Digital Family Histories for Data Mining

Save to myBoK

Abstract

As we move closer to ubiquitous electronic health records (EHRs), genetic, familial, and clinical information will need to be incorporated into EHRs as structured data that can be used for data mining and clinical decision support. While the Human Genome Project has produced new and exciting genomic data, the cost to sequence the human personal genome is high, and significant controversies regarding how to interpret genomic data exist. Many experts feel that the family history is a surrogate marker for genetic information and should be part of any paper-based or electronic health record. A digital family history is now part of the Meaningful Use Stage 2 menu objectives for EHR reimbursement, projected for 2014.

In this study, a secure online family history questionnaire was designed to collect data on a unique cohort of Vietnam-era repatriated male veterans and a comparison group in order to compare participant and family disease rates on common medical disorders with a genetic component. This article describes our approach to create the digital questionnaire and the results of analyzing family history data on 319 male participants.

Keywords: electronic health records, personal health records, Human Genome Project, family medical history

Introduction

The Human Genome Project accomplished whole-genome sequencing in 2003 and resulted in the availability of voluminous data and the development of new genetic tests. In spite of groundbreaking progress, the cost to perform whole-genome sequencing remains high, and the use of the results is controversial. According to the Human Genome Project, "the scientific community continues to debate the best way to deliver [results] to the public and medical communities that are often unaware of their scientific and social implications. While some of these tests have greatly improved and even saved lives, scientists remain unsure of how to interpret many of them."[1] Therefore, we are years away from incorporating genomic information into paper or electronic health records.

In the interim, many experts believe that information derived from a detailed family history will serve as a surrogate for personal genomic information.[2] Information derived from family histories is critical for focusing on medical disorders with a genetic component.[3,4] For example, national guidelines for the screening and management of cancer, diabetes, and cardiovascular disease often include family history information. Unfortunately, obtaining detailed family histories is time consuming, requires some expertise, is not associated with clinician or patient reimbursement, and depends on the quality of information provided by the patient.[5] As a result, many paper-based records, personal health records (PHRs), and electronic health records (EHRs) do not record vital family history that could be used to screen at-risk individuals and their families. Furthermore, when it is recorded, the information is rarely captured as structured data, amenable to computation.

According to a systematic review by Reid et al. (2009), 14 family history questionnaires (FHQs) have been published in the literature, but only 4 covered multiple medical conditions and none were web based. Four studies were validated against a formal pedigree interview, considered the gold standard of genetic study. Nevertheless, the authors concluded that "there are no simple, short generic FHQs suitable for use in primary care practices."[6] New approaches to capture computable family history information must be developed.

Capturing family history information is important because studies have shown that most people are at moderate to strong risk of a medical condition with a genetic component.[7] However, most current medical records lack this information. In spite of the increased adoption of EHRs and PHRs, most commercial applications thus far do not include the ability to capture computable family history information. Data standards have been developed to help represent this important information, but significant limitations exist.[8]

The Health Information Technology for Economic and Clinical Health (HITECH) Act established reimbursement by the Centers for Medicare and Medicaid for the meaningful use of certified EHRs in 2009. Stage 2 meaningful use, which is scheduled to go into effect in 2014, includes a menu objective of having family history recorded as structured data.[9]

The federal government is aware of this void in family history information and has devised several projects to facilitate data collection. My Family Health Portrait is a free, web-based, open-source application that participants and their families can use to create an electronic family history. This tool was developed by the Surgeon General and agencies within the US Department of Health and Human Services. Data standards include the HL7 Family History Model, Logical Observation Identifiers Names and Codes (LOINC), and Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT). Participants can create a family history in about 15 minutes and then download the XML file to their computer or share it with their primary care physician. However, no interpretive features are available to evaluate the results for participants. Because this tool was written using data standards, it was hoped that the results could be incorporated into EHRs and PHRs. The software application is available as a download to individuals, healthcare organizations, and developers.[10] As an example, data from My Family Health Portrait can be exported to RiskApps, a free application to identify and manage women at risk for hereditary breast and ovarian cancer.[11] Another federal program, Family Healthware, is a web-based research tool developed by the Centers for Disease Control and Prevention (CDC) that can be used to assess a person's risk for six disease categories (coronary heart disease, stroke, diabetes, colorectal, breast cancer, and ovarian cancer). A personalized prevention message is generated from collected data.[12]

