# Genetic risk estimation in the Coriell Personalized Medicine Collaborative 

Catharine B. Stack, PhD, Neda Gharani, PhD, Erynn S. Gordon, MS, Tara Schmidlen, MS, Michael F. Christman, PhD, and Margaret A. Keller, PhD


#### Abstract

Purpose: Recent genome wide-association studies have identified hundreds of single nucleotide polymorphisms associated with common complex diseases. With the momentum of these discoveries comes a need to communicate this information to individuals. Methods: The Coriell Personalized Medicine Collaborative is an observational research study designed to evaluate the utility of personalized genomic information in health care. Participants provide saliva samples for genotyping and complete extensive on-line medical history, family history, and lifestyle questionnaires. Only results for diseases deemed potentially actionable by an independent advisory board are reported. Results: We present our methodology for developing personalized reports containing risks for both genetic and nongenetic factors. Risk estimates are given as relative risk, derived or reported from representative peer-reviewed publications. Estimates of disease prevalence are also provided. Presenting risk as relative risk allows for consistent reporting across multiple diseases and across genetic and nongenetic factors. Using this approach eliminates the need for assumptions regarding population lifetime risk estimates. Publications used for risk reporting are selected based on the strength of the design and study quality. Conclusion: Coriell Personalized Medicine Collaborative risk reports demonstrate an approach to communicating risk of complex disease via the web that encompasses risks due to genetic variants along with risks caused by family history and lifestyle factors. Genet Med 2011:13(2):131-139.


Key Words: genetic risk assessment, personalized medicine, family history risk, lifestyle risk, genetic education, relative risk, absolute risk

Hundreds of associations between common genetic variants and complex disease have been discovered with the advent of the genome-wide association study (GWAS). Associations contained within the Catalog of Published GWAS established by the National Human Genome Research Institute currently total up to 3038 based on 620 published studies (as of August 30, 2010). ${ }^{1}$ With this growing resource comes a need to assess the utility of this information and the responsibilities for returning newly discovered genetic results to the public. There is a strong public interest in and anticipation of return of research results to participants. ${ }^{2,3}$ However, the scientific and bioethics communities continue to debate the topic with vocal support for sharing based on ethical principles of respect and autonomy ${ }^{4-6}$ and selective opposition based on the balance of benefits and harms. ${ }^{7-9}$ As acknowledged by Beskow and Burke ${ }^{10}$ as well as Ravitsky and Wilfond, ${ }^{11}$ key to this debate is the context of the

[^0]study and factors such as the clinical utility and personal meaning of results, investigator capabilities, and alternative access.

Historically, clinical risk scores have been based on longterm prospective studies; however, very few GWAS associations have been studied prospectively making traditional approaches to risk estimation difficult in the emerging GWAS era. Although the call for large cohort studies of genetic variants and complex disease has been made, ${ }^{12,13}$ such prospective studies will require substantial resources and time. In the absence of these data, there is a need to move forward in developing methods for communicating the growing body of genetic data to the public and studying its impact on health care. ${ }^{14} \mathrm{~A}$ recent review of Web site content from current health-related direct-to-consumer (DTC) companies ${ }^{15}$ found that only $13 \%$ (2 of 15) of the Web sites reporting results from single nucleotide polymorphism (SNP)-based tests explained how risk was estimated. Just $53 \%$ ( 8 of 15) included scientific evidence or data to support marker selection and only $13 \%$ referenced specific publications used. Common language was not consistently used, and Web sites required high levels of literacy, with an average reading level of grade 15 .

Here, we describe development of the methods used for reporting risk to participants of an observational study designed to examine the utility of personalized genomic information in health care. We begin by providing a brief background on the Coriell Personalized Medicine Collaborative ( $\mathrm{CPMC}^{\circ}$ ) cohort and a discussion of possible risk metrics to communicate genetic risk. The final sections of this article detail the steps that we follow in developing risk reports for CPMC participants.

## CORIELL PERSONALIZED MEDICINE COLLABORATIVE

The CPMC is a multi-institutional effort involving study participants, medical professionals, scientists, ethicists, and information technology experts. An overview of the CPMC project has recently been reported elsewhere. ${ }^{16}$ Briefly, consented participants provide saliva samples, which are genotyped using the Affymetrix 6.0 GeneChip and the Affymetrix Drug Metabolizing Enzymes and Transporters Plus GeneChip in Coriell's CLIA-certified Genotyping and Microarray Center. Using a secure web-based portal, the CPMC provides participants with educational materials, collects participant data through questionnaires, and reports personalized results for potentially actionable health conditions. ${ }^{17}$ Results are reported at periodic intervals, and participants are able to request an in-person or telephone genetic counseling appointment, at no charge, through the webbased portal. Approximately $15 \%$ of CPMC participants have used the free genetic counseling services. ${ }^{18}$

Participant data collected through online medical history, family history, lifestyle questionnaires (MFLQ) includes demographics; medication information; current and past diseases; history of cancer screening; pregnancy; lifestyle behaviors such as smoking, exercise, and alcohol use; and current and past

