Prediagnostic evaluation of multicancer detection tests: Design and analysis considerations

Supplementary Methods

This section calculates the sample size based on previously developed formulas (1).

Number of cases. For cancer j, the number of cases, n_{1j}, and number of positive cases, x_{1j}, approximately solve the Clopper-Pearson (2) equation for a 0.3% lower bound, $CDF[\text{Binomial}(n_{1j}, TPR_{LOW}), x_{1j}] = 0.997$, where CDF is the cumulative distribution function. The standard target of $TPR_{LOW} = 0.70$ is satisfied by 70 cases with 59 positives because $CDF[\text{Binomial}(70, 0.70), 59] = 0.998$. The Scenario 1 sample size tradeoff target of $TPR_{LOW} = 0.20$ is satisfied by 12 cases with 7 positives because $CDF[\text{Binomial}(12, 0.20), 7] = 0.999$. The Scenario 2 sample size tradeoff target of $TPR_{LOW} = 0.50$ is satisfied by 12 cases with 10 positives because $CDF[\text{Binomial}(12, 0.50), 10] = 0.997$.

Number of controls. The required number of controls, n_0, and number of positive controls, x_0, approximately solve the Clopper-Pearson equation for a 97.5% upper bound, $CDF[\text{Binomial}(n_0, FPR_{UPP}), x_0] = 0.025$. The standard target $FPR_{UPP} = 0.03$ is satisfied by 300 controls with 3 positives because $CDF[\text{Binomial}(300, 0.03), 3] = 0.02$. The sample size tradeoff target $FPR_{UPP} = 0.0005$ is satisfied by 7500 controls with 0 positives because $CDF[\text{Binomial}(7500, 0.0005), 0] = 0.023$.

Number of specimen collections. The sample size of N specimen collections with n_{1j} cases approximately solves the Clopper-Pearson equation for a 95% upper bound, $CDF(\text{Binomial}(N, p_j), n_{1j}] = 0.05$, where j corresponds to the rarest cancer. For the standard method with a target of 70 cases the sample size is 720,000 because $CDF[\text{Binomial}(720{,}000, 0.00012), 70] = 0.04$. For the sample size tradeoff method with a target of 12 cases, the sample size is 163000 because $CDF[\text{Binomial}(163{,}000, 0.00012), 12] = 0.047$.

References

1. Baker SG and Kramer BS. Simple methods for evaluating 4 types of biomarkers: surrogate endpoint, prognostic, predictive, and cancer screening. *Biomarker Insights.* 2020; 15:1-8.

2. Clopper, CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934; 26: 404–416.