In addition to providing valuable clinical decision support to clinicians, data from FHQs can also be a valuable resource for research. Yarnell et al. were able to demonstrate that a family history of coronary heart disease and parental longevity are related, but independent of each other and common cardiovascular risk factors, in predicting future coronary events.[13] It is likely that research will eventually "triangulate" information derived from clinical, genetic, and family history data.

In this study, a web-based FHQ was created as a means to quickly populate our research database and allow for data mining. This article discusses the creation of a secure online survey and presents examples of how family history data might be used for clinical research and population health improvement. Given the relatively small number of our survey results, the research approach was to look at common medical diseases with a genetic component. Specifically, our study explored this question: Are nondiabetic participants with a positive family history of diabetes different from nondiabetic participants with no family history of diabetes? Therefore, a positive or negative family history of diabetes was used as an independent variable.

Methodology

Participants

Our study population comprised 430 male repatriated prisoners of war (RPWs) from the Vietnam War era, as well as 118 comparison group participants, matched for gender, age, education, and combat roles in Vietnam. Both groupsvisit the Robert E. Mitchell Center for Prisoner of War Studies, located in Pensacola, FL, on a near-annual basis. The program has existed since 1973, with some RPWs having 40 years of longitudinal physical and psychological data.[14] More than 90 percent of participants are white, were previously aviators, and have four or more years of college education and therefore differ significantly from other veteran populations. In 2012 the average age of the RPWs and the comparison group was 72 years (standard deviation 6.4 years). In spite of severe malnutrition, torture, and solitary confinement, 57 percent of RPWs did not develop any evidence of psychiatric disease during the 40 years of follow-up, while 43 percent developed psychiatric disorders, including posttraumatic stress disorder (PTSD). This study was Institutional Review Board–approved, and all participants signed a consent form.

Survey Development

For content and face validity, an expert panel consisting of a PhD university-based geneticist (H.C.), a private genetic counselor (C.R.), a neuropsychologist (see acknowledgement section), and an experienced internal medicine physician (R.H.) was convened to decide on the appropriate survey design and selection of common medical and psychiatric diseases with a genetic component. The expertise of an information technology privacy and security consultant (B.H.) and a PhD-level researcher (S.L.) was solicited for statistical analysis. A literature review was also undertaken to determine the availability and

relevance of existing FHQs. The FHQ was benchmarked with the recommendations made by the 2008 American Health Information Community's Family Health History Multi-Stakeholder Workgroup.[15]

A commercial survey instrument (SurveyMonkey) was utilized to create the web-based survey.[16] The survey has the following sections:

1. Demographic-type questions including gender, adoption status, twin status, and ethnicity, to be answered by all participants prior to proceeding.
2. Personal health information divided into the following question categories. All categories have a free-text answer option. The number of questions in each category is shown in parentheses. In this section only, participants include the age at diagnosis in a drop-down menu.

    - General condition (8)
    - Heart condition (5)
    - Cancer (14)
    - Brain disease/neurodegenerative disease (6)
    - Mental disorder related to learning disability (2)
    - Mental disorder other than related to learning disability (8)
    - Substance abuse and smoking status (3)

3. Mother's, father's, children's, and siblings' health

    - Begins with these questions: living/deceased Y/N; current age or age of death; served in military Y/N.
    - The questions in the categories from section 2 (above) are again asked (total of 46 questions), but there is no option to record the age at diagnosis.

A decision was made to develop our own FHQ in order to answer specific research questions but have enough flexibility for other studies. Originally, the intent was to modify the open-source application My Family Health Portrait, but several challenges were identified. This program includes psychological disorders but does not include nicotine or alcohol use. Furthermore, the categories for age of death ended at "60 years and older," which would not have been adequate to evaluate parental longevity. Lastly, as the program is written, each individual downloads his or her own family health portrait to a personal computer, making group data aggregation difficult. Our approach to create a customized FHQ was associated with numerous lessons learned, outlined in the following paragraphs.