Fig. 1. CPMC process for risk reporting. A, Candidate genetic variant/health condition associations are identified by searches of public databases such as PubMed, HuGENet, ${ }^{19}$ and the National Human Genome Research Institute GWAS catalog. ${ }^{20}$ The minimum inclusion criteria for candidate genetic variants is a documented association with the disease in a moderately sized study (defined as having 500 or more subjects) with replication in one or more cohorts of the same race, either replicated within a single peer-reviewed publication or published in a separate study. For each health condition, SNPs are prioritized according to the greatest amount of supportive evidence (according to the variant selection hierarchy). SNPs with multiple independent published reports and preferably additional supportive meta-analysis are preferentially chosen. If more than one equally well-supported risk variant is available for a given health condition, the SNP with the largest OR for disease is selected. If the OR estimates do not differ for more than one genetic variant, then an arbitrary selection of the risk variant is made. Currently, only genetic variants contained on the Affymetrix 6.0 platform have been considered for inclusion in the CPMC. Additional variants not included on this platform
 variant associations considered valid and health conditions deemed to be potentially actionable are approved for CPMC risk reporting. C, Publications used for risk reporting are selected based on the strength of the study design and an assessment of study quality. D, RR values are either reported directly from the selected paper (cohort studies) or estimated from the reported OR (case-control studies). When the OR overestimates the estimated RR by $10 \%$ or less, the OR is reported as an estimate of RR.
diseases in first-degree relatives and grandparents. Participants are required to complete the MFLQ before they are eligible to receive results.

Selection of the specific genetic variant, per health condition, for risk reporting is based on a rigorous process of data curation. Peer-reviewed scientific literature and public databases ${ }^{19,20}$ are searched to identify candidate genetic variants associated with common complex diseases for which there is a potential action that can be taken to reduce the disease risk or improve the prognosis. The inclusion criteria and ultimate selection of a given candidate genetic risk variant is depicted in Figure 1, A. Once selected, information on the candidate genetic variants, a summary of the evidence supporting their association with specified health conditions, and a clinical review for each health condition is compiled and submitted for evaluation by an independent advisory board called the Informed Cohort Oversight Board (ICOB).

The ICOB model was adopted from Kohane et al. ${ }^{21}$ The ICOB assesses submissions for each health condition to determine whether the candidate health conditions are potentially actionable. A potentially actionable condition is defined as a condition for which risk can be mitigated by individual action (behavior and lifestyle) or by medical action (screening, preventative treatment, and early intervention). Health condition summaries submitted to the ICOB contain information from medical literature reviews and medical society policy statements. In addition, this board assesses the validity of associations between proposed genetic variant(s) and health condition through review of the summary of genetic association studies provided within the submission document. The board meets twice a year and includes scientists skilled in genetic studies, medical professionals familiar with the use of genetics in medical care, and an ethicist and a pastor who can share insight into the moral and ethical implications of the study, provide valuable feedback on their understanding of genetic results as members of a lay audience, and serve as advocates for the public. Once a health condition and genetic variant have been approved by the

ICOB (as displayed in Fig. 1, B), it is eligible to be reported to CPMC participants.

When developing risk reports and methods for communicating genetic risk, our goals were (1) to present quantitative risk values, based on published associations between genetic variants and complex disease; (2) to present genetic risks paired with other factors, such as family history, lifestyle factors, and comorbidities; and (3) to report results directly from the literature without performing extensive recalculations or applying uncertain assumptions.

## CHOICE OF RISK METRIC

The strength of the association between a genetic variant and disease is typically reported by GWAS in relative terms, using odds ratios (ORs). In this section, we highlight the characteristics of common measures of disease association proposed for reporting genetic associations with complex disease (Table 1) and provide a rationale for using relative risk (RR).

## OR and RR

An OR is the odds of disease in an exposed group (those with the genetic risk variant) divided by the odds of disease in an unexposed group (those without the genetic risk variant) (Table 1). The OR can be estimated using data from case-control studies. For example, suppose a hypothetical case-control GWAS reports an OR of 1.4 for a specific genetic variant, it means that the odds of disease among those with the genetic variant is 1.4 times as large as (or $40 \%$ greater than) the odds of disease in those without the genetic variant.

RRs differ from ORs in that they are ratios of risk. A RR is found by dividing the risk or probability of disease in the exposed group by the risk of disease in the unexposed group (Table 1). These risks of disease are not measured directly from case-control studies; however, when certain conditions are met, the OR from a case-control study may be interpreted as a RR. When the underlying disease prevalence is relatively high (over

Table 1 Measures of association between a genetic variant (G) and disease (D)

| Measure of association | Definition | Measured directly by GWAS (case-control studies)? |
| :---: | :---: | :---: |
| Odds ratio | $\mathrm{OR}=\frac{\operatorname{Pr}(\mathrm{D}+/ \mathrm{G}+) / \operatorname{Pr}(\mathrm{D}-/ \mathrm{G}+)}{\operatorname{Pr}(\mathrm{D}+/ \mathrm{G}-) / \operatorname{Pr}(\mathrm{D}-/ \mathrm{G}-)}$ <br> Ratio of the odds of disease in the group with the genetic variant to the odds of disease in the group without the genetic variant | Yes. In case-control studies, the odds ratio can be found using the ratio of exposure odds based upon the probabilities $\operatorname{Pr}(\mathrm{G}+/ \mathrm{D}+), \operatorname{Pr}(\mathrm{G}+/ \mathrm{D}-)$, $\operatorname{Pr}(\mathrm{G}-/ \mathrm{D}+)$, and $\operatorname{Pr}(\mathrm{G}-/ \mathrm{D}-)$. |
| Relative risk (risk ratio) | $\mathrm{RR}=\frac{\operatorname{Pr}(\mathrm{D}+/ \mathrm{G}+)}{\operatorname{Pr}(\mathrm{D}+/ \mathrm{G}-)}$ <br> Ratio of the risk of disease in the group with the genetic variant to the risk of disease in the group without the genetic variant | No; however, when the underlying disease prevalence is low and the odds ratio is close to the null value of 1 , the odds ratio can be used as an approximation to the relative risk. ${ }^{22}$ |
| Absolute risk | $\mathrm{AR}=\operatorname{Pr}(\mathrm{D}+/ \mathrm{G}+)$ <br> Risk or probability of disease in the group with the genetic variant | No. Most appropriate methods for estimating absolute risk as lifetime risk require population-based disease registry data. ${ }^{23,24}$ Such results apply best to individuals from populations similar to the disease registry population. |
| Population attributable fraction | $\mathrm{PAF}=\frac{\operatorname{Pr}(\mathrm{D}+)-\operatorname{Pr}(\mathrm{D}+\mid \mathrm{G}-)}{\operatorname{Pr}(\mathrm{D}+)}$ <br> Proportion of the total disease in the population that could be removed if the exposure (genetic variant) were eliminated | No. Required assumptions are unlikely to hold for GWAS results. |

