A novel method for unbiased risk prediction and its application to the combined residual risk score incorporating SNPs and the Tyrer-Cuzick_v7 breast cancer model

Elisha Hughes, PhD[1]; Susanne Wagner, PhD[1]; Jerry Lanchbury, PhD[1]; Placede Tshiaba, MS[1]; Shannon Gallagher, MPH[1]; Brian Morris, BS[1]; Johnathan Lancaster, MD, PhD[1]; Kathryn Dalton, DO, FACS[2]; Wade Hedegard, MD[3]; Carol Adami, MD[4]; Mark Robson, MD[5]; Alexander Gutin, PhD[1]

[1]Myriad Genetics, Inc., Salt Lake City, UT; [2]Cape Cod Healthcare, Hyannis, MA; [3]Elizabeth Wende Breast Care, Rochester, NY; [4]Bethesda Health, Boynton Beach, FL; [5]Memorial Sloan Kettering Cancer Center, New York, NY

BACKGROUND

- Breast cancer (BC) risk prediction models should use appropriate methods to avoid "double counting" information shared by correlated risk factors, such as correlations between family history (FHx) and polygenic risk scores (PRS) based on single nucleotide polymorphism (SNP) genotyping.
- Here we describe a novel fixed-stratified (F-S) method for incorporating new factors into an established model.

METHODS

FIXED-STRATIFIED METHOD

- To account for correlation between FHx and PRS, we first construct a univariate model to predict BC based on FHx:

Model 1: BC ~ β_1×FHx

- Next, we construct a bivariate model to estimate the effect of PRS after fixing the univariate FHx effect:

Model 2: BC ~ **1**×(β_1×FHx) + β_2×PRS

- Finally, we stratify and constrain the absolute risk estimates such that the average relative risk due to PRS is 1 within each FHx strata.

SIMULATION STUDIES

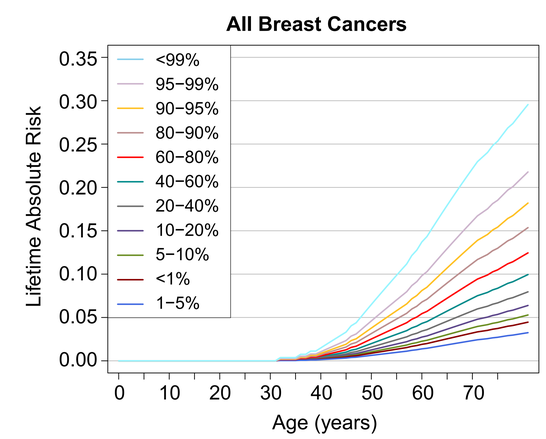
- We simulated a sample of one million women based on the study parameters reported in Mavaddat et al., 2015.[1]
- Continuous variables for BC, FHx and PRS were simulated as multivariate normal with mean zero and a specified covariance matrix.
- BC and FHx were transformed to binary by applying the univariate normal cumulative density function (CDF), followed by the inverse CDF of the Bernoulli distribution.
- For BC, we used a Bernoulli distribution with parameter p=0.5 to achieve roughly 500,000 cases and 500,000 controls.
- For FHx, we used a Bernoulli parameter of p=0.15 so that roughly 15% of the sample would be positive for FHx.

CLINICAL VALIDATION

- Combined residual risk score (cRRS) was defined by the F-S method based on analysis of correlations between SNPs and Tyrer-Cuzick_v7 (TC) factors.
- cRRS was evaluated by pre-specified validation in a prospective cohort from breast imaging centers (N=518).[2]
- Age-adjusted logistic regression was used to evaluate the improved discriminatory accuracy of cRRS over TC.