The initial survey was pilot tested with 20 RPW volunteers routinely seen in the center, who provided valuable input resulting in modifications of the survey, such as including former smoking status in the smoking history. During the design and testing of the familial history survey, changes were made to streamline the user experience and improve data analysis and survey data reliability.

To keep the FHQ concise and decrease the size of the resulting database, information was collected on the participant and first-degree relatives (parents, siblings, and biological children) only. Also, the most common definition of a positive family history is the involvement of one or more first-degree relatives.

Several important privacy factors needed to be taken into consideration in designing and delivering the participant notification, including but not limited to the method of delivery, the content of the message, the point of distribution, and—because the survey is hosted online—the anonymity of participants in regard to cookie permanence and Internet Protocol (IP) address logging at the point of data collection. A concerted effort was made to meet and exceed all of the expectations of research studies related to the HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule, published by the Department of Health and Human Services.[17] The method of delivery chosen was e-mail, a commonplace communication method and one adopted by a majority of the intended participants. The content of the message was an unassuming request for voluntary participation in an online survey. The message did not contain identifying information within the body of the message, and recipients' e-mail addresses were blind copied to reduce the likelihood of identification and the likelihood of unintended consequences in the improbable event that messages would be intercepted or sent in error. Included within the body of each message were two key elements for the survey: (1) the uniform resource locator (URL) used to gain access to the survey and (2) a unique identifier created for the sole purpose of this research. Lastly, the service used for hosting the web-based survey

permitted anonymous data acquisition, allowing evaluators the ability to disable IP and e-mail address logging for survey participants.

The resulting data set is a de-identified collection of familial history results that is tied only to the unique identifier given to participants as part of the notification. A macro was written in the Visual Basic language native to Microsoft Office products. When activated, the macro was designed to launch the local e-mail client on a given host, populate a message with key elements such as e-mail address and the unique identifier code, and send the message to the recipient. The content of this message in no way belied the purpose of the survey or the intention to collect familial medical history data. In spite of this, the chosen technique was seen as a preferable method to the one employed by the survey service, as having both survey results and originating data could constitute a method of re-identification by a third party.

Statistical Methods

Descriptive statistics were used to describe the participant population and to compare the prevalence of common diseases with a genetic component in our participants and their families with national prevalence rates. Prevalence was computed as the percentage of family members in each category (proband, parents, and siblings) with a disease. Because the prevalence of disease was very low in the children, they were not included in our data analyses.

Prior to further analysis, participants with known type 1 or 2 diabetes and any participants with a glycated hemoglobin (A1c) level of 6.5 or above, the cutoff point for diagnosing diabetes by some authorities, were excluded.[18] Diabetics were excluded because the role of family history in type 2 diabetes is well defined in the medical literature and the association with multiple biomarkers is also well defined and strong. Our area of focus was the role of a positive or negative family history of diabetes in the nondiabetic participants, an area less discussed in the medical literature. After the diabetic participants were excluded, 280 participants were left for evaluation, with 73 having a positive family history for diabetes in one or more family members. The metabolic biomarkers used represented the average of the last three tests, collected over the past three to five years. However, some data were missing because of changes in lab methods or tests not performed each year. This resulted in different numbers of data points in the tables displayed. To study the significance of a positive or negative family history of diabetes, two analyses were conducted. First, Mann-Whitney *U* tests were performed to compare the means of the various glycemic measures in participants with a positive family history of diabetes and those with a negative family history. Second, the presence or absence of a family history of diabetes was studied to examine whether family history had an impact on the correlations among several common metabolic markers. The data were assumed to have a nonnormal distribution; therefore, a Spearman correlation was performed. To determine whether the correlations between the two groups were significantly different, a Fisher *r*-to-*z* transformation of the correlations was performed. This is a procedure for computing a confidence interval on the difference between two independent correlations. Calculations were performed using IBM SPSS, version 19.

