

# Evaluation of Risk Scores Derived from the Health Family Tree Program

Yuling Jiang, BS, MS,<sup>1</sup> Catherine J. Staes, BSN, MPH, PhD,<sup>1</sup>

Ted D. Adams, MPH, PhD,<sup>2</sup> Steven C. Hunt, PhD,<sup>2</sup>

<sup>1</sup>Department of Biomedical Informatics, and <sup>2</sup>Division of Cardiovascular Genetics,  
University of Utah, Salt Lake City, UT

## Abstract

*Family health history is an independent risk factor for certain diseases. The Health Family Tree (HFT) was developed and used to document and assess family health history from the families of high school students since 1980. While the risk algorithm of the HFT was initially validated, 20 years of use as a public health tool in the community provides an extremely large dataset for more rigorous validation. A retrospective cohort study was used with the events before the “cut-off” year as the baseline and the events after the “cut-off” year as the follow-up. Baseline data were used in the algorithm to calculate the Family History Score (FHS). Cox proportional hazards model was used to test the dose-response nature of the FHS for predicting incident events. An  $FHS \geq 1$  was determined to be a significant predictor for future development of diabetes, myocardial infarction, and early onset of myocardial infarction.*

## Introduction

A family health history is the description of the genetic relationships and medical history of a family.<sup>1</sup> It reflects shared genetic susceptibility, shared environmental factors, and common behaviors among family members.<sup>2</sup> With increasing recognition of the role of genetic factors in human disease, the family health history can become a tool for improving clinical care and targeting prevention strategies for populations at greater risk of disease.<sup>3</sup> A positive family health history for cancer, diabetes mellitus or heart disease is considered a risk factor for these diseases. In addition, family history has been shown to be an independent risk factor for coronary heart disease, even after adjusting for other known risk factors such as smoking, obesity, cholesterol, blood pressure, and diabetes.

## Background

To document and assess family health history, the Health Family Tree (HFT) program was developed by researchers in Utah and Texas in 1980.<sup>4</sup> The HFT was used to collect health history information from the families of high school students, to apply a family risk score after analyzing the results, and to provide a written report to the students' families about their risk of common diseases and risk factors. In 1986,

Hunt et al. compared different definitions of a positive family history for coronary heart disease (CHD) and hypertension using data from 15,250 Utah families.<sup>4</sup> Definitions based on the number and age of affected first degree relatives were compared with definitions that used a quantitative family history score (FHS) based on observed and expected occurrence of disease. It was concluded that the FHS predicted future disease incidence in unaffected family members better than other definitions.<sup>4</sup> Based on these findings, selected cut-points in the continuous family risk score were defined for classifying families as high risk or not.<sup>4</sup> Since the family health history data was self-reported, a validation study was performed in 1986 to assess the data for early-onset coronary heart disease and hypertension.<sup>4</sup> The researchers found that the risk scores performed well, with a sensitivity (67%), specificity 96%, positive predictive value (79%), and negative predictive value (91%) for coronary heart disease.<sup>4</sup>

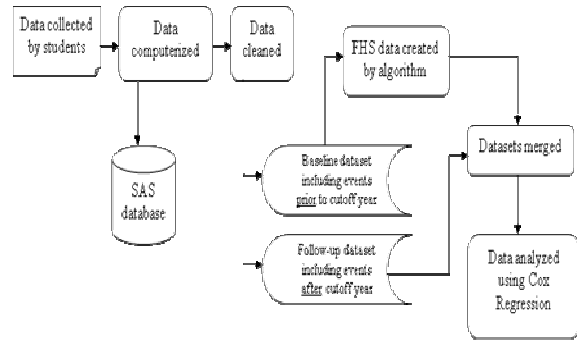
In addition to CHD and hypertension, there are many other health conditions included in the HFT questionnaire. No validation has been done to determine cut-points in the risk scores defining high risk for these diseases. There is a need to determine appropriate cut-points for other common diseases included in the HFT. During 1983 to 2002, the HFT questionnaire was used to collect data for over one million relatives. After 20 years of use as an educational and public health tool in the community, there is an opportunity and need to further evaluate and validate the risk algorithms to meet public health goals. A web-based version of the HFT tool is nearing completion for further use in Utah schools. For this analysis, the aim is to

- analyze data collected in the schools using the HFT between 1983 and 2002,
- evaluate risk scores, and
- determine appropriate cut-points for diabetes mellitus and myocardial infarction (MI).

