References

AUTHOR
R. Heller

R. Heller · Date: 2026-05-17 · Last changed: 2026-05-17

Altman, Douglas G., and J. Martin Bland. 1994. “Statistics Notes: Diagnostic Tests 2: Predictive Values.” *BMJ* 309 (6947): 102. https://doi.org/10.1136/bmj.309.6947.102.

Ancker, Jessica S., Alison Edwards, Sarah Nosal, Diane Hauser, Elizabeth Mauer, and Rainu Kaushal. 2017. “Effects of Workload, Work Complexity, and Repeated Alerts on Alert Fatigue in a Clinical Decision Support System.” *BMC Medical Informatics and Decision Making* 17 (1): 36. https://doi.org/10.1186/s12911-017-0430-8.

Baddeley, Adrian, and Rolf Turner. 2005. “Spatstat: An R Package for Analyzing Spatial Point Patterns.” *Journal of Statistical Software* 12 (6): 1–42. https://doi.org/10.18637/jss.v012.i06.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bellman, R., and K. J. Åström. 1970. “On Structural Identifiability.” *Mathematical Biosciences* 7 (3–4): 329–39. https://doi.org/10.1016/0025-5564(70)90132-X.

Benneyan, J. C., R. C. Lloyd, and P. E. Plsek. 2003. “Statistical Process Control as a Tool for Research and Healthcare Improvement.” *Quality and Safety in Health Care* 12 (6): 458–64. https://doi.org/10.1136/qhc.12.6.458.

Bland, J. Martin, and Douglas G. Altman. 1986. “Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement.” *The Lancet* 327 (8476): 307–10. https://doi.org/10.1016/S0140-6736(86)90837-8.

Bland, J. Martin, and Douglas G. Altman. 2007. “Agreement Between Methods of Measurement with Multiple Observations Per Individual.” *Journal of Biopharmaceutical Statistics* 17 (4): 571–82. https://doi.org/10.1080/10543400701329422.

Broman, Karl W., and Kara H. Woo. 2018. “Data Organization in Spreadsheets.” *The American Statistician* 72 (1): 2–10. https://doi.org/10.1080/00031305.2017.1375989.

Brown, Lawrence D., T. Tony Cai, and Anirban DasGupta. 2001. “Interval Estimation for a Binomial Proportion.” *Statistical Science* 16 (2): 101–33. https://doi.org/10.1214/ss/1009213286.

Bryan, Jennifer. 2018. “Excuse Me, Do You Have a Moment to Talk about Version Control?” *The American Statistician* 72 (1): 20–27. https://doi.org/10.1080/00031305.2017.1399928.

Calfee, Carolyn S., Kevin Delucchi, Polly E. Parsons, B. Taylor Thompson, Lorraine B. Ware, and Michael A. Matthay. 2014. “Subphenotypes in Acute Respiratory Distress Syndrome: Latent Class Analysis of Data from Two Randomised Controlled Trials.” *The Lancet Respiratory Medicine* 2 (8): 611–20. https://doi.org/10.1016/S2213-2600(14)70097-9.

Chari, Tara, and Lior Pachter. 2023. “The Specious Art of Single-Cell Genomics.” *PLOS Computational Biology* 19 (8): e1011288. https://doi.org/10.1371/journal.pcbi.1011288.

Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A Scalable Tree Boosting System.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining, 785–94. https://doi.org/10.1145/2939672.2939785.

Christodoulou, Evangelia, Jie Ma, Gary S. Collins, et al. 2019. “A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models.” *Journal of Clinical Epidemiology* 110: 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004.

Clopper, C. J., and E. S. Pearson. 1934. “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial.” *Biometrika* 26 (4): 404–13. https://doi.org/10.1093/biomet/26.4.404.

Collins, Gary S., Karel G. M. Moons, Paula Dhiman, et al. 2024. “TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods.” *BMJ* 385: e078378. https://doi.org/10.1136/bmj-2023-078378.

Court of Justice of the European Union. 2025a. *Judgment of 4 September 2025, European Data Protection Supervisor v Single Resolution Board, Case C-413/23 P*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62023CJ0413.

Court of Justice of the European Union. 2025b. *Judgment of 4 September 2025, European Data Protection Supervisor v Single Resolution Board, Case C-413/23 P, ECLI:EU:C:2025:645*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62023CJ0413.

Davis, Sharon E., Thomas A. Lasko, Guanhua Chen, et al. 2017. “Calibration Drift in Regression and Machine Learning Models for Acute Kidney Injury.” *Journal of the American Medical Informatics Association* 24 (6): 1052–61. https://doi.org/10.1093/jamia/ocx030.

de Jonghe, Annemarieke, Barbara C. van Munster, J. Carel Goslings, et al. 2014. “Effect of Melatonin on Incidence of Delirium Among Patients with Hip Fracture: A Multicentre, Double-Blind Randomized Controlled Trial.” *Canadian Medical Association Journal* 186 (14): E547–56. https://doi.org/10.1503/cmaj.140495.

