ARTICLE

Using Lifetime Risk Estimates in Personal Genomic Profiles: Estimation of Uncertainty

Quanhe Yang,[1,*] W. Dana Flanders,[3] Ramal Moonesinghe,[2] John P.A. Ioannidis,[4,5] Idris Guessous,[3] and Muin J. Khoury[1]

Personal genome tests are now offered direct-to-consumer (DTC) via genetic variants identified by genome-wide association studies (GWAS) for common diseases. Tests report risk estimates (age-specific and lifetime) for various diseases based on genotypes at multiple loci. However, uncertainty surrounding such risk estimates has not been systematically investigated. With breast cancer as an example, we examined the combined effect of uncertainties in population incidence rates, genotype frequency, effect sizes, and models of joint effects among genetic variants on lifetime risk estimates. We performed simulations to estimate lifetime breast cancer risk for carriers and noncarriers of genetic variants. We derived population-based cancer incidence rates from Surveillance, Epidemiology, and End Results (SEER) Program and comparative international data. We used data for non-Hispanic white women from 2003 to 2005. We derived genotype frequencies and effect sizes from published GWAS and meta-analyses. For a single genetic variant in *FGFR2* gene (rs2981582), combination of uncertainty in these parameters produced risk estimates where upper and lower 95% simulation intervals differed by more than 3-fold. Difference in population incidence rates was the largest contributor to variation in risk estimates. For a panel of five genetic variants, estimated lifetime risk of developing breast cancer before age 80 for a woman that carried all risk variants ranged from 6.1% to 21%, depending on assumptions of additive or multiplicative joint effects and breast cancer incidence rates. Epidemiologic parameters involved in computation of disease risk have substantial uncertainty, and cumulative uncertainty should be properly recognized. Reliance on point estimates alone could be seriously misleading.

Introduction

Recent genome-wide association studies (GWAS) have identified several hundreds of genetic variants associated with many common diseases and traits, and the list is likely to grow rapidly.[1] These newly detected genetic variants are relatively common in the general population, but the associated relative risks for diseases are usually small or moderate.[2,3] At the present time, these identified genetic variants account for a small portion of the individual variation in disease risks, and many more variants remain to be discovered to account for the residual genetic "dark matter."[1] Nevertheless, the identification of common disease susceptibility variants through GWAS has opened the possibility to develop more personalized approaches, e.g., through personal genomic profiles, for risk assessment and common disease prevention.[4–12] By the end of 2008, more than 30 companies were offering direct-to-consumer (DTC) genetic tests for health-related and nonhealth applications in many countries and a few offer whole-genome scans.[13,14] These companies provide consumers with the estimated individual risks for a range of diseases/conditions based on a panel of genetic variants that have been discovered from GWAS and candidate gene studies.

However, before the findings of GWAS are used in everyday practice, many issues need to be studied and many knowledge gaps need to be filled.[3,15–20] Estimation of the lifetime risk of developing a common disease is a part of the clinical validity of genetic testing that examines the accuracy with which a genetic test predicts a clinical outcome.[21] Lifetime risk assessments of anticipated disease occurrence are population dependent and sensitive to the uncertainties and variations in the baseline incidence of disease, genotype frequencies, and risk associated with the genetic variants in different populations, as well as the way that gene-gene and gene-environment risk factors act interactively.[2,13,16,18] Without careful consideration of these issues, lifetime risk estimates could be seriously misleading. However, the companies offering DTC genetic testing provided no information on the extent of uncertainties accompanying these risk estimates.

We quantify the effects of several epidemiologic parameters on the amount of uncertainties in risk estimates for breast cancer (MIM 114480) by using a single or a panel of genetic variants. We also discuss the public health implications of this uncertainty. The issues we discuss are readily applicable to other common diseases of public health significance.