## Methods

A retrospective cohort design was used for this study. Two health conditions included on the data collection form were analyzed: diabetes and MI. Two MI outcomes were tested: all occurrences of MI and early onset of MI. Early onset was defined as the

occurrence of MI at  $\leq 55$  years of age for males and  $\leq 65$  years of age for females. Each family was classified according to their family history of diabetes or MI as of a cut-off year. For each family, only the events that occurred up to the cut-off year were used to define the family history of each disease. Families were then divided by family history category and the unaffected relatives in each family were followed for disease occurrence after the cut-off year until the actual date of data collection. The data analysis process is illustrated in Figure 1.



**Figure 1.** Overview of the data analysis process

As described in the previous publication,<sup>4</sup> the FHS is calculated using the following equation:

$$\text{If } |O - E| > \frac{1}{2} \text{ then,}$$

$$\text{FHS} = \frac{(O - E) - 1/2}{\sqrt{E}} \times \frac{O - E}{O - E}$$

Or, if  $|O - E| \leq \frac{1}{2}$  then  $\text{FHS} = 0$ ,

There are two main variables in the equation. The observed incidence of disease (O) was the observed number of events in the family; the expected number of events (E) was calculated by multiplying the age- and sex-specific person-years for each person in the family by the age- and sex- specific incidence rates of the general population. E was calculated by using the entire HFT database of over one million relatives, which is representative of the northern Utah population. The population incidence rate used to calculate the expected number of events in the family is the number of new cases of disease divided by the person-years. The entire database was used for these incidence rates (1983 to 2002), producing an average rate over the period and not modeling secular trends.

To test the dose-response nature of the FHS, the family histories (as of cut-off year) were divided into five groups to calculate follow-up incidence rates:  $0.5 \leq \text{FHS} < 1$ ,  $1 \leq \text{FHS} < 2$ ,  $\text{FHS} \geq 2$ ,  $\text{FHS} \geq 1$ , and  $\text{FHS} \geq 0.5$ . Differences in diabetes and MI rates between groups were analyzed using a Cox proportional hazards model so that possible confounding variables could be included in the model. Follow-up time was defined as the time since the cut-off year until the onset of the condition (incidence), death, or the year of data collection, whichever was earliest.

All analyses were performed using SAS software. Institutional Review Board approval was obtained for this analysis from the University of Utah.

## Results

### Study population

There were 1,195,599 individuals in the raw dataset. After data cleaning, 174,923 individuals (14.6%) were deleted, including 15,318 (1.3%) individuals who were born after the cut-off year and excluded from the family history score calculations. A total of 35,482 (2.9%) individuals were excluded from the analysis because their family size was too small ( $< 3$  family members over age 20) to calculate a family history score. Therefore, the population under study contains 969,876 individuals, including 481,190 (49.6%) males and 488,686 (50.4%) females. Table 1 and Table 2 show the mean values and standard errors of the study variables for the three FHS family history groups: high ( $\text{FHS} \geq 1.0$ ), medium ( $0.5 \leq \text{FHS} < 1.0$ ), and low ( $\text{FHS} < 0.5$ ) for diabetes and MI. All reported means were determined using information up to the cut-off year. Similar distributions are seen in diabetes and MI high or low risk groups. For both diabetes and MI, the number of families in the high risk group is about 10% of the number of families in the low risk group. But for MI, even though the high risk families averaged 2.8 years older than the low risk families, the affected members within the high risk group were affected at younger ages (4.9 years earlier) than those in the low risk group. The average FHS for diabetes high risk families is 4.01, while the average FHS for diabetes low risk families is 0.008. Similarly, the average FHS for MI high risk families is 2.99, while average FHS for MI low risk families is -0.009.

### Diabetes

Table 3 shows the hazard ratios of different FHS definition groups for diabetes, calculated using Cox Regression. Even for the weakest definition of family history, a positive family history of diabetes is predictive of future diabetes for unaffected family members, both in males and females. Since age could be a confounding factor, we grouped the individual's age at the cut-off year into five groups (20-39, 40-49, 50-59, 60-69, and 70-89), and used the age group as a covariate in the Cox regression model.