Debray, Thomas P. A., Yvonne Vergouwe, Hendrik Koffijberg, et al. 2015. “A New Framework to Enhance the Interpretation of External Validation Studies of Clinical Prediction Models.” *Journal of Clinical Epidemiology* 68 (3): 279–89. https://doi.org/10.1016/j.jclinepi.2014.06.018.

DeLong, Elizabeth R., David M. DeLong, and Daniel L. Clarke-Pearson. 1988. “Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.” *Biometrics* 44 (3): 837–45. https://doi.org/10.2307/2531595.

Diggle, Peter, and Michael G. Kenward. 1994. “Informative Drop-Out in Longitudinal Data Analysis.” *Applied Statistics* 43 (1): 49–93. https://doi.org/10.2307/2986113.

Elwyn, Glyn, Dominick Frosch, Richard Thomson, et al. 2012. “Shared Decision Making: A Model for Clinical Practice.” *Journal of General Internal Medicine* 27 (10): 1361–67. https://doi.org/10.1007/s11606-012-2077-6.

European Commission. 2025. *Proposal for a Regulation Amending Regulations (EU) 2017/745 and (EU) 2017/746 as Regards Simplifying and Reducing the Burden of the Rules on Medical Devices and in Vitro Diagnostic Medical Devices, COM(2025) 1023 Final*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025PC1023.

European Data Protection Board. 2023. *Guidelines 9/2022 on Personal Data Breach Notification Under GDPR, Version 2.0*. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-92022-personal-data-breach-notification-under_en.

European Data Protection Board. 2025. *Guidelines 01/2025 on Pseudonymisation*. https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en.

European Parliament and Council of the European Union. 2010. *Directive 2010/63/EU of 22 September 2010 on the Protection of Animals Used for Scientific Purposes*. https://eur-lex.europa.eu/eli/dir/2010/63/oj/eng.

European Parliament and Council of the European Union. 2016. *Regulation (EU) 2016/679 (General Data Protection Regulation, GDPR)*. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Parliament and Council of the European Union. 2017a. *Regulation (EU) 2017/745 on Medical Devices (MDR)*. https://eur-lex.europa.eu/eli/reg/2017/745/oj.

European Parliament and Council of the European Union. 2017b. *Regulation (EU) 2017/746 on in Vitro Diagnostic Medical Devices (IVDR)*. https://eur-lex.europa.eu/eli/reg/2017/746/oj.

European Parliament and Council of the European Union. 2024a. *Directive (EU) 2024/2853 on Liability for Defective Products*. https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng.

European Parliament and Council of the European Union. 2024b. *Regulation (EU) 2024/1689 Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

European Parliament and Council of the European Union. 2025. *Regulation (EU) 2025/327 on the European Health Data Space (EHDS)*. https://eur-lex.europa.eu/eli/reg/2025/327/oj.

European Parliament and Council of the European Union. 2026. *Regulation (EU) 2026/1744 Amending Regulation (EU) 2024/1689 (Digital Omnibus on AI)*. https://eur-lex.europa.eu/eli/reg/2026/1744/oj.

Federal Office for Information Security (BSI). 2025. *Generative KI-Modelle: Chancen Und Risiken Für Industrie Und Behörden*. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf?__blob=publicationFile&v=7.

Federal Republic of Germany. 2006. *Animal Welfare Act (Tierschutzgesetz, TierSchG), § 7a*. https://www.gesetze-im-internet.de/tierschg/__7a.html.

Federal Republic of Germany. 2018. *Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG), § 27: Processing for Scientific or Historical Research Purposes and for Statistical Purposes*. https://www.gesetze-im-internet.de/bdsg_2018/__27.html.

Federal Republic of Germany. 2021. *Medical Devices User Reporting and Information Ordinance (Medizinprodukte-Anwendermelde- Und Informationsverordnung, MPAMIV), § 3: Reporting Obligation*. https://www.gesetze-im-internet.de/mpamiv/__3.html.

Federal Republic of Germany. 2024a. *Health Data Use Act (Gesundheitsdatennutzungsgesetz, GDNG), § 6: Further Processing of Care Data for Quality Assurance, Patient Safety and Research*. https://www.gesetze-im-internet.de/gdng/__6.html.

Federal Republic of Germany. 2024b. *Health Data Use Act (Gesundheitsdatennutzungsgesetz, GDNG) of 22 March 2024*. https://www.gesetze-im-internet.de/gdng/.

Federal Republic of Germany. 2026. *Act to Modernise Product Liability Law (Gesetz Zur Modernisierung Des Produkthaftungsrechts Und Zur Änderung Weiterer Vorschriften) of 22 October 2026, Article 1: Product Liability Act (ProdHaftG)*. https://www.recht.bund.de/bgbl/1/2026/321/VO.html.

Feng, Jean, Rachael V. Phillips, Ivana Malenica, et al. 2022. “Clinical Artificial Intelligence Quality Improvement: Towards Continual Monitoring and Updating of AI Algorithms in Healthcare.” *Npj Digital Medicine* 5: 66. https://doi.org/10.1038/s41746-022-00611-y.