[1]Office of Public Health Genomics, [2]Office of Minority Health, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA; [3]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA; [4]Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina 45110, Greece; [5]Tufts Clinical and Translational Science Institute and Center for Genetic Epidemiology and Modeling, Tufts Medical Center, and Department of Medicine, Tufts University School of Medicine, Boston, MA 02155, USA
*Correspondence: qay0@cdc.gov
DOI 10.1016/j.ajhg.2009.10.017.

Material and Methods

Estimations of Age-Specific Incidence Rate among Carriers of Genetic Variants

Accurate estimates of age-specific incidence rates and lifetime risk associated with disease-susceptibility genetic variants are ideally derived from properly designed and conducted prospective cohort studies that could take decades and involve substantial costs.[22,23] If such information is not available, one may use population-based disease registries and case-control studies to estimate the age-specific incidence rates and lifetime risk associated with disease-susceptibility genetic variants.[24–29] First, one needs to estimate the age-specific incidence rates among carriers of genetic variants. In brief, for a single genetic variant (SNP), the age-specific incidence rates among carriers can be estimated by Bayes' theorem:

$$P_k(D \mid G) = \frac{P_k(G \mid D)P_k(D)}{[P_k(G \mid D)P_k(D) + P_k(G \mid \overline{D})(1 - P_k(D))]} \quad (1)$$

where $P_k(D|G)$ is the annual risk of disease among people with a genetic variant in the kth age interval, $k = 1,2,3,\ldots n$; $P_k(G|D)$ and $P_k(G|\overline{D})$ are the prevalence of genetic variant among case and control subjects, respectively; and $P_k(D)$ is population incidence rate in the kth age interval.

If estimates of risk ratio for disease among carriers of the genetic variant and the prevalence of risk genotype in the population are available, the age-specific incidence rate among carriers can also be obtained by:

$$P_k(D \mid G) = \frac{\psi_k P_k(D)}{[P_k(G)(\psi_k - 1) + 1]} \quad (2)$$

where ψ_k is risk ratio for disease among carriers of genetic variant in the k_{th} age interval and $P_k(G)$ is the prevalence of the genetic variant in a population. One can also estimate the age-specific incidence rate among carriers of multiple genetic variants (Appendix). The age-specific incidence rate among noncarriers of genetic variant might be obtained by: $P_k(D|\overline{G}) = P_k(D)/[\psi_k P_k(G) + (1 - P_k(G))]$.

Lifetime Risk Models

Lifetime risk of developing breast cancer is derived with life-table methods adjusted for the competing risk of deaths. Estimates of the lifetime risk are computed on the basis of population-based age-specific breast cancer incidence rates and all-causes (excluding breast cancer) mortality rates. Detailed descriptions of the methodology have been published elsewhere.[28–30] In the estimates of lifetime risk among the carriers of the genetic variants, we have replaced the population-based age-specific incidence rates with our estimated age-specific incidence rates as described in the previous section.

With the above proposed approach, lifetime risk for people with a specific genetic profile can be estimated if values of the population disease incidence rates, mortality rates, genotype risk ratio, genotype frequency, and the joint effects of multiple genetic variants are specified. Uncertainty and variation in these parameters (sensitive parameters) would result in uncertainty in the lifetime risk estimates. Among these sensitive parameters, the genotype risk ratio and genotype frequency were derived from the GWAS that involved the uncertainty in their estimates. We defined their effects on lifetime risk estimates as the uncertainty effects. We defined the changes in population incidence rates and the assumed mode of joint effects of multiple genetic variants on lifetime risk estimates as the variation effects because the incidence rates might vary by populations and the joint effect of multiple genetic variants vary by the assumption of mode of interaction.

Uncertainty and Variation of the Sensitive Epidemiologic Parameters

Incidence of Breast Cancer and Mortality Rates. For simplicity of illustration, we focused our analysis on white populations. For the breast cancer incidence rates, we used the age-specific incidence rates and mortality rates of non-Hispanic white women in 2003–2005 U.S. derived from DevCan software (version 6.3.1). DevCan takes cross-sectional counts of cancer incident cases from the Surveillance, Epidemiology, and End Results (SEER) Program conducted by the National Cancer Institute and mortality counts from data collected by the National Center for Health Statistics/CDC. We focused our analysis on the changes in breast cancer incidence rates and assumed that the mortality rates other than breast cancer were the same among the carriers of genetic variants as that in the general population.

