Case Studies in Generating Real World Evidence From Real World Data References

Case Study 1: EHR Data Quality

Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, Denny JC, Peissig PL, Miller AW, Wei WQ, Bielinski SJ. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association*. 2012;19(2):212-8.

Denny JC. Mining electronic health records in the genomics era. *PLoS computational biology*. 2012;8(12):e1002823.

Forrest, C, Margolis, P, Bailey, C, Marsolo, K, Beccaro, M, Finkelstein, J, Milov, D, Vieland, V, Wolf, B, Yu, F, and Kahn, M., PEDSnet: A National Pediatric Learning Health System, *Journal of the American Medical Informatics Association.* 2014; 21: 602-606.

Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*. 2012;13(6):395-405.

Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, Pacheco JA, Tromp G, Pathak J, Carrell DS, Ellis SB. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association.* 2016;23(6):1046-52.

Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*. 1997;146(2):195-203.

Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. 2013;20(1):144-51.

Case Study 2: Combining RCTs and RWD

Rugo HS, Dieras V, Cortes J, Patt D, Wildiers H, O'Shaughnessy J, Zamora E, Yardley DA, Carter GC, Sheffield KM, Li L. Real-world survival outcomes of heavily pretreated patients with refractory HR+, HER2− metastatic breast cancer receiving single-agent chemotherapy—a comparison with MONARCH 1. *Breast Cancer Research and Treatment*. 2020;184(1):161-72.

Burcu M, Dreyer NA, Franklin JM, Blum MD, Critchlow CW, Perfetto EM, Zhou W. Real-world evidence to support regulatory decision-making for medicines: Considerations for external control arms. *Pharmacoepidemiology and Drug Safety*. 2020; 29(10):1228-35.

Carrigan G, Whipple S, Taylor MD, Torres AZ, Gossai A, Arnieri B, Tucker M, Hofmeister PP, Lambert P, Griffith SD, Capra WB. An evaluation of the impact of missing deaths on overall survival analyses of advanced non–small cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiology and Drug Safety*. 2019;28(5):572-81.

Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011; 67(3):1047-56.

Pocock SJ. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*. 1976;29(3):175-88.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1):41-55.

Ventz S, Lai A, Cloughesy TF, Wen PY, Trippa L, Alexander BM. Design and evaluation of an external control arm using prior clinical trials and real-world data. *Clinical Cancer Research*. 2019; 25(16):4993-5001.

Wang C, Li H, Chen WC, Lu N, Tiwari R, Xu Y, Yue LQ. Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *Journal of Biopharmaceutical Statistics*. 2019; 29(5):731-48.

Case Study 3: Distributed analysis

Tong, J., Chen, Z., Duan, R., Lo-Ciganic, W., Lyu, T., Tao, C., Merkel, P., Kranzler, H., Bian, J., Chen, Y.. (2020) Identifying Clinical Risk Factors of Opioid Use Disorder using a Distributed Algorithm to Combine Real-World Data from a Large Clinical Data Research Network. *AMIA Annu Symp Proc.*

Duan R, Boland MR, Moore JH, Chen Y. ODAL: A one-shot distributed algorithm to perform logistic regressions on electronic health records data from multiple clinical sites. *Pacific Symposium on Biocomputing* 2019, 24, 30–41

Duan R, Boland MR, Liu Z, Liu Y, Chang HH, Xu H, Chu H, Schmid CH, Forrest CB, Holmes JH, Schuemie MJ, Berlin J, Moore J, and Chen Y. Learning from electronic health records across multiple sites: A communication-efficient and privacy-preserving distributed algorithm. *Journal of the American Medical Informatics Association*. 2020 Mar;27(3):376-85

Tong J, Duan R, Li R, Scheuemie MJ, Moore JH, Chen Y. Robust-ODAL: Learning from heterogeneous health systems without sharing patient-level data. *Pacific Symposium on Biocomputing 2020*, 25, 695–706.

Duan, R., Chen Z., Tong, J., Luo, C., Lyu, T., Tao, C., Maraganore, D., Bian, J. and Chen, Y.. (2020) Leverage real-world longitudinal data in large clinical research networks for Alzheimer's disease and related dementia. *AMIA Annu Symp Proc*.

Duan R, Luo C, Schuemie MJ, Tong J, Liang CJ, Chang HH, Boland MR, Bian J, Xu H, Holmes JH, Forrest CB, Morton SC, Berlin JA, Moore JH, Mahoney KB, Chen Y. Learning from local to global: An efficient distributed algorithm for modeling time-to-event data. *Journal of the American Medical Informatics Association*. 2020 Jul;27(7):1028-1036.

Edmondson, M.J., Luo, C., Duan, R., Maltenfort, M., Chen, Z., Shults, J., Bian, J., Ryan, P., Forrest, C. and Chen, Y., 2020. Distributed Learning from Multi-Site Observational Health Data for Zero-Inflated Count Outcomes. *medRxiv*.
https://www.medrxiv.org/content/10.1101/2020.12.17.20248194v1

Duan, R., Ning, Y. and Chen, Y. (2021) 'Heterogeneity-aware and communication-efficient distributed statistical inference', *Biometrika*. doi: 10.1093/biomet/asab007.