Finlayson, Samuel G., Adarsh Subbaswamy, Karandeep Singh, et al. 2021. “The Clinician and Dataset Shift in Artificial Intelligence.” *New England Journal of Medicine* 385 (3): 283–86. https://doi.org/10.1056/NEJMc2104626.

Flesch, Rudolf. 1948. “A New Readability Yardstick.” *Journal of Applied Psychology* 32 (3): 221–33. https://doi.org/10.1037/h0057532.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software* 33 (1): 1–22. https://doi.org/10.18637/jss.v033.i01.

Fryback, Dennis G., and John R. Thornbury. 1991. “The Efficacy of Diagnostic Imaging.” *Medical Decision Making* 11 (2): 88–94. https://doi.org/10.1177/0272989X9101100203.

Gallifant, Jack, Majid Afshar, Saleem Ameen, et al. 2025. “The TRIPOD-LLM Reporting Guideline for Studies Using Large Language Models.” *Nature Medicine* 31 (1): 60–69. https://doi.org/10.1038/s41591-024-03425-5.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, et al. 2021. “Datasheets for Datasets.” *Communications of the ACM* 64 (12): 86–92. https://doi.org/10.1145/3458723.

Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian E. Raftery. 2007. “Probabilistic Forecasts, Calibration and Sharpness.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69 (2): 243–68. https://doi.org/10.1111/j.1467-9868.2007.00587.x.

Goddard, Kate, Abdul Roudsari, and Jeremy C. Wyatt. 2012. “Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators.” *Journal of the American Medical Informatics Association* 19 (1): 121–27. https://doi.org/10.1136/amiajnl-2011-000089.

Greenwald, Noah F., Geneva Miller, Erick Moen, Alex Kong, et al. 2022. “Whole-Cell Segmentation of Tissue Images with Human-Level Performance Using Large-Scale Data Annotation and Deep Learning.” *Nature Biotechnology* 40 (4): 555–65. https://doi.org/10.1038/s41587-021-01094-0.

Greshake, Kai, Sahar Abdelnabi, Shailesh Mishra, Christoph Endres, Thorsten Holz, and Mario Fritz. 2023. “Not What You’ve Signed up for: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection.” *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security*, 79–90. https://doi.org/10.1145/3605764.3623985.

Griffith, Gareth J., Tim T. Morris, Matthew J. Tudball, et al. 2020. “Collider Bias Undermines Our Understanding of COVID-19 Disease Risk and Severity.” *Nature Communications* 11: 5749. https://doi.org/10.1038/s41467-020-19478-2.

Grimm, Volker, Steven F. Railsback, Christian E. Vincenot, et al. 2020. “The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism.” *Journal of Artificial Societies and Social Simulation* 23 (2): 7. https://doi.org/10.18564/jasss.4259.

Guermazi, Ali, Chadi Tannoury, Andrew J. Kompel, Akira M. Murakami, et al. 2022. “Improving Radiographic Fracture Recognition Performance and Efficiency Using Artificial Intelligence.” *Radiology* 302 (3): 627–36. https://doi.org/10.1148/radiol.210937.

Hartigan, J. A., and M. A. Wong. 1979. “Algorithm AS 136: A K-Means Clustering Algorithm.” *Applied Statistics* 28 (1): 100–108. https://doi.org/10.2307/2346830.

Hennig, Christian. 2007. “Cluster-Wise Assessment of Cluster Stability.” *Computational Statistics & Data Analysis* 52 (1): 258–71. https://doi.org/10.1016/j.csda.2006.11.025.

Hennig, Christian. 2026. *Fpc: Flexible Procedures for Clustering*. https://doi.org/10.32614/CRAN.package.fpc.

Hernán, Miguel A., Sonia Hernández-Díaz, and James M. Robins. 2004. “A Structural Approach to Selection Bias.” *Epidemiology* 15 (5): 615–25. https://doi.org/10.1097/01.ede.0000135174.63482.43.

Hernán, Miguel A., John Hsu, and Brian Healy. 2019. “A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks.” *CHANCE* 32 (1): 42–49. https://doi.org/10.1080/09332480.2019.1579578.

Hernán, Miguel A., and James M. Robins. 2016. “Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available.” *American Journal of Epidemiology* 183 (8): 758–64. https://doi.org/10.1093/aje/kwv254.

Hirshberg, Asher, Bradford G. Scott, Thomas Granchi, et al. 2005. “How Does Casualty Load Affect Trauma Care in Urban Bombing Incidents? A Quantitative Analysis.” *The Journal of Trauma: Injury, Infection, and Critical Care* 58 (4): 686–95. https://doi.org/10.1097/01.ta.0000159243.70507.86.

Hirshberg, Asher, Michael Stein, and Raphael Walden. 1999. “Surgical Resource Utilization in Urban Terrorist Bombing: A Computer Simulation.” *The Journal of Trauma: Injury, Infection, and Critical Care* 47 (3): 545–50. https://doi.org/10.1097/00005373-199909000-00020.