The incidence of breast cancer varies greatly in countries of majority white populations, from 43 in Eastern Europe to 125 per 100,000 women in the U.S.[31,32] To examine the effect of changes in breast cancer incidence rates on lifetime risk estimates, we took the age-specific incidence rates of non-Hispanic white women in the U.S. as the high end and downward adjusted the incidence rates by 3-fold (to approximately match the incidence rate in Eastern Europe) as the low end; this defined the variation range of breast cancer incidence rates.

Relative Risks. Many GWAS have reported only allelic odds ratios. For simplicity of illustration, we calculated and used the dominant genotype odds ratio as derived from the reported risk allele frequency and allelic odds ratio. We assumed Hardy-Weinberg equilibrium (HWE) for both controls and cases to convert allelic odds ratios to genotype odds ratios. Even though HWE might not hold for cases in reality, we made this assumption to illustrate our method.[33] The odds ratio from case-control association studies is used as a proxy of the population-level risk ratio. For more common conditions, one could apply the simple method to correct the odds ratio to better approximate risk ratio in population.[34] We selected five SNPs that were robustly associated with breast cancer risk and replicated by GWAS and subsequent replication studies in multiple populations: SNP (rs2981582) in *FGFR2* gene (MIM 176943) on chromosomes 10q26, SNP (rs3803662) in *TNRC9* gene (TOX3 [MIM 611416]) on 16q12, SNP (rs13387042) on 2q35, SNP (rs889312) in *MAP3K1* gene (MIM 600982) on 5q11, and SNP (rs3817198) in *LSP1* gene (MIM 153432) on 11p15.[35–37]

We used random effects models to calculate the summary odds ratio and 95% prediction interval (PI) based on the studies with available data for each of the five SNPs.[38] The meta-analyses included 19 studies of white populations derived from Easton et al. and Stacey et al. studies' supplementary information.[35,37] Data in four studies on rs3803662 in *TNRC9* were missing in Easton et al.,[35] and we also included five studies on rs3803662 in *TNRC9* from the Stacey et al. study (total 20 studies).[37] For SNP rs13387024 on 2q35, we included five studies from the Stacey et al. study.[37] These case-control studies included the population-based, hospital-based, and convenient samples. To properly account for the heterogeneity in relative risk estimates across the populations, we calculated the 95% PI that provided the most appropriate measure of uncertainty of the summary odds ratio estimates, especially when estimates are to be extrapolated to people beyond those tested in the original research study; this is exactly the case when a test is going to be used in the wider

population.[38] PI is calculated by taking into account the variance among the studies, τ^2, and is given by:

$$\mu \pm t^{\alpha}_{k-2}\sqrt{\tau^2 + SE(\mu)^2}$$

where t^{α}_{k-2} is the $100(1-\alpha/2)$ percentile of the t distribution with k-2 degrees of freedom.[38] However, the test for heterogeneity of the odds ratio indicated that the point estimate of $\tau^2 = 0$ for all five SNPs included in our study, so the 95% PI approximates 95% CI (any difference is due only to the difference of the t value with k-2 degrees of freedom versus the z distribution value for the $(1-\alpha/2)$ percentile).[39] We used the lower and upper 95% PI as the uncertainty range of the effect size of the odds ratio. The 95% PI shows the range within which the true odds ratio is likely to lie in 95 of 100 populations similar to those where data are already available. Even this range is already based on a conservative assumption, because it assumes that the populations genotyped in the case-control studies that have identified these associations were representative of all the white populations. If we assume $\tau^2 = 0.05$ (a low) and $\tau^2 = 0.25$ (a modest heterogeneity) that typically cannot be excluded by the 95% CIs of the τ^2 for SNPs arising from GWAS investigations, then the PI would be considerably wider (not shown in detail; available upon request).

