Comparison of Variant Classification Algorithms Incorporating Clinical and Family History for Breast and Ovarian Cancer

Jacob Clifford, PhD; Mary Helen Black, PhD; Aaron Elliott, PhD; Hsiao-Mei Lu, PhD; Shuwei Li, PhD
Ambry Genetics, Aliso Viejo, CA

Background

- Pathogenic mutations are more likely to occur in high-risk individuals while benign variants are unrelated to personal and family history.
- Two approaches have been developed to assess variant pathogenicity based on personal and family history: logistic regression model (LRM) and history weighting algorithm (HWA).
- To assess differences in performance and clinical utility, we compared the two methods for classification of variants in genes associated with breast and/or ovarian cancer.
- Trained models can be used to predict pathogenicity or provide evidence for variant assessment, especially for reclassifying variants of uncertain significance (VUS).

Methods

1. Data Collection
   - We collected genetic sequence data, and personal (phx) and family history (fhx) for 94,562 Caucasian patients referred for genetic testing between 2014 and 2016.
   - For this analysis, we included 80,750 probands carrying only one reportable (pathogenic or VUS) variant and 59,831 probands with no reportable variants.

2. Statistical Approaches
   - LRM:
     \[
     \log \frac{P(G=1|X)}{1-P(G=1|X)} = \beta_0 + \beta_1 X
     \]
   - HWA:
     \[
     P(G=1) = \frac{(n_G + n_{\text{error}})}{(n_G + n_{\text{error}} + n_{\text{variant}} + 1 - P_{\text{error}}(X))}
     \]
   - Where:
     - \(n_G\): # non-mutant carriers satisfying X
     - \(n_{\text{error}}\): # non-mutant carriers satisfying X
     - \(n_{\text{variant}}\): # variant carriers satisfying X
     - \(P_{\text{error}}(X)\): variant prevalence

3. Comparing performance of LRM and HWA
   - Explanatory variables were breast cancer diagnosis age, ovarian cancer diagnosis age, male breast cancer, weighted counts of affected family members, and a severity-weighted fhx score.
   - Applied 2-fold and 10-fold stratified cross-validation for each scenario.

4. Variant Assessment
   - Proband Odds Ratio (OR) of carrying pathogenic mutation:
     \[
     OR = \frac{P(G=1|X)}{1-P(G=1|X)} \frac{1-P_{\text{error}}}{P_{\text{error}}}
     \]
   - LRM:
     \[
     \log \frac{P(G=1|X)}{1-P(G=1|X)} = \beta_0 + \beta_1 X
     \]
   - HWA:
     \[
     P(G=1) = \frac{(n_G + n_{\text{error}})}{(n_G + n_{\text{error}} + n_{\text{variant}} + 1 - P_{\text{error}}(X))}
     \]

Results

- For models including phx variables only, both HWA and LRM performed similarly (AUC range: 0.57-0.78 and 0.58-0.79; data not shown).
- For models including phx and severity-weighted fhx score (5-variable model), LRM performed HWA for some genes (Table 1).
- For models including phx and breast and ovarian cancer-specific fhx (7-variable model), LRM performed HWA for most genes (Table 1).
- High penetrance genes consistently performed better than low penetrance genes regardless of model.
- 158 VUSs were to be benign/VLB based on variant-specific thresholds (Fig 1) given the 7-variable LRM model. As shown in Table 2.15 of the VUSs also meet or exceed the \(\eta_{\text{min}}\) required (Fig 2).

Table 1. Model Performance for 5- and 7-Variable Scenarios with 2-Fold Cross-Validation

<table>
<thead>
<tr>
<th>Gene</th>
<th>HWA AUC</th>
<th>LRM AUC</th>
<th>P-value</th>
<th>HWA AUC</th>
<th>LRM AUC</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1</td>
<td>0.80</td>
<td>0.81</td>
<td>0.33</td>
<td>0.72</td>
<td>0.84</td>
<td>9.8 x 10^11</td>
</tr>
<tr>
<td>BRCA2</td>
<td>0.72</td>
<td>0.77</td>
<td>7.0 x 10^-6</td>
<td>0.65</td>
<td>0.77</td>
<td>8.2 x 10^-10</td>
</tr>
<tr>
<td>CDH1</td>
<td>0.65</td>
<td>0.67</td>
<td>0.70</td>
<td>0.69</td>
<td>0.86</td>
<td>9.8 x 10^-10</td>
</tr>
<tr>
<td>TP53</td>
<td>0.64</td>
<td>0.73</td>
<td>0.03</td>
<td>0.69</td>
<td>0.77</td>
<td>0.04</td>
</tr>
<tr>
<td>PALB2</td>
<td>0.69</td>
<td>0.70</td>
<td>0.55</td>
<td>0.54</td>
<td>0.66</td>
<td>8.4 x 10^-4</td>
</tr>
<tr>
<td>PTEN</td>
<td>0.75</td>
<td>0.72</td>
<td>0.45</td>
<td>0.75</td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>ATM</td>
<td>0.62</td>
<td>0.65</td>
<td>0.37</td>
<td>0.68</td>
<td>0.61</td>
<td>0.17</td>
</tr>
<tr>
<td>NBN</td>
<td>0.62</td>
<td>0.69</td>
<td>0.69</td>
<td>0.64</td>
<td>0.60</td>
<td>0.16</td>
</tr>
<tr>
<td>CHEK2</td>
<td>0.57</td>
<td>0.66</td>
<td>3.0 x 10^-7</td>
<td>0.70</td>
<td>0.66</td>
<td>0.52</td>
</tr>
</tbody>
</table>

- \*Variant model includes breast cancer diagnosis age, ovarian cancer diagnosis age, male breast cancer, multiple primaries of breast cancer, and severity-weighted fhx score.
- \#Variable model includes breast cancer diagnosis age, ovarian cancer diagnosis age, male breast cancer, multiple primaries of breast cancer, and 39 phx variables, weighted number of relatives with breast cancer diagnosis, breast cancer diagnosis, and affected with ovarian cancer.
- \%Patients were treated with breast cancer chemotherapy.

Conclusions

- These data demonstrate the performance of each model under a variety of data availability scenarios; LRM performed HWA for most scenarios across most genes.
- We combined the strength of LRM and variant-specific thresholds to reclassify VUSs.
- This may have important implications for variant assessment.

References