## Myocardial infarction

For MI, a Cox Regression analysis was performed on both all-age MI and early-onset of MI (male $\leq$ 55, female $\leq$  65). For males, positive family history of MI was predictive for MI at any age among unaffected family members, with the exception of the risk group defined by FHS $\geq$ 0.5. For females, all definitions of a positive family history of MI were predictive for MI at any age among unaffected family members (Table 4). Furthermore, family history of MI was a better predictor for early-onset MI than MI at any age. All definitions of positive family history for MI were predictive of future early onset MI for all unaffected family members (Table 5).

Tables 4 and 5 show the results of MI risk with the covariate of baseline age included in the models for MI at any age and early onset of MI (male $\leq$ 55, female $\leq$  65). Positive family history of MI was still predictive for MI at any age among unaffected male family members. All definitions of positive family history were predictive of MI at any age among female family members. All definitions of positive family history are predictive of future early onset MI for both male and female unaffected family members.

**Table 1.** Characteristics (mean $\pm$ SE) of the diabetes risk groups defined by family history

|                           | Diabetes Risk Group |                 |                    |
|---------------------------|---------------------|-----------------|--------------------|
|                           | High                | Medium          | Low                |
| # of families             | N=742               | N=653           | N=69,724           |
| Family* size              | 10.3 $\pm$ 3.5      | 6.7 $\pm$ 2.9   | 13.4 $\pm$ 4.8     |
| # of affected             | 2.25 $\pm$ 0.61     | 1.00 $\pm$ 0.05 | 0.02 $\pm$ 0.12    |
| Avg. age-affected         | 37.5 $\pm$ 25.9     | 37.9 $\pm$ 26.5 | 32.5 $\pm$ 25.5    |
| Avg. age-all              | 31.1 $\pm$ 19.9     | 36.4 $\pm$ 15.4 | 30.6 $\pm$ 18.2    |
| Avg. family history score | 4.01 $\pm$ 3.33     | 0.83 $\pm$ 0.12 | 0.008 $\pm$ 0.0718 |

\*One family includes the student, siblings, mother, maternal aunts and uncles, and grandparents. A second family includes the student, siblings, father, paternal aunts and uncles, and grandparents.

**Table 2.** Characteristics (mean $\pm$ SE) of the Myocardial Infarction risk groups defined by family history

|                           | Diabetes Risk Group |                 |                     |
|---------------------------|---------------------|-----------------|---------------------|
|                           | High                | Medium          | Low                 |
| # of families             | N=751               | N=1250          | N=69,118            |
| Family* size              | 9.5 $\pm$ 3.4       | 6.5 $\pm$ 3.0   | 13.2 $\pm$ 4.9      |
| # of affected             | 2.16 $\pm$ 0.46     | 1.01 $\pm$ 0.13 | 0.05 $\pm$ 0.23     |
| Avg. age-affected         | 45.2 $\pm$ 25.2     | 43.8 $\pm$ 23.1 | 50.1 $\pm$ 27.6     |
| Avg. age-all              | 33.3 $\pm$ 20.3     | 36.1 $\pm$ 15.4 | 30.5 $\pm$ 18.3     |
| Avg. family history score | 2.99 $\pm$ 2.23     | 0.81 $\pm$ 0.14 | -0.009 $\pm$ 0.1316 |

\*Same family definition as in table 1.

## Discussion

An FHS $\geq$ 1 was a significant predictor for future development of diabetes, MI, and early onset of MI. With age as a covariate, even the weakest definition, FHS $\geq$  0.5, predicted unaffected family member's risk of developing diabetes or MI (at any age or early-onset). Without age as covariate, an FHS  $\geq$  0.5 still predicted the individual's risk of developing diabetes or early-onset MI. These findings are consistent with the findings from the earlier study that only addressed coronary heart disease and hypertension and only used two years of data collected by the HFT program.<sup>4</sup> Cut-points rather than continuous risk scores were used, as cut-points provide categorization benefits for field operations. High risk groups can be better identified for screening by cut-points.

Data collected by the HFT program and the calculation of a family history score takes advantage of information about the family size, age of persons in the family, and the frequency of disease in the general population. This information is more comprehensive than the information often used to assess family history of common disease (e.g., the numbers of affected first degree relatives). Future analysis of the data could include behaviors such as smoking and exercise as covariates to the Cox proportional hazards model, to examine the effect of environmental factors. The quality of data pertaining to relatives' behaviors as reported by third parties should be further examined.