Hoffrage, Ulrich, and Gerd Gigerenzer. 1998. “Using Natural Frequencies to Improve Diagnostic Inferences.” *Academic Medicine* 73 (5): 538–40. https://doi.org/10.1097/00001888-199805000-00024.

Hoffrage, Ulrich, Samuel Lindsey, Ralph Hertwig, and Gerd Gigerenzer. 2000. “Communicating Statistical Information.” *Science* 290 (5500): 2261–62. https://doi.org/10.1126/science.290.5500.2261.

Holmer, Amadeus, Jörg Marotz, Philip Wahl, Michael Dau, and Peer W. Kämmerer. 2018. “Hyperspectral Imaging in Perfusion and Wound Diagnostics – Methods and Algorithms for the Determination of Tissue Parameters.” *Biomedical Engineering / Biomedizinische Technik* 63 (5): 547–56. https://doi.org/10.1515/bmt-2017-0155.

Holtzman, Ari, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. *The Curious Case of Neural Text Degeneration*. https://doi.org/10.48550/arXiv.1904.09751.

Horst, Fabian, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Müller, and Wolfgang I. Schöllhorn. 2019. “Explaining the Unique Nature of Individual Gait Patterns with Deep Learning.” *Scientific Reports* 9 (1): 2391. https://doi.org/10.1038/s41598-019-38748-8.

Hubert, Lawrence, and Phipps Arabie. 1985. “Comparing Partitions.” *Journal of Classification* 2 (1): 193–218. https://doi.org/10.1007/BF01908075.

Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, et al. 2012. *Statistical Disclosure Control*. Wiley. https://doi.org/10.1002/9781118348239.

Hurlbert, Stuart H. 1984. “Pseudoreplication and the Design of Ecological Field Experiments.” *Ecological Monographs* 54 (2): 187–211. https://doi.org/10.2307/1942661.

Hyndman, Rob J., and George Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3rd ed. OTexts. https://otexts.com/fpp3/.

Hyndman, Rob J., and Yeasmin Khandakar. 2008. “Automatic Time Series Forecasting: The forecast Package for R.” *Journal of Statistical Software* 27 (3): 1–22. https://doi.org/10.18637/jss.v027.i03.

Hyndman, Rob J., and Anne B. Koehler. 2006. “Another Look at Measures of Forecast Accuracy.” *International Journal of Forecasting* 22 (4): 679–88. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Inouye, Sharon K., Sidney T. Bogardus, Peter A. Charpentier, et al. 1999. “A Multicomponent Intervention to Prevent Delirium in Hospitalized Older Patients.” *New England Journal of Medicine* 340 (9): 669–76. https://doi.org/10.1056/NEJM199903043400901.

International Committee of Medical Journal Editors. 2026. *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. https://www.icmje.org/recommendations/.

International Council for Harmonisation. 2019. *ICH E9(R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials*. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Ji, Ziwei, Nayeon Lee, Rita Frieske, et al. 2023. “Survey of Hallucination in Natural Language Generation.” *ACM Computing Surveys* 55 (12): 1–38. https://doi.org/10.1145/3571730.

Jones, Spencer S., Alun Thomas, R. Scott Evans, et al. 2008. “Forecasting Daily Patient Volumes in the Emergency Department.” *Academic Emergency Medicine* 15 (2): 159–70. https://doi.org/10.1111/j.1553-2712.2007.00032.x.

Jurafsky, Daniel, and James H. Martin. 2026. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*. 3rd (online draft). https://web.stanford.edu/~jurafsky/slp3/.

Justice, Amy C., Kenneth E. Covinsky, and Jesse A. Berlin. 1999. “Assessing the Generalizability of Prognostic Information.” *Annals of Internal Medicine* 130 (6): 515–24. https://doi.org/10.7326/0003-4819-130-6-199903160-00016.

Kapoor, Sayash, and Arvind Narayanan. 2023. “Leakage and the Reproducibility Crisis in Machine-Learning-Based Science.” *Patterns* 4 (9): 100804. https://doi.org/10.1016/j.patter.2023.100804.

Kilkenny, Carol, Nick Parsons, Ed Kadyszewski, et al. 2009. “Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals.” *PLoS ONE* 4 (11): e7824. https://doi.org/10.1371/journal.pone.0007824.

Kuo, Rachel Y. L., Conrad Harrison, Terry-Ann Curran, Benjamin Jones, et al. 2022. “Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis.” *Radiology* 304 (1): 50–62. https://doi.org/10.1148/radiol.211785.

Kurland, Brenda F., Laura L. Johnson, Brian L. Egleston, and Paula H. Diehr. 2009. “Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims.” *Statistical Science* 24 (2): 211–22. https://doi.org/10.1214/09-STS293.

Laird, Nan M., and James H. Ware. 1982. “Random-Effects Models for Longitudinal Data.” *Biometrics* 38 (4): 963–74. https://doi.org/10.2307/2529876.