about $10 \%$ ) or the OR is not close to the null value of 1 , ORs should not be interpreted as RR because they may markedly overstate risk. ${ }^{22}$ In the hypothetical GWAS example, if the underlying risk of disease is $10 \%$ or less, the OR approximates the RR, and one may use the OR to conclude the risk of disease in those with the genetic variant is 1.4 times (or $40 \%$ greater than) the risk of disease in those without the variant. As disease risks are easier to interpret than disease odds, ORs are commonly interpreted as RRs, ${ }^{25}$ even though this interpretation is not always appropriate. ${ }^{22,26}$

Advantages for reporting genetic risk in relative terms include (1) the risk value can be reported either directly (RR) or after approximation (OR) from the published study describing the association; (2) ORs and RRs can be reported for nongenetic risk factors using a similar construct; and (3) measures of relative effect are generally stable across populations, ${ }^{27}$ such that these risks are likely to be relevant to populations with different baseline risk. The primary disadvantage of using a measure of relative effect to describe risk is relative effects can be difficult to conceptualize and describe to a lay audience. ${ }^{28} \mathrm{An}$ important limitation of using a measure of relative effect to communicate risk is that information regarding the baseline risk is absent. For very rare diseases with strong risk factors, RRs will be large and absolute risk (AR) will be small and, presenting a measure of relative effect alone, may exaggerate risk. ${ }^{28}$

## Absolute risk

The format commonly used by DTC companies is a presentation of AR. AR or probability of disease (Table 1) cannot be measured directly in case-control studies, except in the rare situation when the sampling fractions for cases and controls are known. ${ }^{29}$ Methods exist for estimating AR as lifetime risk using RR, age-stratified disease incidence, and mortality rates from population-based disease registries or similar data sources $^{23,30,31}$. Estimates found using this method are most applicable to individuals who are similar to the population used to derive lifetime risk estimates (with respect to baseline disease risk and country of residence, for example). Other, less supported approaches have also been used to estimate AR without using age-stratified rates or accounting for competing risks.

The greatest advantage of using AR to communicate genetic risk is its ease of interpretation. The greatest disadvantage of using $A R$ is that methods used to translate measures of relative effect to AR require prospective disease registry data, which are not consistently available for all diseases. AR estimates are very sensitive to the input values of lifetime risk and thus may not be relevant across population groups. Yang et al. ${ }^{30}$ recently demonstrated the importance of accounting for the cumulative uncertainty and variation in the input parameters when deriving absolute disease risk based on genetic variants using this methodology. Others have noted differences in estimates of AR for a single disease supplied by DTCs, stemming from variation in the lifetime risk of disease estimates used. ${ }^{32,33}$

## Population attributable fraction

Population attributable fraction (PAF) is a measure of the proportion of disease burden in the population that could be eliminated if a given causal exposure was eliminated. PAF can be estimated using estimates of RR and exposure prevalence, provided the relationship between exposure and disease are causal, and there is no confounding of the exposure-disease association. ${ }^{34}$ The assumption of a causal association between a given genetic variant and disease is not well supported for
variants identified through GWAS, except where function has been demonstrated. PAF is not an ideal choice for risk communication as it is commonly misinterpreted ${ }^{35}$ and miscalculated ${ }^{36}$ in the literature.

## CPMC RISK REPORTING

We chose to report risk to study participants as RR because it allowed us to present risks for all factors (genetic, family history, and lifestyle) using the same metric and did not require population estimates of disease incidence. A sample risk summary page of a CPMC risk report is given in Figure 2. Solid discs represent the RR for the participant's risk group, and vertical cylinders depict the range of RR values across risk groups. On-line risk reports are organized using a tabbed approach, with separate tabs for disease condition information, risk results, limitations, methods, and links to request genetic counseling, review educational material, or register to attend a free disease specific educational seminar cohosted by a physician expert and CPMC genetic counselor. Current risk reports present genetic variant risk based on a single SNP because of the lack of validated multigenic models with robust prediction. The following sections describe our procedures for creating risk reports.

## Genetic variant risk: study selection

The first step in genetic risk reporting is selecting the published study from which risk data will be used, as illustrated in Figure 1, C. Peer-reviewed literature reporting associations between the approved genetic variant and complex disease are further curated, starting with the studies identified in the initial genetic variant selection step (Fig. 1, A) and adding new search results from public databases ${ }^{19,20}$ and their references. Potential publications are grouped by their design and ranked according to their ability to provide representative and valid estimates of association.