Genotype Frequencies. The GWAS publications did not contain detailed information to estimate the dominant genotype frequencies for each specific study.[35,37] We derived the dominant genotype frequencies and their 95% confidence intervals (CI) of SNP (rs2981582) in the *FGFR2* gene, SNP (rs3803662) in the *TNRC9* gene, SNP (rs889312) in the *MAP3K1* gene, and SNP (rs3817198) in the *LSP1* gene from the Carcia-Closas et al. study.[40] The study provided the pooled genotype frequencies of four genetic variants from 17 case-control studies of white populations with more than 23,000 control subjects. For SNP (rs13387042) on 2q35, we derived the genotype frequency from the Antoniou et al. study of common genetic variants and breast cancer risk.[41] The study provided the pooled genotype frequency from 33 study centers of white populations of *BRCA1* and *BRCA2* mutation carriers. We assumed the independence between *BRCA1/BRCA2* mutation and the SNP on 2q35 and used the unaffected (breast cancer) carriers (n = 4268) to derive the dominant genotype frequency and 95% CI. Neither study provided detailed information for each specific study to calculate 95% PI. We used the lower and upper 95% CI as the uncertainty range of genotype frequency.

Joint Effects of Multiple Risk Variants. For the panel of five genetic variants, we considered three scenarios: (1) women who carry no risk genotype; (2) women who carry three risk genotypes (SNP [rs2981582] in *FGFR2* gene, SNP [rs3803662] in *TNRC9* gene, and SNP [rs13387042] on 2q35); and (3) women who carry all five genetic variants. We calculated the joint effect of three or five SNPs on additive scale as: $R_{add} = R_1 + R_2 + R_3 - 2$, or $R_{add} = R_1 + R_2 + R_3 + R_4 + R_5 - 4$, where R_1, R_2,…,R_5, are risk ratio of disease for subjects with risk genotype compared with subjects without risk genotype,[42,43] and we repeated the same calculation with the lower and upper 95% PI of risk ratio for each genetic variant as the lower and higher risks range of joint effect of three or five SNPs on additive scale. We defined this range of risk ratios as the variation range of joint effect on the additive scale. For the joint effect on the multiplicative scale, we estimated $R_{mult} = R_1{*}R_2{*}R_3$ or $R_{mult} = R_1{*}R_2{*}R_3{*}R_4{*}R_5$,[42,43] estimated the lower and upper risks of joint effect as for the additive model, and defined them as the variation range on multiplicative scale. The five SNPs have no linkage disequilibrium among them, so independence of effects is reasonable.

Uncertainty Analysis

For uncertainty analysis, we started with one SNP in the *FGFR2* gene (rs2981582) that is the most significantly associated SNP with breast cancer risk. It also has the largest effect size among the five SNPs we considered.[35] First, we estimated age-specific incidence rates and lifetime risk of developing breast cancer among *FGFR2* carriers and noncarriers. Then, we examined separately the effect of changes in each sensitive parameter one at a time, e.g., changes in breast cancer incidence rates, genotype risk ratio, and frequency followed by the combination of changes in these parameters on lifetime risk estimate. For examining the effect of change in each sensitive parameter on lifetime risk, we held the other parameters at their point estimates. For example, to examine the effect of changes in the breast cancer incidence rate, we used formula (2) by keeping the genotype risk ratio and frequency at their point estimates and unchanged, and first, used the age-specific breast cancer incidence rates of U.S. non-Hispanic women in 2003–2005 to estimate the age-specific incidence rates and the lifetime risk for the carriers of genetic variant and noncarriers. Then we replaced the U.S. non-Hispanic women's rates with the 3-fold downward-adjusted age-specific breast cancer incidence rates, and the difference between these two estimates was documented. We repeated the similar analysis for the genotype risk ratio and frequency. However, because we used the pooled estimate across multiple studies for the genotype frequencies, the 95% CI is too narrow to demonstrate any substantial impact of the genotype frequency on risk estimate. Ideally, one would use data from each study and combine these frequencies with random effects to calculate the 95% PI covering the full range of uncertainty. To further examine the impact of uncertainty of genotype frequency and risk ratio (we also included risk ratio as a comparison of the relative impact of changes in genotype frequency and risk ratio on lifetime risk), we assumed that the genotype frequency and risk ratio were 10% and 20% lower or higher than the point estimate and documented the changes in lifetime risk estimates among the carriers of *FGFR2* genetic variant.