Lazic, Stanley E. 2010. “The Problem of Pseudoreplication in Neuroscientific Studies: Is It Affecting Your Analysis?” *BMC Neuroscience* 11: 5. https://doi.org/10.1186/1471-2202-11-5.

Lazic, Stanley E., Charlie J. Clarke-Williams, and Marcus R. Munafò. 2018. “What Exactly Is ‘N’ in Cell Culture and Animal Experiments?” *PLOS Biology* 16 (4): e2005282. https://doi.org/10.1371/journal.pbio.2005282.

Leeflang, Mariska M. G., Anne W. S. Rutjes, Johannes B. Reitsma, Lotty Hooft, and Patrick M. M. Bossuyt. 2013. “Variation of a Test’s Sensitivity and Specificity with Disease Prevalence.” *Canadian Medical Association Journal* 185 (11): E537–44. https://doi.org/10.1503/cmaj.121286.

Leek, Jeffrey T., Robert B. Scharpf, Héctor Corrada Bravo, David Simcha, et al. 2010. “Tackling the Widespread and Critical Impact of Batch Effects in High-Throughput Data.” *Nature Reviews Genetics* 11 (10): 733–39. https://doi.org/10.1038/nrg2825.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, et al. 2020. *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. https://doi.org/10.48550/arXiv.2005.11401.

Little, John D. C. 1961. “A Proof for the Queuing Formula: $L = \lambda W$.” *Operations Research* 9 (3): 383–87. https://doi.org/10.1287/opre.9.3.383.

Little, Roderick J., Ralph D’Agostino, Michael L. Cohen, Kay Dickersin, Scott S. Emerson, et al. 2012. “The Prevention and Treatment of Missing Data in Clinical Trials.” *New England Journal of Medicine* 367 (14): 1355–60. https://doi.org/10.1056/NEJMsr1203730.

Liu, Xiaoxuan, Samantha Cruz Rivera, David Moher, et al. 2020. “Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension.” *Nature Medicine* 26 (9): 1364–74. https://doi.org/10.1038/s41591-020-1034-x.

Lu, Guolan, and Baowei Fei. 2014. “Medical Hyperspectral Imaging: A Review.” *Journal of Biomedical Optics* 19 (1): 010901. https://doi.org/10.1117/1.JBO.19.1.010901.

Marcantonio, Edward R., Jonathan M. Flacker, R. John Wright, and Neil M. Resnick. 2001. “Reducing Delirium After Hip Fracture: A Randomized Trial.” *Journal of the American Geriatrics Society* 49 (5): 516–22. https://doi.org/10.1046/j.1532-5415.2001.49108.x.

Mathis, Alexander, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, et al. 2018. “DeepLabCut: Markerless Pose Estimation of User-Defined Body Parts with Deep Learning.” *Nature Neuroscience* 21 (9): 1281–89. https://doi.org/10.1038/s41593-018-0209-y.

Matthews, J. N. S., Douglas G. Altman, M. J. Campbell, and Patrick Royston. 1990. “Analysis of Serial Measurements in Medical Research.” *BMJ* 300 (6719): 230–35. https://doi.org/10.1136/bmj.300.6719.230.

McInnes, Leland. 2026. *Using UMAP for Clustering*. https://umap-learn.readthedocs.io/en/latest/clustering.html.

McInnes, Leland, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. “UMAP: Uniform Manifold Approximation and Projection.” *Journal of Open Source Software* 3 (29): 861. https://doi.org/10.21105/joss.00861.

Medical Device Coordination Group. 2025. *MDCG 2019-11 Rev. 1: Qualification and Classification of Software – Regulation (EU) 2017/745 and Regulation (EU) 2017/746*. https://health.ec.europa.eu/system/files/2025-06/mdcg_2019_11_en.pdf.

Miller, Mark Roman, Fernanda Herrera, Hanseul Jun, James A. Landay, and Jeremy N. Bailenson. 2020. “Personal Identifiability of User Tracking Data During Observation of 360-Degree VR Video.” *Scientific Reports* 10: 17404. https://doi.org/10.1038/s41598-020-74486-y.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, et al. 2019. “Model Cards for Model Reporting.” *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–29. https://doi.org/10.1145/3287560.3287596.

Monks, Thomas, Christine S. M. Currie, Bhakti Stephan Onggo, Stewart Robinson, Martin Kunc, and Simon J. E. Taylor. 2019. “Strengthening the Reporting of Empirical Simulation Studies: Introducing the STRESS Guidelines.” *Journal of Simulation* 13 (1): 55–67. https://doi.org/10.1080/17477778.2018.1442155.

Moons, Karel G. M., Johanna A. A. Damen, Tabea Kaul, et al. 2025. “PROBAST+AI: An Updated Quality, Risk of Bias, and Applicability Assessment Tool for Prediction Models

Using Regression or Artificial Intelligence Methods.” *BMJ* 388: e082505. https://doi.org/10.1136/bmj-2024-082505.

Mould, D. R., and R. N. Upton. 2012. “Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development.” *CPT: Pharmacometrics & Systems Pharmacology* 1 (9): e6. https://doi.org/10.1038/psp.2012.4.