Next, those studies ranked within the highest available design tier are evaluated for study quality based on published recommendations. ${ }^{38,39}$ Study quality items include disease definition (was it reported, was it carried out consistently in all study subjects, and was it objective and clinically accepted), genotyping methods (were they reported, were they carried out consistently, and was the variant genotyped directly and not imputed), and population stratification (were analyses stratified by ancestry or adjusted for population structure). Meta-analyses that do not report results by ethnic group are generally not selected, unless justification for the pooling is provided. Studies must report an estimate of relative effect (RR, OR, or hazard ratio). If only allelic ORs are reported, Hardy-Weinberg equilibrium and additive effects are assumed to determine ORs by genotype. If no studies in the highest tier of study design meet study quality and reporting criteria, the process is repeated using all studies from the next highest study design tier. If more than one study meets the quality and reporting criteria, the largest study is used. An exception to these rules exists. If there is one prospective study in a representative US population providing estimates of both genetic and nongenetic risk factors, then this study is given preference.

## Genetic variant risk: risk estimation

The specific RR value is found using the selected study as shown in Figure 1, D. When the selected study is a cohort study, estimates of RR are usually reported directly (and hazard ratios are reported as RRs). If the cohort study only reports ORs, then $R R$ is estimated using the known relationship, $R R=O R /[(1-$


Fig. 2. Sample CPMC risk summary report. A sample risk report is shown for type 2 diabetes. Results are for a hypothetical 50 -year-old Caucasian woman with a family history of disease, body mass index between $25 \mathrm{and} 30 \mathrm{~kg} / \mathrm{m}^{2}$, and heterozygous for the genetic risk variant (rs7754840). The risk summary includes RRs for family history (magenta disc), nongenetic factors (blue disc), and genetic variant (green disc), based on information supplied by the participant in the MFLQ. The height of the gray cylinders depicts the range of risk. The tabs for individual "Genetic Variant Risk," "Family History Risk," and "Other Risk" contain only the cylinders within that category with supporting text; interpretative text appears when participants mouse over individual cylinders. An estimate of disease prevalence for the age, gender, and ethnicity of the participant is provided below the risk plot. The "About" tab contains information on the disease prevalence and heritability, and the "Genetic Variant" tab contains information on the genotype frequency. The "Methods" and "Limitations" tabs contain technical information, and the "What do I do now?" tab supplies links to supporting web pages, an educational video, or request for genetic counseling. Additional sample reports may be viewed from the CPMC Web site ${ }^{37}$ by selecting one of the two sample individuals and then selecting the "View My Result" button for a specific disease.
$\left.\left.p_{0}\right)+p_{0} \times \mathrm{OR}\right]$, where $p_{0}$ is the prevalence of disease in the unexposed group. ${ }^{40}$ When the selected study is a case-control study, RR is estimated based on the reported OR and an estimate for the prevalence of disease in the unexposed group. Specifically, $\mathrm{RR}=\mathrm{OR} /[(1-p)+p \times \mathrm{OR}]$, where $p$ is the prevalence of disease in the population from which the cases and controls were sampled. ${ }^{40}$ When the OR overstates the RR by $10 \%$ or less, the OR is reported as an estimate of $R R$. In exceptional situations when $A R$ is available from a cohort study, and RR alone may not be meaningful, we report AR along with RR (e.g., in male C282Y homozygotes, the AR of hemochromatosis is $33-57 \%$ and the RR for this association is 27).

## Genetic risk estimates for non-Caucasians

As the majority of published studies are conducted in populations of Caucasian descent, we initially determine risk estimates for genetic variants in Caucasian populations. If studies in non-Caucasian populations are available, then we report race/ ethnic appropriate risk estimates for the participant. If race/ ethnic specific studies are not available, or if the studies are of poor quality or have indeterminate results, then we present risk estimates from the Caucasian population. Individuals of mixed race or ethnicity are presented risk estimates from the Caucasian population. In all cases, results are presented with a statement of the population used for risk estimation.

If there are multiple studies reporting associations in nonCaucasian populations, one study is selected using the same strategy described above. Because the ICOB does not assess the robustness of genetic risk associations across all populations, additional study criteria are applied to non-Caucasian datasets. Study size is generally required to be at least 200 cases, and studies including more than one race or ethnicity must use analyses to account for population stratification.

## Nongenetic risk determination

Risk estimates provided for other factors such as family history and lifestyle or environmental factors are derived or reported from valid and representative peer-reviewed publications. Nongenetic factors are included if they meet two criteria. First, the factor must be collected by the MFLQ. Second, the factor must be an established disease risk, included in multiple disease review articles and consistently associated with disease. For example, blood pressure and cholesterol levels are established risk factors for coronary artery disease ${ }^{41}$ but are not included in risk reports because blood pressure and cholesterol values are not collected in the MFLQ because of poor selfreport reliability. Alternatively, diabetes status is a strong predictor of coronary artery disease collected in the MFLQ and therefore is included in the coronary artery disease risk report. For risk factors meeting these two criteria, a peer-reviewed publication is selected following the strategy described above. As most of the nongenetic risk factors have been studied prospectively, RRs are generally reported directly from the selected publication. In cases where RRs are not reported, they are estimated from ORs using the methods described above.

A summary of RR values and studies reporting associations for disease conditions released to study participants as of February 2010 are given in Table 2.

## Disease prevalence, genotype frequency, and heritability

A summary of disease prevalence by race, age, and gender is included in each risk report to provide context for RR values. When available, prevalence estimates are based on the US population and come from federal disease registries (e.g., CDC, NCHS, and SEER) or from peer-reviewed literature.

