Gene+environment risk models: whys and hows

Peter Kraft
Professor of Epidemiology and Biostatistics
Harvard T.H. Chan School of Public Health
• Genes are important
• The environment is important
• Risk models that include both can be useful
Risk models using E

- Framingham risk model (CVD)
- Breast cancer risk assessment tool
- Chronic and acute risk of VTE
Risk models using E

- Framingham risk model (CVD)
- Breast cancer risk assessment tool
- Chronic and acute risk of VTE
- Prostate cancer
Does adding G help?
$e$, $g$, $g + e$
Training Models Including G & E

What functional form should the model have? What variables should be included? How should they be coded?

Data ("A")
In

Model ("Y_A(X)")
Out
Training Models Including G & E

What functional form should the model have? What variables should be included? How should they be coded?

Linear or logistic regression assuming additive effects within and across genetic and environmental factors
What functional form should the model have? What variables should be included? How should they be coded?

\[ E[g(Y)] = a_1 E_1 + \ldots + a_k E_k + b_1 G_1 + \ldots + b_k G_L \]
Training Models Including G & E

What about non-linear effects (aka interactions)? School of hard knocks: linear terms pick up most of the signal, potential gains from including non-linear terms swamped by degrees of freedom.
What about non-linear effects (aka interactions)?
School of hard knocks: linear terms pick up most of the signal, potential gains from including non-linear terms swamped by degrees of freedom.

This model choice has implications for extrapolated risks in the tails.
The assumed log-additive relative risk model has notable implications for individuals in the tails of the genetic risk distribution, more work is needed to verify that this model is a good fit \textit{in the tails}.
Polygenic models used in disease risk prediction have genetic risk scores constructed using odds ratios from initial discovery publications. However, the statistically significant deviations from additivity may leave some scope for improvement. This deviation may reappear due to the fact that the external odds ratios used to derive the multiplicative risk score were overestimated due to the curse of dimensionality. In sensitivity analyses where we adjusted all coefficients by a constant and this model could be utilized to provide some support for the use of a multiplicative model for the combination of 19 SNPs suggests the presence of ad-

Tests for pairwise interactions between the 23 SNPs failed to show statistically significant departures from multiplicativity in the tails of the risk distribution are projections beyond the a priori evidence for the same. Our results for marginal effects of most of the SNPs were not substantially different from results of previous studies that assumed log-additive SNP effects. More-
Testing calibration of risk models at extremes of disease risk

MINSUN SONG
Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20850, USA

PETER KRAFT, AMIT D. JOSHI
Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

MYRTO BARRDAHL
Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

NILANJAN CHATTERJEE*
Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20850, USA
chattern@mail.nih.gov

SUMMARY
Risk-prediction models need careful calibration to ensure they produce unbiased estimates of risk for subjects in the underlying population given their risk-factor profiles. As subjects with extreme high or low risk may be the most affected by knowledge of their risk estimates, checking the adequacy of risk models at the extremes of disease risk distribution is very important for clinical applications. We propose a new approach to test model calibration targeted toward extremes of disease risk distribution where standard goodness-of-fit tests may lack power due to sparseness of data. We construct a test statistic based on model residuals summed over only those individuals who pass high and/or low risk thresholds and then maximize the test statistic over different risk thresholds. We derive an asymptotic distribution for the max-test statistic based on analytic derivation of the variance–covariance function of the underlying Gaussian process. The method is applied to a large case–control study of breast cancer to examine joint effects of common single nucleotide polymorphisms (SNPs) discovered through recent genome-wide association studies. The analysis clearly indicates a non-additive effect of the SNPs on the scale of absolute risk, but an excellent fit for the linear-logistic model even at the extremes of risks.

Keywords: Case–controls studies; Gene–gene and gene–environment interactions; Genome-wide association studies; Goodness-of-fit tests; Polygenic score; Risk stratification.
Training Models Including G & E

What about non-linear effects (aka interactions)? School of hard knocks: linear terms pick up most of the signal, potential gains from including non-linear terms swamped by degrees of freedom.