Mould, D. R., and R. N. Upton. 2013. “Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development – Part 2: Introduction to Pharmacokinetic Modeling Methods.” *CPT: Pharmacometrics & Systems Pharmacology* 2 (4): e38. https://doi.org/10.1038/psp.2013.14.

Nagendran, Myura, Yang Chen, Christopher A. Lovejoy, et al. 2020. “Artificial Intelligence Versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies.” *BMJ* 368: m689. https://doi.org/10.1136/bmj.m689.

NC3Rs. 2026. *The 3Rs*. https://nc3rs.org.uk/who-we-are/3rs.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations.” *Science* 366 (6464): 447–53. https://doi.org/10.1126/science.aax2342.

OWASP Foundation. 2024. *OWASP Top 10 for LLM Applications 2025*. https://genai.owasp.org/llm-top-10/.

Paszke, Adam, Sam Gross, Francisco Massa, et al. 2019. “PyTorch: An Imperative Style, High-Performance Deep Learning Library.” *Advances in Neural Information Processing Systems 32*, 8024–35. https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research* 12: 2825–30. https://jmlr.org/papers/v12/pedregosa11a.html.

Peng, Roger D. 2011. “Reproducible Research in Computational Science.” *Science* 334 (6060): 1226–27. https://doi.org/10.1126/science.1213847.

Percie du Sert, Nathalie, Viki Hurst, Amrita Ahluwalia, et al. 2020. “The ARRIVE Guidelines 2.0: Updated Guidelines for Reporting Animal Research.” *PLOS Biology* 18 (7): e3000410. https://doi.org/10.1371/journal.pbio.3000410.

Petzold, Linda. 1983. “Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations.” *SIAM Journal on Scientific and Statistical Computing* 4 (1): 136–48. https://doi.org/10.1137/0904010.

Quarto. n.d. *Creating a Book*. https://quarto.org/docs/books/.

R Core Team. 2026. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rackauckas, Christopher, and Qing Nie. 2017. “DifferentialEquations.jl – a Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia.” *Journal of Open Research Software* 5 (1): 15. https://doi.org/10.5334/jors.151.

Raue, A., C. Kreutz, T. Maiwald, et al. 2009. “Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood.” *Bioinformatics* 25 (15): 1923–29. https://doi.org/10.1093/bioinformatics/btp358.

Riley, Richard D., Joie Ensor, Kym I. E. Snell, et al. 2020. “Calculating the Sample Size Required for Developing a Clinical Prediction Model.” *BMJ* 368: m441. https://doi.org/10.1136/bmj.m441.

Roberts, David R., Volker Bahn, Simone Ciuti, et al. 2017. “Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure.” *Ecography* 40 (8): 913–29. https://doi.org/10.1111/ecog.02881.

Roberts, Michael, Derek Driggs, Matthew Thorpe, et al. 2021. “Common Pitfalls and Recommendations for Using Machine Learning to Detect and Prognosticate for COVID-19 Using Chest Radiographs and CT Scans.” *Nature Machine Intelligence* 3 (3): 199–217. https://doi.org/10.1038/s42256-021-00307-0.

Robin, Xavier, Natacha Turck, Alexandre Hainard, et al. 2011. “pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves.” *BMC Bioinformatics* 12: 77. https://doi.org/10.1186/1471-2105-12-77.

Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre de Montjoye. 2019. “Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models.” *Nature Communications* 10: 3069. https://doi.org/10.1038/s41467-019-10933-3.

Rousseeuw, Peter J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Journal of Computational and Applied Mathematics* 20: 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Rubin, Donald B. 1976. “Inference and Missing Data.” *Biometrika* 63 (3): 581–92. https://doi.org/10.1093/biomet/63.3.581.

Sale, Mark, and Eric A. Sherer. 2015. “A Genetic Algorithm Based Global Search Strategy for Population Pharmacokinetic/Pharmacodynamic Model Selection.” *British Journal of Clinical Pharmacology* 79 (1): 28–39. https://doi.org/10.1111/bcp.12179.

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. “Ten Simple Rules for Reproducible Computational Research.” *PLoS Computational Biology* 9 (10): e1003285. https://doi.org/10.1371/journal.pcbi.1003285.

Sargent, Robert G. 2013. “Verification and Validation of Simulation Models.” *Journal of Simulation* 7 (1): 12–24. https://doi.org/10.1057/jos.2012.20.

Schapiro, Denis, Hartland W. Jackson, Swetha Raghuraman, Jana R. Fischer, et al. 2017. “histoCAT: Analysis of Cell Phenotypes and Interactions in Multiplex Image Cytometry Data.” *Nature Methods* 14 (9): 873–76. https://doi.org/10.1038/nmeth.4391.

Schillinger, Dean, John Piette, Kevin Grumbach, et al. 2003. “Closing the Loop: Physician Communication with Diabetic Patients Who Have Low Health Literacy.” *Archives of Internal Medicine* 163 (1): 83–90. https://doi.org/10.1001/archinte.163.1.83.