Genotype frequencies, based on the participant's self-reported race and ethnicity, are also provided within each report, when available. Frequency values are based on the study population in which genetic risk estimates were determined, another peer-reviewed publication, or a relevant HapMap population (African ancestry in Southwest USA or Yoruba for African American, for example). For ethnic groups not present in HapMap, another closely related ethnic group or a Caucasian group is selected as the default, and frequencies from the default group (with disclosure) are reported. Published estimates of heritability, generally from family or twin studies, are also reported to illustrate the multifactorial nature of the diseases being reported. Disease condition pages provide both graphical and text descriptions for heritability and disease prevalence. Results pages use schematics with large grids of individual people ${ }^{57}$ to depict genotype frequencies.

## DISCUSSION

We have developed a unique method for reporting genetic risk of complex disease to individuals, whereby, in contrast to

DTC genetic testing companies, genetic risks are presented with estimates of risks caused by nongenetic factors. Results are reported for potentially actionable diseases only, following an assessment made by an independent advisory board. We present risk as RRs, which are generally stable across populations and do not require estimates of population lifetime disease risks. Although in adopting this approach, we trade some conceptual simplicity for accuracy and transparency, interviews with study participants suggest participants have a good understanding of their results. ${ }^{58}$ Our methodology allows us to present multiple risk factors within the same framework, across multiple diseases, and to multiple ethnic groups.

The Multiplex Initiative (MI) also used a deliberate approach to determine a select group of genetic associations to report to participants. ${ }^{59,60}$ They used an interdisciplinary and iterative approach to determine a set of 15 genetic variants and 8 common diseases meeting specific criteria, derived by an external working group of scientists. Similar to our study, associations were required to show consistent replication in adequately sized studies, and health conditions were evaluated based on existing prevention strategies. Unlike our study, MI required diseases to be adult onset and required genetic variants to have frequencies of $5 \%$ or more and be associated with a minimum $10 \%$ change in risk. ${ }^{60}$ Participants of MI received one risk report through the mail with a follow-up telephone call and reports did not include nongenetic risks. ${ }^{61}$ In our study, on-line results are delivered dynamically and reports contain both genetic and nongenetic risks.

A number of DTC companies report genetic risk to customers for a panel of complex diseases. Most of these companies present risk as AR and most provide risk estimates for one complex disease based on multiple genetic variants, under the assumption all variants act independently. A recent comparison of results for five individuals by two DTC companies found $50 \%$ or less agreement in risk interpretations. ${ }^{32}$ Disagreement primarily resulted from (1) using different values for average population disease risk and (2) the selection of genetic markers. Another comparison of genetic risks supplied by three DTC companies also noted the impact of differences in assumed average lifetime disease risk on estimates of absolute disease risk. ${ }^{33}$ Specifically, assumed lifetime risks for heart attack varied from 21 to $47 \%$ in men and from 7 to $25 \%$ in women, across three DTC companies. ${ }^{33}$ There is a direct relationship between assumed population disease risk and final AR estimates. An inappropriate population disease risk can lead to invalid and misleading estimates of AR. By reporting risk in relative terms, our study avoids this problem of selecting an appropriate population lifetime disease risk estimate and gains the ability to cite risk estimates directly from the peer-reviewed literature. Furthermore, unlike most DTC companies, we collect detailed medical, family history and lifestyle information and incorporate some of this information into risk reports, so as to communicate the multifactorial nature of complex diseases.

The REVEAL and ongoing ClinSeq studies have also reported genetic risk of a complex disease to individuals, even though both were designed with a more narrow disease focus than the CPMC. REVEAL was designed to examine the behaviors and psychological impact of communicating APOE genotype to first-degree relatives of patients with Alzheimer disease. ${ }^{62}$ With access to data from a large prospective study of first-degree relatives of patients with Alzheimer disease, REVEAL investigators were able to present risk using risk curves of AR of Alzheimer disease by age. The ClinSeq project targets individuals at varying risk for atherosclerotic disease, with the goal of sequencing approximately 400 genes believed to be

Table 2 Genetic variant, family history, and other risk factor relative risks for CPMC risk reports released to participants as of February 2010