For the multiple genetic variants, we used formulas (1) and (2) in the Appendix to estimate the effects of additive versus multiplicative joint effects of different combination of genetic variants on the age-specific incidence rates and lifetime risk of developing breast cancer.

To account for the uncertainties and variations in the sensitive epidemiologic parameters, we used the Monte Carlo simulation methods.[44,45] We assumed that the observed age-specific breast cancer incidences and mortality rates follow the Poisson distribution (when the observed breast cancer incidences or the deaths by age group were large [n > 1000], and we used the normal distribution to approximate the Poisson distribution). For genotype risk ratio and frequency, we assumed a normal distribution with the mean being the point estimate and standard deviation given by the difference of the lower (L) and upper (U) 95% boundaries divided by 3.92, because of the relatively large sample size for many GWA studies.[35–37] For joint effects of multiple SNPs, we calculated the risk ratios of joint effect on additive or multiplicative scales as defined above. We took the lower 95% PI of risk ratios or the upper 95% PI of different combination of multiple SNPs as the lower 95% CI or upper 95% CI of the joint risk ratio on additive or multiplicative scale, respectively. We assumed a normal distribution with the mean being the point estimate and standard

Table 1. Dominant Genotype Frequencies and Odds Ratios for Association between Five Selected SNPs and Breast Cancer Risk in White Populations

Locus	SNP	Dominant Genotype Frequency[a] %	95% CI	Dominant Genotype OR[b]	95% PI[c]
FGFR2	rs2981582	61.7	61.1–62.3	1.35	1.29–1.41
TNRC9	rs3803662	45.8	45.2–46.5	1.28	1.23–1.33
2q35	rs13387042	75.9	74.6–77.2	1.28	1.14–1.43
MAP3K1	rs889312	47.6	46.9–48.2	1.15	1.10–1.20
LSP1	rs3817198	51.7	51.0–52.3	1.10	1.06–1.15

Meta-analyses included 19 studies of white populations derived from the Easton et al. and Stacey et al. studies.[37,39] Data in four studies on rs3803662 in *TNRC9* was missing in the Easton et al. study,[37] so we included additional five studies from Stacy et al. study.[39] Meta-analysis of SNP rs13387024 on 2q35 included five studies from the Stacey et al. study.[39] Abbreviations: CI, confidence interval; PI, prediction interval.

[a] Dominant genotype frequencies of the five selected SNPs were derived from the Carcia-Closas et al. and Antoniou et al. studies.[42,43]

[b] Dominant genotype odds ratio (OR) was calculated with the meta-analysis of the fixed effect models.

[c] Prediction interval was calculated via the Higgins et al. approximation method.[40]

deviation given by the difference of the lower (L) and upper (U) 95% boundaries divided by 3.92.

The Crystal Ball software (version 2000.2, Decisioneering, Inc., Denver, CO) was used for the Monte Carlo uncertainty analysis with 10,000 draws from the above defined distributions for the model parameters. We reported the lower 2.5 and upper 97.5 percentiles of the total simulation distribution as 95% "simulation interval" (SI).[46]

Results

Table 1 presents the dominant genotype frequencies (95% CI) and risk ratios (OR) and 95% PIs for the association between five selected SNPs and breast cancer risk estimated from multiple studies of white populations.[35,37,40,41] The dominant genotype risk ratio ranged from 1.10 (95% PI 1.06–1.15) for rs3817198 in *LSP1* to 1.35 (95% PI 1.29–1.41) for rs2981582 in *FGFR2*. The frequency of carriers ranged from 45.8% (95% CI 45.2–46.5) for rs3803662 in *TNRC9* to 75.9% (95% CI 74.6–77.2) for rs13387042 on 2q35.