Scrucca, Luca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. 2016. “Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.” *The R Journal* 8 (1): 289–317. https://doi.org/10.32614/RJ-2016-021.

Seelig, Julian, Raban A. Heller, Julian Hackler, et al. 2020. “Selenium and Copper Status – Potential Signposts for Neurological Remission After Traumatic Spinal Cord Injury.” *Journal of Trace Elements in Medicine and Biology* 57: 126415. https://doi.org/10.1016/j.jtemb.2019.126415.

Seok, Junhee, H. Shaw Warren, Alex G. Cuenca, et al. 2013. “Genomic Responses in Mouse Models Poorly Mimic Human Inflammatory Diseases.” *Proceedings of the National Academy of Sciences* 110 (9): 3507–12. https://doi.org/10.1073/pnas.1222878110.

Seyyed-Kalantari, Laleh, Haoran Zhang, Matthew B. A. McDermott, Irene Y. Chen, and Marzyeh Ghassemi. 2021. “Underdiagnosis Bias of Artificial Intelligence Algorithms Applied to Chest Radiographs in Under-Served Patient Populations.” *Nature Medicine* 27 (12): 2176–82. https://doi.org/10.1038/s41591-021-01595-0.

Shmueli, Galit. 2010. “To Explain or to Predict?” *Statistical Science* 25 (3): 289–310. https://doi.org/10.1214/10-STS330.

Shokri, Reza, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. “Membership Inference Attacks Against Machine Learning Models.” *2017 IEEE Symposium on Security and Privacy (SP)*, 3–18. https://doi.org/10.1109/SP.2017.41.

Singhal, Karan, Shekoofeh Azizi, Tao Tu, et al. 2023. “Large Language Models Encode Clinical Knowledge.” *Nature* 620 (7972): 172–80. https://doi.org/10.1038/s41586-023-06291-2.

Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, et al. 2020. “Racial Bias in Pulse Oximetry Measurement.” *New England Journal of Medicine* 383 (25): 2477–78. https://doi.org/10.1056/NEJMc2029240.

Soetaert, Karline, Thomas Petzoldt, and R. Woodrow Setzer. 2010. “Solving Differential Equations in R: Package deSolve.” *Journal of Statistical Software* 33 (9): 1–25. https://doi.org/10.18637/jss.v033.i09.

Sounderajah, Viknesh, Ahmad Guni, Xiaoxuan Liu, et al. 2025. “The STARD-AI Reporting Guideline for Diagnostic Accuracy Studies Using Artificial Intelligence.” *Nature Medicine* 31: 3283–89. https://doi.org/10.1038/s41591-025-03953-8.

Souppaya, Murugiah, Karen Scarfone, and Donna Dodson. 2022. *Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities*. NIST SP 800-218. National Institute of Standards; Technology. https://doi.org/10.6028/NIST.SP.800-218.

Spiegelhalter, David J., Keith R. Abrams, and Jonathan P. Myles. 2004. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley. https://doi.org/10.1002/0470092602.

Squair, Jordan W., Matthieu Gautier, Claudia Kathe, Mark A. Anderson, et al. 2021. “Confronting False Discoveries in Single-Cell Differential Expression.” *Nature Communications* 12 (1): 5692. https://doi.org/10.1038/s41467-021-25960-2.

Stenum, Jan, Cristina Rossi, and Ryan T. Roemmich. 2021. “Two-Dimensional Video-Based Analysis of Human Gait Using Pose Estimation.” *PLOS Computational Biology* 17 (4): e1008935. https://doi.org/10.1371/journal.pcbi.1008935.

Steyerberg, Ewout W., Andrew J. Vickers, Nancy R. Cook, et al. 2010. “Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures.” *Epidemiology* 21 (1): 128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2.

Stringer, Carsen, Tim Wang, Michalis Michaelos, and Marius Pachitariu. 2021. “Cellpose: A Generalist Algorithm for Cellular Segmentation.” *Nature Methods* 18 (1): 100–106. https://doi.org/10.1038/s41592-020-01018-x.

Swanson, Marianne, Pauline Bowen, Amy Wohl Phillips, Dean Gallup, and David Lynes. 2010. *Contingency Planning Guide for Federal Information Systems*. NIST SP 800-34 Rev. 1. National Institute of Standards; Technology. https://doi.org/10.6028/NIST.SP.800-34r1.

Sweeney, Latanya. 2002. “K-Anonymity: A Model for Protecting Privacy.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (5): 557–70. https://doi.org/10.1142/S0218488502001648.

Takao, Keizo, and Tsuyoshi Miyakawa. 2015. “Genomic Responses in Mouse Models Greatly Mimic Human Inflammatory Diseases.” *Proceedings of the National Academy of Sciences* 112 (4): 1167–72. https://doi.org/10.1073/pnas.1401965111.

Tashman, Leonard J. 2000. “Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review.” *International Journal of Forecasting* 16 (4): 437–50. https://doi.org/10.1016/S0169-2070(00)00065-0.