| Disease condition | Risk factor | Relative risk ${ }^{a}$ | Study design | Multivariable adjustment terms | Reference |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Coronary artery disease | ```Genetic variant region between CDKN2A and CDKN2B (rs1333049)``` | $\begin{aligned} & 1.3 \text { (CG vs. GG), } \\ & 1.7 \text { (CC vs. GG) } \end{aligned}$ | Meta-analysis of casecontrol studies | None | 42 |
|  | Family history ${ }^{b}$ | 1.4 (males), 1.2 (females) | Prospective | Age, blood pressure, cholesterol, smoking, glucose intolerance, weight, father and mother's age at death, and left ventricular hypertrophy | 43 |
|  | Diabetes | $\begin{aligned} & 1.7 \text { (males), } \\ & 2.4 \text { (females) } \end{aligned}$ | Prospective | Age, blood pressure, smoking, and cholesterol | 44 |
|  | Current smoking | 2.1 (males), 2.6 (females) | Prospective | Age, blood pressure, diabetes, and cholesterol | 44 |
| Type 2 Diabetes | $\begin{aligned} & \text { Genetic variant } \\ & \text { CDKAL1 } \\ & \text { (rs7754840) } \end{aligned}$ | $\begin{aligned} & 1.2 \text { (CG vs. GG), } \\ & 1.3 \text { (CC vs. GG) } \end{aligned}$ | Prospective | Age, age squared, and sex | 45 |
|  | Family history ${ }^{b}$ | 1.9 | Prospective | Age, gender, and body mass index | 46 |
|  | Body mass index | $\begin{aligned} & 2.3(25-29.9 \text { vs. } \\ & <25) ; 5.9(\geq 30 \\ & \text { vs. }<25) \end{aligned}$ | Prospective | Age, gender, and family history | 46 |
| Iron Overload/ Hemochromatosis | $\begin{aligned} & \text { Genetic variant } H F E \\ & (\text { rs1800562 }) \end{aligned}$ | GG:0-4\% (males), $0-1 \%$ (females); AG:0-5\% (males), 0-1\% (females); AA: 33-57\% (males), 3-15\% (females) |  | N/A | 47 |
| Prostate cancer | ```Genetic variant 8q24.21 intergenic region (rs16901979)``` | $\begin{aligned} & 1.5(\mathrm{CA}+\mathrm{AA} \text { vs. } \\ & \mathrm{CC}) \end{aligned}$ | Population-based casecontrol | Age, geographic region, family history, and four other genetic variants | 48 |
|  | Family history ${ }^{b}$ | 1.9 | Prospective | Age, body mass index, height, smoking, physical activity, diabetes, race, and diet | 49 |
| Melanoma | $\begin{aligned} & \text { Genetic variant } P I G U \\ & \text { (rs910873) } \end{aligned}$ | $\begin{aligned} & 1.7 \text { (CT vs. CC), } \\ & 3.0 \text { (TT vs. CC) } \end{aligned}$ | Case-control | None | 50 |
|  | Family History ${ }^{\text {b }}$ | 2.2 | Prospective | Age, gender, history of severe sunburn, moles on arms/ lower legs, and hair color | 51 |
| Age-related macular degeneration | $\begin{gathered} \text { Genetic variant } \\ \text { LOC387715 } \\ \text { (rs10490924) } \end{gathered}$ | 2.4 (GT vs. GG), <br> 6.0 (TT vs. GG) | Meta-analysis of casecontrol studies | None | 52 |
|  | Family history ${ }^{b}$ | 3.9 | Population-based casecontrol | Age, gender, and current smoking | 53 |
|  | Smoking | 1.4 (former vs. never), 2.1 (current vs. never) | Meta-analysis of cohort studies | Age and gender | 54 |
| Lupus | $\begin{aligned} & \text { Genetic variant } \\ & \text { STAT4 } \\ & (\mathrm{rs} 3821236) \end{aligned}$ | $\begin{aligned} & 1.4 \text { (GA vs. GG), } \\ & 2.0 \text { (AA vs. GG) } \end{aligned}$ | Case-control | None | 55 |

Table 2 Continued

| Disease condition | Risk factor | Relative risk $^{a}$ | Study design | Multivariable adjustment terms | Reference |
| :--- | :---: | :---: | :---: | :---: | :---: |
|  | Family history ${ }^{b}$ | 4.1 (1 first-degree <br> relative) 11.3 <br> $(\geq 2$ first-degree <br> relatives) | Case-control | Age, sex, and family size |  |
|  | Current smoking | 1.5 | Meta-analysis of cohort <br> and case-control <br> studies | Varied by study |  |

${ }^{a^{\prime}}$ Absolute risk reported for iron overload/hemochromatosis.
${ }^{b}$ Positive family history defined as one or both parents dying of coronary artery disease, one or both parents with type 2 diabetes, one or more first-degree relatives with melanoma, biological father or any brothers diagnosed with prostate cancer, one or more first-degree relatives with age-related macular degeneration, and one or two or more first-degree relatives with a history of any of 11 autoimmune diseases.
involved in atherosclerosis. ${ }^{63}$ Results will only be returned for those variants/mutations associated with highly penetrant Mendelian disorders after confirmation in a CLIA approved laboratory, ${ }^{63}$ and sequencing results will be used to identify new variants influencing risk of atherosclerosis. The REVEAL and ClinSeq studies illustrate the current climate of genetic testing for complex diseases, where testing protocols focus on known highly penetrant genes and available risk data.

Framing risk in relative or absolute terms has been shown to impact patient decision making. A systematic review of studies investigating how patients make treatment decisions concluded that patients were more likely to choose a treatment when the treatment effect was presented in terms of RR reduction when compared with AR reduction. ${ }^{64}$ Possible explanations for why patients perceive a greater treatment benefit when the effect is presented in relative terms include the tendency for patients to ignore the underlying risk of disease ${ }^{65}$ in addition to a lack of understanding of the reference group used in deriving a relative effect. ${ }^{28}$ This literature suggests that in ideal situations, when accurate estimates of AR are available, risk should be communicated using absolute numbers with visual aids. ${ }^{28,57,66}$ Because our objective is to report both genetic and nongenetic factors across multiple health conditions, we could not consistently derive accurate estimates of AR. Rather than base risks on possibly inappropriate values of lifetime risk, or limit results to genetic risk factors only, we chose to present more generally applicable, RR values. Given the challenges of communicating risk in relative terms, we include detailed descriptions for RR within the reports and provide educational material on the Web site. As noted previously, we provide estimates of disease prevalence within each risk page to help participants interpret their results in context and provide access to genetic counseling at no charge.

