images from the same animal?” leads from 10 Hyperspectral imaging via 19 From one centre or species to the next to 27 Model passport, study report and decision. “The support log contains case text” leads through 23 Data protection along the entire data flow to 25 Security risks of learning and generative systems and on to the restart decision in 20 Pilot operation, impact and monitoring.

Companion books and courses

General statistics is repeated here only as far as a case requires. For the methods themselves, the chapters link to the relevant sections of:

- Methods in R – a method reference for researchers and reviewers, with diagnostics, red flags and reporting templates (including TRIPOD-AI and ARRIVE);
- Scientometrics in R – case studies, labs and reference material on research literature, reproducible pipelines and responsible metrics;
- Strategy in R – simulation with explicit assumptions and decisions under uncertainty;
- CTTIR Courses – a four-part biostatistics curriculum in R whose progression the course for this book follows;
- CTTIR Tutorials – stand-alone tutorials on foundations, regression, time series, Bayesian statistics, machine learning and clinical biostatistics.

Each chapter ends with a “Further reading” section that points to the matching sections.

Programming

The complete teaching path is written in R (R Core Team 2026). Base R is sufficient for the elementary examples; mixed models, differential equations and process simulation use well-established packages such as **lme4**, **deSolve** and **simmer**. Python or Julia appear only where a data pipeline, research weights, solvers or interfaces offer a concrete advantage; a box then describes the original implementation and what must be identical for a fair comparison. Software and legal status of this edition are recorded in Appendix A — Edition notes, software and imprint.

History of this book

Work on the cases, worked examples and chapter structure began in February 2025 and was extended and revised regularly until July 2026. This edition reflects the legal, literature and software status of November 2026.

Citation and licence

> Heller, R. (2026). *Clinical Data Science in R: Methods in Clinical AI and Translational Trauma Research.* https://cttir.github.io/clinical-data-science-in-r/

Text and figures are licensed under CC BY 4.0, code under the MIT licence. Please report corrections through the repository issues.

This book does not replace individual medical, legal or statistical advice.