Team SimPy. n.d. *SimPy: Discrete Event Simulation for Python*. https://simpy.readthedocs.io/en/latest/.

Textor, Johannes, Benito van der Zander, Mark S. Gilthorpe, Maciej Liśkiewicz, and George T. H. Ellison. 2016. “Robust Causal Inference Using Directed Acyclic Graphs: The R Package ‘Dagitty’.” *International Journal of Epidemiology* 45 (6): 1887–94. https://doi.org/10.1093/ije/dyw341.

Tibshirani, Robert, and Guenther Walther. 2005. “Cluster Validation by Prediction Strength.” *Journal of Computational and Graphical Statistics* 14 (3): 511–28. https://doi.org/10.1198/106186005X59243.

Ucar, Iñaki, Bart Smeets, and Arturo Azcorra. 2019. “simmer: Discrete-Event Simulation for R.” *Journal of Statistical Software* 90 (2): 1–30. https://doi.org/10.18637/jss.v090.i02.

Van Calster, Ben, David J. McLernon, Maarten van Smeden, Laure Wynants, and Ewout W. Steyerberg. 2019. “Calibration: The Achilles Heel of Predictive Analytics.” *BMC Medicine* 17: 230. https://doi.org/10.1186/s12916-019-1466-7.

Varma, Sudhir, and Richard Simon. 2006. “Bias in Error Estimation When Using Cross-Validation for Model Selection.” *BMC Bioinformatics* 7: 91. https://doi.org/10.1186/1471-2105-7-91.

Vasey, Baptiste, Myura Nagendran, Bruce Campbell, et al. 2022. “Reporting Guideline for the Early Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI.” *BMJ* 377: e070904. https://doi.org/10.1136/bmj-2022-070904.

Vassilev, Apostol, Alina Oprea, Alie Fordyce, Hyrum Anderson, Xander Davies, and Maia Hamin. 2025. *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*. NIST AI 100-2e2025. National Institute of Standards; Technology. https://doi.org/10.6028/NIST.AI.100-2e2025.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. 2017. *Attention Is All You Need*. https://doi.org/10.48550/arXiv.1706.03762.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. Springer. https://doi.org/10.1007/978-0-387-21706-2.

Vickers, Andrew J., and Elena B. Elkin. 2006. “Decision Curve Analysis: A Novel Method for Evaluating Prediction Models.” *Medical Decision Making* 26 (6): 565–74. https://doi.org/10.1177/0272989X06295361.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17: 261–72. https://doi.org/10.1038/s41592-019-0686-2.

Walt, Stéfan van der, Johannes L. Schönberger, Juan Nunez-Iglesias, et al. 2014. “Scikit-Image: Image Processing in Python.” *PeerJ* 2: e453. https://doi.org/10.7717/peerj.453.

Walters, William H., and Esther Isabelle Wilder. 2023. “Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT.” *Scientific Reports* 13: 14045. https://doi.org/10.1038/s41598-023-41032-5.

Wickham, Hadley. 2014. “Tidy Data.” *Journal of Statistical Software* 59 (10): 1–23. https://doi.org/10.18637/jss.v059.i10.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18.

Wilson, Edwin B. 1927. “Probable Inference, the Law of Succession, and Statistical Inference.” *Journal of the American Statistical Association* 22 (158): 209–12.

https://doi.org/10.1080/01621459.1927.10502953.
Wilson, Greg, Jennifer Bryan, Karen Cranston, et al. 2017. “Good Enough Practices in Scientific Computing.” *PLoS Computational Biology* 13 (6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510.
Wong, Andrew, Erkin Otles, John P. Donnelly, et al. 2021. “External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients.” *JAMA Internal Medicine* 181 (8): 1065–70. https://doi.org/10.1001/jamainternmed.2021.2626.
World Health Organization. 2024. *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*. World Health Organization. https://www.who.int/publications/i/item/9789240084759.
Worp, H. Bart van der, David W. Howells, Emily S. Sena, et al. 2010. “Can Animal Models of Disease Reliably Inform Human Studies?” *PLoS Medicine* 7 (3): e1000245. https://doi.org/10.1371/journal.pmed.1000245.
Zech, John R., Marcus A. Badgeley, Manway Liu, et al. 2018. “Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study.” *PLOS Medicine* 15 (11): e1002683. https://doi.org/10.1371/journal.pmed.1002683.
Ziemann, Mark, Yotam Eren, and Assam El-Osta. 2016. “Gene Name Errors Are Widespread in the Scientific Literature.” *Genome Biology* 17: 177. https://doi.org/10.1186/s13059-016-1044-7.
Zimmerman, Kip D., Mark A. Espeland, and Carl D. Langefeld. 2021. “A Practical Solution to Pseudoreplication Bias in Single-Cell Studies.” *Nature Communications* 12 (1): 738. https://doi.org/10.1038/s41467-021-21038-1.
Zou, Hui, and Trevor Hastie. 2005. “Regularization and Variable Selection via the Elastic Net.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 67 (2): 301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.