

## Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2022, Section Cancer Informatics

Liu R, Rizzo S, Whipple S, Pal N, Pineda AL, Lu M, Arnieri B, Lu Y, Capra W, Copping R, Zou J

**Evaluating eligibility criteria of oncology trials using real-world data and AI**  
*Nature* 2021 Apr;592(7855):629-33

A computational framework called Trial Pathfinder was developed and evaluated on a series of non-small cell lung cancer (NSCLC) trials to determine whether

changing trial eligibility criteria would change the results. The simulations were conducted using a large real-world data source of 61,094 patients with advanced NSCLC. The authors determined that relaxing eligibility criteria would result in small changes to hazard ratios, while potentially expanding the pool of eligible patients by at least two-fold. This has major implications for the future design of cancer clinical trials.

Yang DX, Khera R, Miccio JA, Jairam V, Chang E, Yu JB, Park HS, Krumholz HM, Aneja S

**Prevalence of missing data in the National Cancer Database and association with overall survival**

**JAMA Netw Open** 2021 Mar 1;4(3):e211793

The authors assessed the prevalence of missing data in a very large registry of patients with cancer, and whether missingness in itself was prognostic. They found substantial missingness, e.g., with 71% of patients with NSCLC missing data for variables of interest. When compared to patients with complete case data, they found a statistically significant difference in 2-year overall survival in lung, breast, and prostate cancer. This study demonstrates the importance of metadata characteristics in the conduct and evaluation of real-world data registry studies.