Results

Four hundred forty-seven participants were contacted by e-mail for the survey. Ten participants preferred to take the survey using paper forms. Three hundred nineteen participants successfully completed the survey (paper and electronic) for a survey completion rate of 72 percent. The survey was open for approximately one year to allow for data collection and follow-up reminders. The prevalence data in Table 1 were based on all 319 respondents and their first-degree relatives. For Table 2 and Table 3, we identified 38 participants whose records contained the ICD-9-CM code for type 2 diabetes, and they were excluded from data analysis. Therefore, data were analyzed for 280 nondiabetic participants, 207 of whom had a negative family history and 73 of whom had a positive family history for diabetes. Among the 73 participants without diabetes and with a positive family history for diabetes, 25 of 73 mothers (34 percent) and 29 of 73 fathers (40 percent) had diabetes. In 48 of 146 parents (33 percent), both parents were nondiabetic, and in 10 of 146 parents (7 percent), both parents were diabetic.

Table 1
Prevalence of Seven Medical Diseases with a Genetic Component, Compared to National Statistics

Condition	Participant	Mother	Father	Brother	Sister	National Statistics
Diabetes	11.6	12.5	9.7	8.9	8.2	26.9[a]

Myocardial infarction	8.1	10.6	32.6	11.7	3.9	18.3–24.7[b]
Stroke	2.1	19.1	10.6	1.8	2.0	6.5–10.5[c]
Hypertension	48.2	33.0	28.9	27.3	17.8	71.2[d]
Lung cancer	1.6	5.0	6.0	3.3	3.1	0.33–0.65[e]
Prostate cancer	12.8	NA	9.7	3.8	NA	4.8–10.8[f]
Colon or rectal cancer	0.63	4.1	3.4	2.1	1.1	0.85–1.7[g]

[a]Ages 65 and older, male and female, all ethnicities. American Diabetes Association. "Executive Summary: Standards of Medical Care in Diabetes—2010." *Diabetes Care* 33 (2010): S4–S10.
[b]Ages 65–74 and 75 and older, male and female, all ethnicities. Centers for Disease Control and Prevention. "Vital Signs: Prevalence, Treatment and Control of Hypertension—United States, 1999–2002 and 2005–2008." *Morbidity and Mortality Weekly Report (MMWR).* February 4, 2011. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6004a4.htm (accessed August 13, 2013).
[c]Ages 65–74 and 75 and older, male and female, all ethnicities. Centers for Disease Control and Prevention. "Vital Signs: Prevalence, Treatment and Control of Hypertension—United States, 1999–2002 and 2005–2008." *Morbidity and Mortality Weekly Report (MMWR).* February 4, 2011. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6004a4.htm (accessed August 13, 2013).
[d]Ages 65 and older, white, male and female, on treatment. National Cancer Institute. "Surveillance Epidemiology and End Results." Available at http://seer.cancer.gov/ (accessed January 10, 2013).
[e]Ages 60–69 and 70–79, male, white. National Diabetes Information Clearinghouse. "National Diabetes Statistics, 2011." Available at http://www.diabetes.niddk.nih.gov/dm/pubs/statistics (accessed January 3, 2013).
[f]Ages 60–69 and 70–79, male, white. National Diabetes Information Clearinghouse. "National Diabetes Statistics, 2011." Available at http://www.diabetes.niddk.nih.gov/dm/pubs/statistics (accessed January 3, 2013).
[g]Ages 60–69 and 70–79, male, white. National Diabetes Information Clearinghouse. "National Diabetes Statistics, 2011." Available at http://www.diabetes.niddk.nih.gov/dm/pubs/statistics (accessed January 3, 2013).

Table 1 includes common medical conditions with a genetic component recorded in the survey as listed by the participant and the participant's family and compared to national prevalence statistics.[19-22] In general, the prevalence of diabetes and cardiovascular disorders with a genetic component was lower in our cohort. The prevalence of cancer, particularly lung cancer, was higher in our cohort, but that finding might have been due to our small sample size. The national prevalence rates for noncancer conditions included both men and women, while the prevalence rates of cancer conditions were for only elderly white males.

