Title: Development and validation of the PREMMplus clinical prediction model for multigene hereditary cancer risk assessment.

Matthew B. Yurgelun[1,2,3], Hajime Uno[1,2], C. Sloane Furniss[1], Chinedu I. Ukaegbu[1], Miki Horiguchi[1,2], Israel Gomy[1], Anu B. Chittenden[1], Tara Fehlmann[1], Amal Yussuf[4], Holly LaDuca[4], Judy E. Garber[1,2,3], Sapna Syngal[1,2,3]

[1.] Dana-Farber Cancer Institute, Boston, MA, USA
[2.] Harvard Medical School, Boston, MA, USA
[3.] Brigham & Women's Hospital, Boston, MA, USA
[4.] Ambry Genetics, Aliso Viejo, CA

Background: Current clinical prediction models provide syndrome-specific numeric estimates of an individual's likelihood of having a specific hereditary cancer syndrome (e.g., $PREMM_5$ for Lynch syndrome; BRCAPRO for *BRCA1/2*). With the emergence of multigene panel testing, there is a need to evaluate individuals' risk of carrying a pathogenic variant in a diverse array of cancer susceptibility genes in parallel. This study's aim was to develop and validate the PREMMplus clinical prediction model for multigene cancer risk assessment.

Methods: PREMMplus was developed in a cohort of 7296 individuals who had undergone germline multigene panel testing at a single academic cancer center. Logistic regression models with LASSO regularization were used to examine candidate predictive variables – including age, sex, ethnicity, and personal/family history of cancer – to provide a numeric estimate of an individual's likelihood of carrying a pathogenic/likely pathogenic germline variant in one of 19 cancer susceptibility genes (11 "Category 1" [*APC*, *BRCA1/2*, *CDH1*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, biallelic *MUTYH*, *PMS2*, and *TP53*] and 8 "Category 2" genes [*ATM*, *BRIP1*, *CDKN2A*, *CHEK2*, *PALB2*, *PTEN*, *RAD51C*, and *RAD51D*]). Model performance was validated in an independent dataset of 14845 individuals who had undergone multigene panel testing at a commercial laboratory.

Results: Using clinical characteristics, including personal/family history of 18 cancers plus colorectal adenoma burden, PREMMplus demonstrated excellent ability to predict pathogenic variants in Category 1 genes and acceptable performance with the addition of 8 Category 2 genes. Decision curve analyses support multigene germline testing for individuals predicted to have ≥2.5% probability of harboring a pathogenic germline variant in any Category 1 gene (Table). PREMMplus was well-calibrated and demonstrated comparable performance in the external validation dataset.

Conclusions: PREMMplus is the first validated risk assessment model to quantify an individual's likelihood of carrying pathogenic variants in a wide diversity of genes associated with established high risks of various cancers. Individuals with PREMMplus scores ≥2.5% should be referred for multigene panel testing. As expected, PREMMplus's discriminatory capacity was reduced with the inclusion of moderate penetrance cancer risk genes.

Cohort	Outcome	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	# Needed to Test to Detect 1 Carrier	AUC (95% CI)
Development	11 genes	95.2%	20.4%	6.5%	98.7%	15.5	0.74 (0.71-0.77)
	19 genes	90.1%	20.4%	10.7%	95.1%	10.0	0.67 (0.64-0.69)
Validation	11 genes	85.1%	31.1%	5.8%	97.7%	17.3	0.68 (0.66 – 0.71)
	19 genes	77.5%	31.2%	10.3%	93.2%	9.7	0.60 (0.58 - 0.62)