The study population was the family members of students enrolled in the health education classes from high schools throughout Utah. The religious background of the majority of the state's citizens encourages recording detailed family histories, and family pedigrees in Utah are also typically larger than in other states.<sup>5</sup> In general, Utah's population has lower levels of alcohol and tobacco use than the U.S. population at large. However, a previous study showed that data from Texas students participating in the HFT project had similar results compared to Utah.<sup>6</sup> Larger families help in assessing family history risk and the larger the family the more important it is to calculate an expected number of events in assessing a family history. In addition to the successful use of the FHS in a multi-ethnic Texas population, the NHLBI Family Heart Study successfully used the score to assess family history in its five field centers in Alabama, North Carolina, Massachusetts, Minnesota, and Utah.

The analysis for this project may have limitations. While great effort was made during the high school

teachers' instruction to insure the most accurate data possible, the data were still either self-reported or reported by a relative and not verified externally. While this may be potentially problematic and either over- or under- report health conditions, this is likely to be a common feature for any family history tool. Recall bias may be a potential problem. It is possible families with higher rates of recent disease may be more likely to recall past events among family members than families with lower rates of past disease. This could result in falsely elevating the significance because families with higher recent rates will have higher recall for similar events. However, a prior study did not find evidence for this bias in the HFT program.<sup>4</sup>

In summary, the method outlined herein used to objectively score family health history can be used to identify high risk families in which unaffected individuals are currently at increased risk of developing diabetes and MI in the future. Based on this validated algorithm, we believe the web-based family history tool, Health Family Tree, can be successfully used in schools and in the community to educate families about their risk of disease and what they can do to reduce that risk. Other diseases listed on the Health Family Tree should also be validated (as detailed in this report) to complete the validation of the HFT program as a method of assigning family risk in order to help families reduce their risk of disease.

**Acknowledgments**

This project was funded by the Utah Department of Health, Chronic Disease Genomics Program Community Mini-Grant Program. One author (CJS) was partially supported by CDC Center of Excellence in Public Health Informatics grant #8P01HK000030 and National Library of Medicine, training grant #LM007124.

**References**

1. Bennett R. The practical guide to the genetic family history. New York: Wiley-Liss 1999.
2. Yoon PW SM, Peterson-Oehlke KL, Gwinn M, Faucett A, Khoury MJ. . Can family history be used as a tool for public health and preventive medicine? *Genet Med*2002;4:304-10.
3. Trotter TL, Martin HM. Family history in pediatric primary care. *Pediatrics*2007;120 Suppl 2:S60-5.
4. Hunt SC, Williams RR, Barlow GK. A comparison of positive family history definitions for defining risk of future disease. *J Chronic Dis*1986;39(10):809-21.
5. Johnson J WJ, Coats K, Giles RT, Larsen L, Adams T, et al. Family High Risk Program 1983-1999 [unpublished report in the Chronic Disease Genomic Program, Utah Department of Health]. Salt Lake City (UT): Utah Department of Health; 2004.
6. Johnson J, Giles RT, Larsen L, et al. Utah's Family High Risk Program: bridging the gap between genomics and public health. *Prev Chronic Dis*2005;2(2):A24.

**Table 3.** Hazard ratios for diabetes, by gender and FHS category

|                                    | 0.5≤FHS<1   | 1≤FHS<2     | FHS≥2       | FHS≥1       | FHS≥0.5     |
|------------------------------------|-------------|-------------|-------------|-------------|-------------|
| % of individuals with positive FHS | 5.16        | 0.08        | 0.51        | 0.59        | 5.69        |
| Without Age as covariate           |             |             |             |             |             |
| Male                               |             |             |             |             |             |
| Hazard ratio                       | 1.82        | 3.00        | 3.95        | 3.82        | 2.01        |
| 95% CI                             | (1.69,1.95) | (1.93,4.65) | (3.39,4.60) | (3.31,4.41) | (1.89,2.15) |
| Female                             |             |             |             |             |             |
| Hazard ratio                       | 1.67        | 3.68        | 4.39        | 4.28        | 1.91        |
| 95% CI                             | (1.56,1.79) | (2.52,5.37) | (3.80,5.06) | (3.75,4.90) | (1.79,2.03) |
| With Age as a covariate            |             |             |             |             |             |
| Male                               |             |             |             |             |             |
| Hazard ratio                       | 2.11        | 2.69        | 4.64        | 4.31        | 2.32        |
| 95% CI                             | (1.96,2.27) | (1.73,4.17) | (3.98,5.40) | (3.73,4.98) | (2.18,2.48) |
| Female                             |             |             |             |             |             |
| Hazard ratio                       | 1.89        | 3.29        | 5.05        | 4.74        | 2.14        |
| 95% CI                             | (1.76,2.02) | (2.26,4.80) | (4.38,5.82) | (4.14,5.41) | (2.01,2.28) |