Table 2 displays the statistically significant mean differences among metabolic markers in nondiabetic participants with a positive versus a negative family history of diabetes. Only BMI and insulin levels demonstrated a statistically significant difference (small effect size), with higher BMI and insulin levels noted in participants with a positive family history of diabetes.

Table 2
Mean Comparisons, Standard Deviations, and Effect Sizes of Metabolic Biomarkers for Nondiabetic Participants with Negative and Positive Family Histories for Diabetes (Mann-Whitney U Test)

	Negative Family History for Diabetes			**Positive Family History for Diabetes**		
	Mean	**SD**	***N***	**Mean**	**SD**	***N***
BMI[a]	27	3.5	207	28.2	4.09	73
TRG	122.2	75.2	207	119	57	73
Insulin[b]	7.3	5.2	192	8.5	4.8	71
HDL cholesterol	52	14.6	207	50	11.7	73

Hemoglobin A1c	5.5	0.29	174	5.5	0.38	68	
CRP	1.14	2.19	185	0.81	1.49	68	

Abbreviations: BMI: body mass index; TRG: triglycerides; HDL: high density lipoprotein; CRP: C-reactive protein.

[a]M BMI U(280) = 8,782, $z = 2.06$, $p = .04$, $r = .12$ (small effect size)
[b]M Insulin U(263) = 7,994, $z = 2.15$, $p = .03$, $r = .13$ (small effect size)

Table 3 shows the statistically significant associations of metabolic markers in nondiabetic participants with a positive or negative family history of diabetes. We found strong positive correlations among triglycerides and BMI, insulin and BMI, triglycerides and insulin, and triglycerides and C-reactive protein. We found strong negative correlations among HDL cholesterol and BMI, triglycerides and HDL cholesterol, and insulin and HDL cholesterol. Glycated hemoglobin (A1c) was not included in the table because it was not significantly correlated with the other biomarkers. A strong negative correlation between HDL cholesterol and C-reactive protein was noted, but only in the participants with a negative family history of diabetes. It is not known why this association was not seen in participants with a positive family history of diabetes.

Table 3
Spearman Correlations between Common Metabolic Biomarkers in Nondiabetic Participants with Negative or Positive Family Histories of Diabetes

Biomarkers	**Negative Family History of Diabetes**			**Positive Family History of Diabetes**				
Correlations	***r***	***p*** [a]	***N***	***r***	***p*** [b]	***N***	***z*** [b]	***p***
TRG–BMI	0.26	.000	207	0.37	.001	73	–0.89	0.37
Insulin–BMI	0.54	.000	191	0.53	.000	71	0.1	0.92
HDL–BMI	–0.32	.000	207	–0.24	.045	73	–0.67	0.50
TRG–insulin	0.43	.000	192	0.54	.000	71	–0.99	0.32
TRG–HDL	–0.48	.000	207	–0.43	.000	73	–0.51	0.61
TRG–CRP	0.18	.014	185	0.25	.038	68	–0.52	0.60
Insulin–HDL	–0.47	.000	192	–0.34	.004	71	-1.13	0.26
HDL–CRP	–0.20	.006	185	–0.13	—	68	–0.51	0.61

Abbreviations: BMI: body mass index; TRG: triglycerides; HDL: high density lipoprotein; CRP: C-reactive protein.

[a]Correlation is significant at $p < .05$ (two-tailed).
[b]Fisher *r*-to-*z* transformation.

Discussion

The creation of a secure online survey permitted a more rapid means of populating family history information into our database compared to waiting for participants to visit the research center on an annual basis. Importantly, using a well-recognized web-based survey instrument allowed for easier participation and data collection. Having family history data in a digital format resulted in easy importation into our database and into a statistical software package for analysis.