Constraining GxE effects may improve model fit (assuming SNP-E interactions are mostly in the same direction). Increasing exposure variance may as well.
Recent advances in our understanding of common genetic markers associated with a broad range of human traits have provided a wealth of new data for genetic association studies. Hypothetical studies of exposure captured by two genetic factors, such as single nucleotide polymorphisms (SNPs) or a multi-SNP genetic risk score, have shown promise in identifying genetic factors that influence outcomes. However, the practical application of such genetic risk scores is limited by the variability in sampled exposures and genetic susceptibility.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.

The discussion of exposure misclassification in Stenzel et al. raises philosophical and increasingly important considerations. The relative lack of accuracy of current exposure assessment methods is inevitable. But on a practical level, many of the exposures we can measure and on which we could intervene are too expensive to measure directly in epidemiologic studies. Instead, epidemiologists incorporate inexpensive proxies—a practice which is only likely to increase, as epidemiologists incorporate different streams of Big Data into their studies. The results of Stenzel et al. suggest that the utility of designs that sample from a larger cohort based on an exposure subset of subjects whose exposures will be measured using a more expensive ''gold standard'' technology.
What if G is Mediated thru E?
What if G is Mediated thru E?

If all of G’s effect is mediated thru E and we’ve measured E without error, then genetic terms $b_1,\ldots,b_K$ go to 0.
What if G is Mediated thru E?

The truth is more like this.
CVD

Breast Cancer

Note that at finite sample size, the optimal NRI appears encouraging, a large number of women would in fact be misclassified under either score. The continuous NRI values of the prevalence and proportion of null SNPs. Similar to the development of cancer is 92% for the environmental score alone, but is 86% for the combined score, whereas the sensitivity for alternative values of the prevalence and proportion of null SNPs. These results are compatible with those of previous studies by screening the half of the population with highest scores. Sizes nearly half of cases could be detected by screening the other sensitivity analyses in the supplementary tables, these

Table 7 shows the proportion of cases within the top 10%,

Reflecting applications in screening, the breast cancer literature by Mavaddat et al., with 0.99 controls per case as in the Breast Cancer Association Consortium.

For a proportion of the population

\[ q = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \]

Here the example of height is used to illustrate a relation

\[ P(X | Y) = \frac{P(Y | X) P(X)}{P(Y)} \]

with 2.05 controls per case as in the CARDIoGRAMplusC4D

![CVD and Breast Cancer graphs](image-url)

Dudbridge (2018) *Genet Epidemiol*
So, is it useful?

- Spreading risk distribution
- Identifying subgroups where G is actionable
Can increase the gradient of predicted risks by including $G$, but utility will depend on context—e.g. on the “action threshold” where expected benefits of intervention outweigh risks.
Adding G can identify 2x as many folks with RR > 2
Adding G can identify >10x as many folks with RR>2
Figure 1. Cutaneous squamous cell carcinoma risk with increasing Polygenic Risk Score.
Women < 40 years old

Preliminary results from CARRIERS consortium, presented at AACR 2019
Gao, Couch, Goldgar, Nathanson, Couch, Kraft et al.
Determining clinical utility

• RCTs

• In the absence of large, expensive, and time consuming RCTs, we can simulate effectiveness using a model of disease natural history w/ or w/o intervention
Determining clinical utility

Common Inputs:
- Risk Factors
- Screening Behavior
- Diffusion of New Treatments

Simulation or Analytical Model:
- Cancer Models

Common Outputs: Costs and Benefits of Interventions:
- Examples of outputs:
  - Mortality
  - Quality Adjusted Life Years
  - Over Diagnosis
  - Direct Medical Costs

https://cisnet.cancer.gov/modeling/
Determining clinical utility

Example: stratified breast cancer screening

(preliminary work from CISNET breast cancer working group
—van den Broek et al.)
Bi-annual screening starting at age 50
Life years gained: 118
Breast cancer deaths averted: 6.7

Life years gained: 196
Breast cancer deaths averted: 9.7
More can be better

- Qx risk factors
- Mammographic density
- PRS
- Emerging biomarkers
Misc Issues

• Implementing “complicated” E models — good locally, maybe not globally
• Including biology in risk models
The risk I took was calculated, but man, am I bad at math.