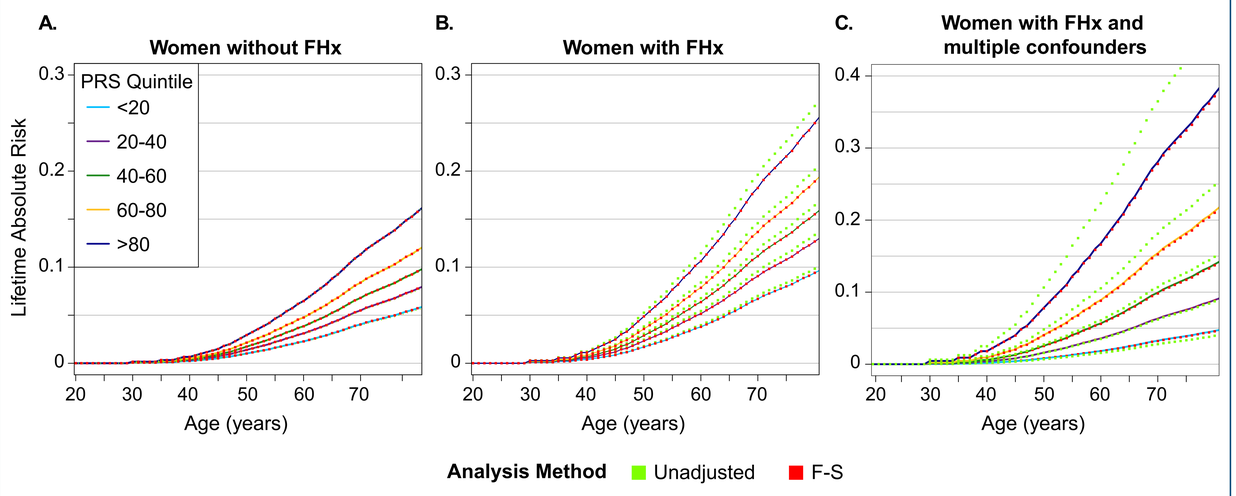
SIMULATION STUDIES

Figure 1. Simulation of Mavaddat et al. cumulative absolute risks.



- A simulated version of the Mavaddat et al. dataset exactly matched all published univariate and bivariate odds ratios for PRS and FHx.
- Based on the simulated dataset, we visually matched the published figure of cumulative absolute risk values (Figure 1).
- For women without a FHx of BC, risk estimates based on PRS and FHx were accurate regardless of whether the risk estimates had been adjusted for confounding (Figure 2A).
- The F-S method produced correct risk estimates for women with a FHx of BC, while unadjusted risk estimates overestimated the true risk (Figure 2B).
- With more extreme correlation between multiple risk factors, the F-S method was accurate while unadjusted risk estimates severely overestimated actual risk (Figure 2C).
- Simulation studies confirmed that F-S risk estimates are unbiased, while methods disregarding correlation are biased.

Figure 2. Simulation of Mavaddat et al. cumulative absolute risks with FHx.



CLINICAL VALIDATION

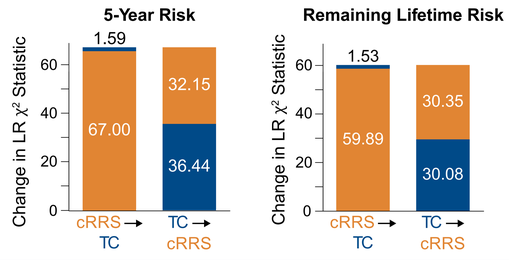
- cRRS and TC were both highly significantly associated with both a 5-year and remaining lifetime risk of BC (Table 1).
- In multivariate models including cRRS, TC, and age, cRRS provided significantly improved discriminatory accuracy over TC for both 5-year ($p=1.4\times10^{-8}$) and remaining lifetime risks ($p=3.7\times10^{-8}$; Figure 3).

Table 1. Age-Stratified Regression Analysis (N=518).

BC Risk Prediction Model	Odds Ratio (95% CI)*	p-value
5-Year Risk of Breast Cancer		
Tyrer-Cuzick	2.10 (1.63, 2.74)	1.4×10^{-9}
cRRS	2.58 (2.06, 3.34)	2.4×10^{-16}
Remaining Lifetime Risk of Breast Cancer		
Tyrer-Cuzick	1.98 (1.54, 2.58)	4.5×10^{-8}
cRRS	2.47 (1.93, 3.18)	1.9×10^{-14}

Breast cancer associations of risk prediction models based on bivariate weighted logistic regression models adjusted for age.
*Odds ratios and Wald Confidence Interval (CIs) are reported per weighted standard deviation of log odds of BC risk prediction models in unaffected controls.

Figure 3. Multivariate Analysis (N=518).



CONCLUSIONS

- A F-S method accurately combines correlated risk factors for a comprehensive assessment of breast cancer.
- cRRS may facilitate more accurate application of breast cancer prevention and screening strategies through improved discriminatory accuracy.

REFERENCES

1. Mavaddat N, et al. *JNCI J Natl Cancer Inst.* 2015;107(5):djv036.
2. Dalton K, *et al. J Clin Oncol.* 2018;36.no15_suppl.1507.

Presented at Basser Symposium May 7th, 2019