Digital family histories can be used for a variety of purposes as demonstrated by the examples provided in this article. Table 1 compared the prevalence of common medical conditions with a genetic component in our cohort with national prevalence rates for these conditions. The results revealed that, on average, our cohort of military veterans did not clearly match a typical

cross-section of the US population, most likely because they represented a higher socioeconomic status, as discussed above. For example, the prevalence of type 2 diabetes and hypertension in our participants and their families was much lower than the national average. We were particularly interested in whether nondiabetic participants with a positive family history of diabetes were different from nondiabetic participants with no family history of diabetes. In our cohort, participants with a family history of diabetes had higher insulin levels and BMIs than those with a negative family history. As a result of these preliminary data, several associations will be analyzed in more detail in future research. Our study did not show any differences in the anticipated correlations among common metabolic markers, with the exception that we found no statistically significant relationship between HDL cholesterol and C-reactive protein in participants with a positive family history of diabetes, for unclear reasons.

Digital family histories have the potential to enhance population health and become part of future clinical decision support in EHR systems. One study that looked at the effect of adding a systematic family history to a cardiovascular risk assessment showed a significant increase in high-risk patients identified. [23] Unique populations, like ours, can be analyzed to see if the proband or first-degree relatives have an increased or decreased prevalence of medical disorders compared with local, state, or national statistics. Integrating family histories into EHRs could alert clinicians if a patient is at increased risk of a medical condition with a genetic component. Algorithms embedded in the EHR might generate an alert to clinicians that the patient is at increased risk of a future condition based on a strong family history. In addition, new risk score calculators that combine family history data with biomarkers to calculate risk scores for multiple medical conditions are likely to be developed. An existing risk score calculator for type 2 diabetes that includes family history could be easily be part of clinical decision support in EHR systems.[24,25]

The transition to Meaningful Use Stage 2 measures projected for 2014 will change how clinicians and hospitals must address family histories. If the family history measure is adopted, clinicians and hospitals must record family histories for first-degree relatives as structured data for more than 20 percent of all unique patients seen. To accomplish this, appropriate vocabularies and standards such as SNOMED-CT and HL7 Version 3 Standard: Clinical Genomics; Pedigree must be used.[26] Patient portals that are integrated with EHRs will likely provide an option for patients to upload and edit their family histories.

For common diseases, self-reported family histories correlate reasonably well with review of actual medical records. According to a review by Yoon et al, the accuracy of reported family histories is high, with excellent sensitivities and specificities.[27] Family histories also correlate moderately well with formal pedigree studies.[28]

Several limitations of the family history should be pointed out, however. A systematic review of family history was sponsored by the Agency for Healthcare Research and Quality in 2009 and was the subject of a National Institutes of Health conference that year. The review evaluated five major research questions and drew the following conclusions: (1) family history definitions demonstrated suboptimal accuracy in predicting disease risk in individuals; (2) reports of relatives without a disease tend to be more accurate than reports of relatives with a disease; (3) it is not known if risk assessment based on family history will affect preventive behaviors, such as smoking cessation; (4) there is limited evidence to show that personalized risk assessment causes adverse outcomes; and (5) there is inadequate evidence on the optimal means to collect and report family history in primary care.[29]

Limitations of our study should also be mentioned. Our patient population consisted of primarily elderly white men with a high socioeconomic status; therefore, results may not be generalizable to other populations consisting of younger participants, both genders, multiple ethnicities, and other socioeconomic statuses. Furthermore, because our cohort comprised only male participants, whereas the national prevalence statistics for diabetes, hypertension, stroke, and heart attack included both genders, they are not entirely comparable.

More research is needed to determine how family history data can be optimally used and integrated with EHRs. The creation and adoption of a standardized FHQ to be used by primary care clinicians is essential because they are the most likely healthcare providers to inquire about family history. Existing standards for family history data need to be thoroughly tested and validated. Ultimately, computable family history data will need to be integrated with personal risk factors, laboratory tests, and genetic profiles in the EHR so that clinical decision support tools can be designed and tested. Lastly, clinicians will need to decide whether to input family histories at each visit or to use a FHQ such as the one we designed to upload results in batches.

