**Appendix**

**Methods:** Let Z = 1 denote the use of the RS assay and Z=0 for not, and let Crepresent the vector of all variables with no missing values and the intercept. Let Y be the status of chemotherapy use with some potentially missing values, and let Ry = 1 if Y is observed and Ry = 0 if otherwise. Let Xdenote the vector of all other variables with some missing values. Let X-j = (X1, …, Xj-1, Xj+1, …, Xk) denote the collection of the k-1 variables in X except Xj; and let Rj be the response indicator of Xj with Rj = 1 if Xj is observed and Xj = 0 if Xj is missing for any j. The observed parts of Y and X are represented by Yobs and Xobs and the missing parts by Ymis and Xmis where Xobs = {x1obs, x2obs, …, xkobs} and Xmis = {x1mis, x2mis, …, xkmis}. In a causal framework, we hypothesize that there are two potential outcomes for each patient, with Y1 denoting her chemotherapy selection if RS testing occurs (Z=1) and Y0 otherwise (Z=0). If we know both Y1 and Y0, we can infer whether a test causes a change in chemotherapy selection by comparing Y1 to Y0, and calculate any causal quantity, such as causal rate differences in chemotherapy use. However, in a general practice setting, we only observe either Y0 or Y1 and not both because a patient either does or does not receive a test. There are four possible patterns of the pair Y0 and Y1: (Y0=0, Y1=0) for the never-chemotherapy group, (Y0=0, Y1=1) for the chemotherapy-encouraged, (Y0=1, Y1=0) for the chemotherapy-discouraged, and (Y0 =1, Y1 =1) for the always-chemotherapy group.

We assume that the data will be missing at random.1 We further assume that: 1) independence of potential outcomes given observed data; 2) stable unit treatment value assumption, requiring an individual’s chemotherapy use unaffected by RS assay status of others; 3) no unmeasured confounders. These assumptions are commonly made in the causal inference literature, and will be essential to facilitating identification and estimation. We propose to impute the missing values in X and the missing potential outcome (either Y0 or Y1) iteratively variable-by-variable using a sequence of conditional regression models until convergence: P(Xj |C, X-j, Z, Y0, Y1), P(Y0| X, C), P(Y1| C, X).2,3 To account for uncertainty about the imputation of the unknown missing values, we will conduct 5 imputations. Then, we will obtain estimates and variances of quantities of interest based on each imputed data. Finally, we will apply Rubin’s formula to combine these estimates and variances across 5 imputed data sets.4 Multiple imputation techniques have been used in literature to conduct causal inference.5,6 However, it has not been used for patient classification and effect quantifications as in our context.

1. Little RJA and Rubin DB: Statistical Analysis with Missing Data (2nd ed). Hoboken, NY, John Wiley & Sons, 2002.
2. van Buuren S: Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research16:219-242, 2007.
3. Raghunathan T, Lepkowski J, VanHoewyk J, Solenberger P: A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology 27:85-96, 2001.
4. Rubin DB: Multiple Imputation for Nonresponse in Surveys. Hoboken, NY, John Wiley & Sons, 1987.
5. Westreich D, Edwards JK, Cole SR, et al: Imputation approaches for potential outcomes in causal inference. International Journal of Epidemiology 44:1731-1737, 2015.
6. Bondarenko I and Raghunathan T: Multiple Imputation for Causal Inference. In JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association 3934-3944, 2010. Retrieved from https://www.amstat.org/sections/srms/proceedings/y2010/Files/308027\_59882.pdf.