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

Matthew B. Yurgelun,[1,2,3] Hajime Uno,[1] C. Sloane Furniss,[1] Chinedu I. Ukaegbu,[1] Miki Horiguchi,[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; [2]Harvard Medical School, Boston, MA; [3]Brigham & Women's Hospital, Boston, MA; [4]Ambry Genetics, Aliso Viejo, CA
E-mail: Matthew_Yurgelun@dfci.Harvard.edu
Funded by the NIH R01CA132829

Background:

- Syndrome-specific prediction models estimate individuals' likelihood of specific hereditary cancer syndromes (e.g., $PREMM_5$ for Lynch syndrome; BRCAPRO for *BRCA1/2*)
- Emergence of multigene panel testing (MGPT) → need to evaluate risk for diverse array of cancer risk genes in parallel
- **Primary Aim**: To develop and validate the PREMMplus clinical prediction model for multigene cancer risk assessment

Methods:

- Model development cohort:
  - Consecutive cohort of 7296 unrelated individuals undergoing MGPT at Dana-Farber from January 1, 2013 through December 31, 2017
- External validation cohort:
  - Consecutive cohort of 20,000 unrelated probands undergoing MGPT through Ambry Genetics from 2015-16
  - Excluded those with missing age/family history data → 14,845 individuals

Model Development Cohort: 7296 consecutively ascertained, unrelated individuals (and multi-generational pedigree) undergoing clinical MGPT through DFCI Cancer Genetics Clinic

↓

Evaluated various combinations of cancer susceptibility genes based on clinical relevance, coverage within Development Cohort, # mutation carriers

↓

For each gene set (Table 1), we developed a logistic regression model to predict mutation in any of the genes in the set based on demographic and personal/family cancer history data

↓

LASSO (L1-regularization) was used to determine the weight of each clinical predictor variable → Weights of less important predictors automatically become 0

↓

ROC-AUC (cross-validation estimate), # of clinical predictors, and clinical judgement were used to select the final model for each set of genes

↓

Integrated individual models → Assessed performance at 90% sensitivity with ROC-AUC, specificity, positive-/negative-predictive value (PPV and NPV), and # needed to test to find a mutation carrier (NNT)

↓

Validated performance of the integrated models using **External Validation Cohort** of 14,845 individuals undergoing MGPT through a commercial laboratory (Table 2)

Take Home Points:

- PREMMplus is the first validated risk assessment model for multi-syndromic hereditary cancer risk assessment
- Can be used to identify individuals who should undergo MGPT
- Risk predictions based on personal and family history of **18 cancer types**, plus personal history of adenomas
- Ability to continually expand model and include additional genes
  - Trade-off between inclusion of moderate-penetrance genes and reduced discriminatory capacity

Results:

- Genes grouped as category I (Lynch syndrome or high-penetrance) or category II (moderate-penetrance)
- Development and validation cohorts well-balanced for demographic features and clinical histories

Table 1: Pathogenic Variants and Gene Sets Used for PREMMplus Development

Category	Genes	# Carriers: Development	# Carriers: Validation
I	*MLH1, MSH2, MSH6, PMS2, EPCAM, APC,* biallelic *MUTYH*	106	274
I	*BRCA1, BRCA2, CDH1*	272	408
I	*TP53*	25	23
II	*RAD51C, RAD51D*	22	51
II	*CHEK2, PTEN*	140	287
II	*ATM*	85	190
II	*CDKN2A, PALB2*	56	129
Any I	All 11 category I genes	400	705
Any I and II	All above 18 genes	706	1362

- PREMMplus generates risk predictions based off personal/family history of 18 different cancers plus colorectal adenoma burden

Table 2: Performance of the PREMMplus Model in Development and Validation Cohorts

Cohort	Outcome	Sensitivity	Specificity	PPV	NPV	NNT	ROC-AUC (95% CI)
Development	11 genes	90.0%	34.8%	7.4%	98.4%	13.5	0.74 (0.71-0.77)
	18 genes	90.0%	23.9%	10.6%	96.0%	9.4	0.67 (0.65-0.69)
Validation	11 genes	90.0%	22.8%	5.5%	97.9%	18.3	0.69 (0.66-0.71)
	18 genes	90.0%	17.8%	9.8%	94.8%	10.2	0.62 (0.60-0.64)

- PREMMplus was well-calibrated
- Excellent performance for identifying high-penetrance/Lynch carriers; acceptable performance with addition of moderate-penetrance genes
- Comparable performance in external dataset

Future Directions:

- Expansion of PREMMplus to include additional genes (e.g., *BRIP1*)
  - Trade-off between inclusion of moderate-penetrance genes and reduced discriminatory capacity
- Additional validations in other clinical datasets
- Evaluate ability to integrate model in routine clinical care