Conclusion

Family history is an important part of any medical record and is a potentially valuable tool for disease prediction, prevention, and research. We are moving toward genetic information being part of all medical record systems, but obstacles remain, such as cost, incomplete data standards, and the fact that we have only begun to include family histories in EHRs. Family history information should be readily available in a computable format so that clinical decision support tools can remind clinicians of important testing and risk assessment needs. Unfortunately, no standard, simple generic FHQ is available for common use in primary care, with or without an EHR system.

A web-based FHQ was developed as part of this research study to help evaluate our unique cohort. Further work is needed to determine and validate the optimal family history core questions, the best methods to collect this information, how to integrate computable family history information into EHRs, interoperable data standards, and future clinical decision support tools.

Funding

Support was provided by the Office of Naval Research (ONR) under the Force Health Protection Future Naval Capabilities program. (N0001411AF00002).

Acknowledgments

We would like to thank Justice Mbizo, DrPH; Jeffrey Moore, PhD; Georgina Palombo, MBA; and Alison Fields, BA, for their contributions to this manuscript.

A copy of the FHQ is available upon request to the corresponding author, Dr. Hoyt, at robert.hoyt@med.navy.mil.

Contributor Information

Robert Hoyt, MD, FACP, is the Director of the Medical Informatics Program in the School of Allied Health and Life Sciences at the University of West Florida in Pensacola, FL. He is also a clinical researcher at the Robert E. Mitchell Center for Prisoner of War Studies in Pensacola, FL.

Steven Linnville, PhD, is a researcher at the Robert E. Mitchell Center for Prisoner of War Studies in Pensacola, FL

Hui-Min Chung, PhD, is a geneticist in the School of Allied Health and Life Sciences at the University of West Florida in Pensacola, FL.

Brent Hutfless, MS, is an information technology security manager at Austal USA, LLC, in Mobile, AL.

Courtney Rice, MS, CGC, is a genetic counselor in the TriHealth Cancer Institute in Cincinnati, OH.

Notes

[1] Human Genome Project. “Medicine and the New Genetics.” http://www.ornl.gov/sci/techresources/Human_Genome/medicine/medicine.shtml(accessed January 5, 2013).

[2] Kmiecik, T., and D. Sanders. *Integration of Genetic and Familial Data into Electronic Medical Records and Healthcare Processes*. February 2, 2009. Available at http://www.surgery.northwestern.edu/docs/KmiecikSandersArticle.pdf(accessed January 10, 2012).

[3] Rich, E. C., et al. “Reconsidering the Family History in Primary Care.” *Journal of General Internal Medicine* 19 (2004): 273–80.

[4] Guttmacher, A. E., et al. “The Family Health History—More Important Than Ever.” *New England Journal of Medicine* 351 (2004): 2333–36.

[5] Suther, S., and P. Goodson. “Barriers to the Provision of Genetic Services by Primary Care Physicians: A Systematic Review of the Literature.” *Genetic Medicine* 5 (2003): 70–76.

[6] Reid, G. T., et al. “Family History Questionnaires Designed for Clinical Use: A Systematic Review.” *Public Health Genetics* 12 (2009): 73–83.

[7] O’Neill, S. M., et al. “Familial Risk for Common Diseases in Primary Care: The Family Healthware Impact Trial.” *American Journal of Preventive Medicine* 36, no. 6 (2009): 506–14.

[8] Melton, G. B., et al. “Evaluation of Family History Information within Clinical Documents and Adequacy of HL7 Clinical Statement and Clinical Genomics Family History Models for its Representation: A Case Report.” *Journal of the American Medical Informatics Association* 17 (2010): 337–40.

[9] Centers for Medicare and Medicaid Services. “Stage 2.” Available at http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html (accessed January 20, 2013).

[10] “My Family Health Portrait.” Available at https://familyhistory.hhs.gov (accessed January 28, 2013).

[11] Hughes riskApps. Cancer Risk Assessment Software. Available at http://www.hughesriskapps.com/(accessed February 4, 2013).