There are additional limitations associated with our methods. Specifically, the reported RR estimates for genetic and nongenetic risk factors may not be independent, or comparable, since they may come from different statistical models with different sets of covariates. Family history and genetics in addition to family history and lifestyle factors likely overlap, a point highlighted to participants in our reporting. Another current limitation is the focus on single genetic variants because of the lack of validated multigenic models. However, the risk reporting approach was designed to be adapted to include multigenic models as they are described and validated in the literature. Finally, as additional lifestyle and medical history factors are integrated into CPMC questionnaires, the ability to report other nongenetic risks will expand, highlighting the dynamic nature of the study.

There is an urgent need to develop responsible and clinically meaningful methods for reporting genetic risk to individuals. A number of approaches to communicating genetic risk to individuals have been developed and used. Our reporting is unique because we report genetic risk with other risks based on family history and lifestyle factors, to better communicate the multifactorial nature of complex diseases. This approach can be applied to multigenic risk models, as they are developed and validated. In addition, the report format is amenable to integration into an electronic health record or personal health record. By creating a dynamic system for reporting both genetic and nongenetic risks of potentially actionable diseases using transparent methods, we strive to move the field of personalized medicine forward.

## ACKNOWLEDGMENTS

This research was supported by Grants from the William G. Rohrer Foundation and the RNR Foundation. We acknowledge the members of the Informed Cohort Oversight Board (Erin O'Shea, Jennifer Hoheisel, Marc Lenburg, Steve Murphy, Kenneth Offit, David Pellman, Charles Rotini, Reverend Floyd White, and Ellis Neufeld) for their critical role in the study. We thank Coriell staff from the Information Systems Department and the Genotyping and Microarray Center for their contributions to the project.

## REFERENCES

1. Hindorff LA, Ha J, Mehta JP, Manolio TA. A catalog of published genomewide association studies. Available at: www.genome.gov/gwastudies. Accessed August 30, 2010.
2. Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. Public expectations for return of results from large-cohort genetic research. $A m J$ Bioeth 2008;8:36-43.
3. Wellcome Trust and Medical Research Council. Public perceptions of the collection of human biological samples. Summary report. London, UK: Wellcome Trust and Medical Research Council, 2000.
4. Rothstein MA. Tiered disclosure options promote the autonomy and wellbeing of research subjects. Am J Bioeth 2006;6:20-21.
5. Shalowitz DI, Miller FG. Disclosing individual results of clinical research: implications of respect for participants. JAMA 2005;294:737-740.
6. Partridge AH, Winer EP. Informing clinical trial participants about study results. JAMA 2002;288:363-365.
7. Fuller BP, Kahn MJ, Barr PA, et al. Privacy in genetics research. Science 1999;285:1359-1361.
8. Beskow LM, Burke W, Merz JF, et al. Informed consent for populationbased research involving genetics. JAMA 2001;286:2315-2321.
9. Clayton EW, Ross LF. Implications of disclosing individual results of clinical research. JAMA 2006;295:37.
10. Beskow LM, Burke W. Offering individual genetic research results: context matters. Sci Transl Med 2010;2:38cm20.
11. Ravitsky V, Wilfond BS. Disclosing individual genetic results to research participants. Am J Bioeth 2006;6:8-17.
12. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006;7:812-820.
13. Manolio TA. Cohort studies and the genetics of complex disease. Nat Genet 2009;41:5-6.
14. McBride CM, Brody LC. Point: genetic risk feedback for common disease time to test the waters. Cancer Epidemiol Biomarkers Prev 2007;16:17241726.
15. Lachance CR, Erby LA, Ford BM, Allen VC Jr, Kaphingst KA. Informational content, literacy demands, and usability of websites offering healthrelated genetic tests directly to consumers. Genet Med 2010;12:304-312.
16. Keller M, Gordon ES, Stack CB, et al. The Coriell Personalized Medicine Collaborative: a prospective study of the utility of personalized medicine. Personalized Med 2010;7:301-317.
17. CPMC. Available at: http://cpmc.coriell.org. Accessed June 15, 2010.
18. Schmidlen T, Gordon ES. Genomic literacy: Emerging themes among genetic counseling inquires from participants of the Coriell Personalized Medicine Collaborative. Presented at the 8th Annual Education Conference of the National Society of Genetic Counselors, Atlanta, Georgia. November 2009.
19. HuGENet. Available at: http://www.cdc.gov/genomics/hugenet/default.htm. Accessed June 15, 2010.
20. NHGRI GWAS catalog. Available at: http://www.genome.gov/gwastudies/. Accessed June 15, 2010.
21. Kohane IS, Mandl KD, Taylor PL, Holm IA, Nigrin DJ, Kunkel LM. Medicine. Reestablishing the researcher-patient compact. Science 2007;316: 836-837.
22. Davies HT, Crombie IK, Tavakoli M. When can odds ratios mislead? BMJ 1998;316:989-991.
23. Yang Q, Khoury MJ, Coughlin SS, Sun F, Flanders WD. On the use of population-based registries in the clinical validation of genetic tests for disease susceptibility. Genet Med 2000;2:186-192.
24. Yang Q, Flanders WD, Moonesinghe R, Ioannidis JP, Guessous I, Khoury MJ. Using lifetime risk estimates in personal genomic profiles: estimation of uncertainty. Am J Hum Genet 2009;85:786-800.
25. Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. $N$ Engl J Med 1999;341:279-283.
26. Jewell NP. Statistics for epidemiology. Boca Raton, Florida: Chapman \& Hall; 2004.
27. Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in metaanalyses. Int J Epidemiol 2002;31:72-76.
28. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ 2003;327:741-744.
29. Schlesselman JJ. Case-control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
30. Yang Q, Flanders WD, Moonesinghe R, Ioannidis JP, Guessous I, Khoury MJ. Using lifetime risk estimates in personal genomic profiles: estimation of uncertainty. Am J Hum Genet 2009;85:786-800.
31. Gail MH. Estimation and interpretation of models of absolute risk from epidemiologic data, including family-based studies. Lifetime Data Anal 2008;14:18-36.
32. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature 2009;461:724-726.
33. Swan M. Multigenic condition risk assessment in direct-to-consumer genomic services. Genet Med 2010;12:279-288.
34. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88:15-19.
35. Janssens AC, Gwinn M, Valdez R, Narayan KM, Khoury MJ. Predictive genetic testing for type 2 diabetes. BMJ 2006;333:509-510.
36. Severi G, Byrnes GB, Hopper JL. Five genetic variants associated with prostate cancer. N Engl J Med 2008;358:2739-2740; author reply 2741.
37. CPMC Demo Accounts. Available at: http://cpmc.coriell.org/Demo/DemoPeople.aspx. Accessed June 15, 2010.
38. Attia J, Ioannidis JP, Thakkinstian A, et al. How to use an article about genetic association: C: What are the results and will they help me in caring for my patients? JAMA 2009;301:304-308.
39. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol 2008;37:120-132.
40. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 1998;280:1690-1691.
41. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel

WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-1847.
42. Schunkert H, Gotz A, Braund P, et al. Repeated replication and a prospective meta-analysis of the association between chromosome 9 p21.3 and coronary artery disease. Circulation 2008;117:1675-1684.
43. Myers RH, Kiely DK, Cupples LA, Kannel WB. Parental history is an independent risk factor for coronary artery disease: the Framingham Study. Am Heart J 1990;120:963-969.
44. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180-187.
45. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008;359:2208-2219.
46. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068-1074.
47. Allen KJ, Gurrin LC, Constantine CC, et al. Iron-overload-related disease in HFE hereditary hemochromatosis. N Engl J Med 2008;358:221-230.
48. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. $N$ Engl J Med 2008;358:910-919.
49. Giovannucci E, Liu Y, Platz EA, Stampfer MJ, Willett WC. Risk factors for prostate cancer incidence and progression in the health professionals fol-low-up study. Int J Cancer 2007;121:1571-1578.
50. Brown KM, Macgregor S, Montgomery GW, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 2008;40: 838-840.
51. Cho E, Rosner BA, Feskanich D, Colditz GA. Risk factors and individual probabilities of melanoma for whites. J Clin Oncol 2005;23:2669-2675.
52. Conley YP, Jakobsdottir J, Mah T, et al. CFH, ELOVL4, PLEKHA1 and LOC387715 genes and susceptibility to age-related maculopathy: AREDS and CHS cohorts and meta-analyses. Hum Mol Genet 2006;15:3206-3218.
53. Smith W, Mitchell P. Family history and age-related maculopathy: the Blue Mountains Eye Study. Aust N Z J Ophthalmol 1998;26:203-206.
54. Cong R, Zhou B, Sun Q, Gu H, Tang N, Wang B. Smoking and the risk of age-related macular degeneration: a meta-analysis. Ann Epidemiol 2008;18: 647-656.
55. Taylor KE, Remmers EF, Lee AT, et al. Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. PLoS Genet 2008;4:e1000084.
56. Costenbader KH, Kim DJ, Peerzada J, et al. Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis. Arthritis Rheum 2004;50: 849-857.
57. Paling J. Strategies to help patients understand risks. BMJ 2003;327:745748.
58. Bernhardt BGS, Griffin G, Pyeritz R, Wawak L, Zayac C, Gordon E. It's not like judgment day: public understanding of and reactions to personalized genomic risk information. Paper presented at American Society for Human Genetics Annual Meeting, Washington, DC, 2010.
59. McBride CM, Alford SH, Reid RJ, Larson EB, Baxevanis AD, Brody LC. Putting science over supposition in the arena of personalized genomics. Nat Genet 2008;40:939-942.
60. Wade CH, McBride CM, Kardia SL, Brody LC. Considerations for designing a prototype genetic test for use in translational research. Public Health Genomics 2010;13:155-165.
61. McBride CM, Alford SH, Reid RJ, Larson EB, Baxevanis AD, Brody LC. Characteristics of users of online personalized genomic risk assessments: implications for physician-patient interactions. Genet Med 2009;11:582-587.
62. Cupples LA, Farrer LA, Sadovnick AD, Relkin N, Whitehouse P, Green RC. Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. Genet Med 2004;6:192-196.
63. Biesecker LG, Mullikin JC, Facio FM, et al. The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. Genome Res 2009;19:1665-1674.
64. Moxey A, O’Connell D, McGettigan P, Henry D. Describing treatment effects to patients. J Gen Intern Med 2003;18:948-959.
65. Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM. The framing effect of relative and absolute risk. $J$ Gen Intern Med 1993;8:543-548.
66. Carling CL, Kristoffersen DT, Montori VM, et al. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. PLoS Med 2009;6:e1000134.


[^0]:    From the Coriell Institute for Medical Research, Camden, New Jersey.
    Catharine B. Stack, PhD, Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103. E-mail: cstack@coriell.org.
    Disclosure: The authors declare no conflict of interest.
    Submitted for publication June 29, 2010.
    Accepted for publication October 9, 2010.
    Published online ahead of print January 12, 2011
    DOI: 10.1097/GIM.0b013e318201164c