Note: All p values were less than 0.001.

**Table 4.** Hazard ratios for MI, by gender and FHS category

|                                    | 0.5≤FHS<1   | 1≤FHS<2       | FHS≥2       | FHS≥1       | FHS≥0.5      |
|------------------------------------|-------------|---------------|-------------|-------------|--------------|
| % of individuals with positive FHS | 6.83        | 0.20          | 0.50        | 0.70        | 7.43         |
| Without Age as covariate           |             |               |             |             |              |
| Male                               |             |               |             |             |              |
| Hazard ratio                       | 0.85        | 3.19          | 2.13        | 2.43        | 1.00         |
| 95% CI                             | (0.79,0.91) | (2.57,3.96)   | (1.79,2.52) | (2.13,2.79) | (0.93,1.06)* |
| Female                             |             |               |             |             |              |
| Hazard ratio                       | 1.19        | 2.17          | 2.00        | 2.05        | 1.26         |
| 95% CI                             | (1.09,1.31) | (1.44,3.27)   | (1.53,2.62) | (1.64,2.57) | (1.16,1.37)  |
| With Age as a covariate            |             |               |             |             |              |
| Male                               |             |               |             |             |              |
| Hazard ratio                       | 1.54        | 3.30          | 3.29        | 3.29        | 1.75         |
| 95% CI                             | (1.43,1.66) | (2.66,4.09)   | (2.77,3.91) | (2.87,3.76) | (1.64,1.88)  |
| Female                             |             |               |             |             |              |
| Hazard ratio                       | 1.23        | 1.80          | 2.54        | 2.26        | 1.31         |
| 95% CI                             | (1.13,1.35) | (1.20,2.71)** | (1.94,3.32) | (1.81,2.83) | (1.20,1.43)  |

\* p value = 0.9056

\*\*p value = 0.0049

Note: p values were less than 0.001, if not mentioned.

**Table 5.** Hazard ratios for early onset (male ≤ 55, female ≤ 65) of MI, by sex and FHS category

|                                    | 0.5≤FHS<1   | 1≤FHS<2     | FHS≥2       | FHS≥1       | FHS≥0.5     |
|------------------------------------|-------------|-------------|-------------|-------------|-------------|
| % of individuals with positive FHS | 6.83        | 0.20        | 0.50        | 0.70        | 7.43        |
| Without Age as covariate           |             |             |             |             |             |
| Male                               |             |             |             |             |             |
| Hazard ratio                       | 2.28        | 5.63        | 5.67        | 5.67        | 2.59        |
| 95% CI                             | (2.09,2.50) | (4.14,7.67) | (4.64,6.93) | (4.78,6.71) | (2.39,2.81) |
| Female                             |             |             |             |             |             |
| Hazard ratio                       | 1.32        | 3.21        | 3.23        | 3.22        | 1.47        |
| 95% CI                             | (1.18,1.49) | (2.02,5.10) | (2.41,4.32) | (2.52,4.13) | (1.32,1.64) |
| With Age as a covariate            |             |             |             |             |             |
| Male                               |             |             |             |             |             |
| Hazard ratio                       | 2.19        | 5.78        | 5.55        | 5.61        | 2.51        |
| 95% CI                             | (2.00,2.40) | (4.24,7.86) | (4.54,6.79) | (4.74,6.65) | (2.31,2.72) |
| Female                             |             |             |             |             |             |
| Hazard ratio                       | 1.34        | 2.91        | 3.54        | 3.33        | 1.49        |
| 95% CI                             | (1.19,1.51) | (1.83,4.62) | (2.64,4.73) | (2.60,4.27) | (1.34,1.67) |

Note: All p values were less than 0.001.