[12] Centers for Disease Control and Prevention. Genomics Translation: Family Healthware ™. Available at http://www.cdc.gov/genomics/famhistory/famhx.htm (accessed January 16, 2013).

[13] Yarnell, J., et al. “Family History, Longevity, and Risk of Coronary Heart Disease: The PRIME Study.” *International Epidemiology Association* 32 (2003): 71–77.

[14] Robert E. Mitchell Center for Prisoner of War Studies. Available at http://www.med.navy.mil/sites/nmotc/rpow/Pages/default.aspx (accessed January 16, 2013).

[15] Feero, W. G., et al. “New Standards and Enhanced Utility for Family History Information in the Electronic Health Record: An Update from the American Health Information Community’s Family Health History Multi-Stakeholder Workgroup.” *Journal of the American Medical Informatics Association* 15 (2008): 723–28.

[16] SurveyMonkey. Available at http://www.surveymonkey.com (accessed January 4, 2013).

[17] National Institutes of Health. “Clinical Research and the HIPAA Privacy Rule.” Available at http://privacyruleandresearch.nih.gov/clin_research.asp (accessed January 5, 2013).

[18] American Diabetes Association. “Executive Summary: Standards of Medical Care in Diabetes—2010.” *Diabetes Care* 33 (2010): S4–S10.

[19] National Diabetes Information Clearinghouse. “National Diabetes Statistics, 2011.” Available at http://www.diabetes.niddk.nih.gov/dm/pubs/statistics (accessed January 3, 2013).

[20] National Cancer Institute. “Surveillance Epidemiology and End Results.” Available at http://seer.cancer.gov/ (accessed January 10, 2013).

[21] Centers for Disease Control and Prevention. “Vital Signs: Prevalence, Treatment and Control of Hypertension—United States, 1999–2002 and 2005–2008.” *Morbidity and Mortality Weekly Report (MMWR).* February 4, 2011. Available at http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6004a4.htm (accessed August 13, 2013).

[22] Centers for Disease Control and Prevention. “Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2011.” *Vital and Health Statistics* 10, no. 256 (December 2012). Available at http://www.cdc.gov/nchs/data/series/sr_10/sr10_256.pdf accessed August 13, 2013.

[23] Quereshi, N., et al. "Effect of Adding Systematic Family History Enquiry to Cardiovascular Disease Risk Assessment in Primary Care." *Annals of Internal Medicine* 156 (2012): 253–62.

[24] Bang, Heejun, et al. "Development and Validation of a Patient Self-Assessment Score for Diabetes Risk." *Annals of Internal Medicine* 151 (2009): 775–83.

[25] American Diabetes Association. "Are You at Risk for Type 2 Diabetes?" Diabetes Risk Test. Available at http://www.diabetes.org/diabetes-basics/prevention/diabetes-risk-test (accessed March 20, 2013).

[26] Centers for Medicare and Medicaid Services. "Stage 2."

[27] Yoon, P., et al. "Developing Family Healthware, a Family History Screening Tool to Prevent Common Chronic Diseases." *Preventing Chronic Disease* 6, no. 1 (2009). Available at http://www.cdc.gov/pcd/issues/2009/jan/07_0268.htm (accessed August 13, 2013).

[28] Quereshi, N., et al. "Collecting Genetic Information in Primary Care: Evaluating a New Family History Tool." *Family Practice* 22, no. 6 (2005): 663–69.

[29] Quereshi, N. *Family History and Improving Health* (AHRQ Publication No. 09-E016). Evidence Report/Technology Assessment No. 186. Rockville, MD: Agency for Healthcare Research and Quality, August 2009. Available at http://www.ncbi.nlm.nih.gov/books/NBK32554/ (accessed February 7, 2013).

Article citation:
Hoyt, Robert; Linnville, Steven; Chung, Hui-Min; Hutfless, Brent; Rice, Courtney. "Digital Family Histories for Data Mining" *Perspectives in Health Information Management* (Fall, October 2013).

Driving the Power of Knowledge

Copyright 2022 by The American Health Information Management Association. All Rights Reserved.