Lifetime Risk Associated with a Single Genetic Variant

Table 2 lists the data and the estimated age-specific incidence rates and lifetime risk (and 95% simulation interval [SI]) of developing breast cancer from birth among carriers of *FGFR2* variant, noncarriers, and in the overall population. Assuming constant genotype frequency and risk ratio across age group, the estimated lifetime risk of developing breast cancer from birth among carriers of *FGFR2* variant and noncarriers by age 80 years was about 2.3% higher and 1.8% lower than the average risk in population (15.7% and 11.6% versus 13.4%), respectively.

Holding other parameters at their point estimates, the lifetime risk of developing breast cancer from birth to age 80 years reduced from 15.7% to 5.3% among carriers of *FGFR2* variant, and from 11.6% to 3.9% among noncarriers, respectively, if the population incidence rates were 3-fold lower (Figure 1A). Changes in genotype risk ratio (lower or upper 95% PI) in the population have limited impact on lifetime risk estimates (Figure 1B) and changes in genotype frequency (lower or upper 95% CI) have negligible impact because of the narrow 95% CI (Figure 1C). However, changes in combinations of these parameters resulted in 5.2% to 15.9% lifetime risk of developing breast cancer from birth to before age 80 years among carriers of *FGFR2* variant, and 3.8% to 12.1% lifetime risk among noncarriers (Figure 1D). Difference in population breast cancer incidence rates was the largest contributor of variation in lifetime risk estimates.

A 10% assumed variation in the genotype frequency of *FGFR2* genetic variant (55.5% to 67.9%) produced a 0.5% difference in lifetime risk from birth to before age 80 years, and a 20% variation (49.4% to 74.0%) produced a 1.0% difference (Figure 2A). A 10% (1.22 to 1.42) or 20% (1.08 to 1.62) variation in genotype risk ratio produced 1.0% or 2.0% difference in lifetime risk from birth to before age 80 years, respectively (Figure 2B).

Lifetime Risk among Women with Multiple Genetic Variants

Approximately 22% or 5% of non-Hispanic white women were carriers of three or all five selected genetic variants, respectively. The estimated additive joint effect of three genetic variants was 1.9 (95% CI 1.7–2.2) versus 2.2 (95% CI 1.8–2.7) for the multiplicative model. The corresponding comparison was 2.2 (95% CI 1.8–2.5) versus 2.8 (95% CI 2.1–3.7) for five genetic variants. Within the same incidence rates, both age-specific incidence rates and lifetime risk from birth were substantially higher assuming multiplicative joint effect of multiple genetic variants than that of additive joint effects (Table 3). The estimated lifetime risk of developing breast cancer from birth to before age 80 years among women who carry three genetic variants ranged from 5.8% (95% SI 5.5–6.1) assuming additive joint effect and 3-fold lower breast cancer incidence rates to 18.7% (95% SI 17.3–20.1) assuming multiplicative joint effects of multiple genetic variants and U.S. non-Hispanic women breast cancer incidence rates in 2003–2005. The corresponding lifetime risk estimates ranged from 6.1% (95% SI 5.7–6.5) to 21.0% (95% SI 18.8–23.3) among women who carry all five genetic variants (Table 3).

Discussion

There are many issues to be considered in the evaluation of personal genomic profiles including analytical and clinical validity, clinical utility, and ethical, legal, social, and policy implications.[3,10,12,13,15–20,47] Risk assessment is part of clinical validity studies in genetic testing. Our results suggest that extreme caution is needed when using genetic

Table 2. Lifetime Risk of Developing Breast Cancer from Birth among Carriers of *FGFR2* Variants, Noncarriers, and in General Population

				FGFR2 Variant Carriers		Noncarriers		U.S. Non-Hispanic White Women in 2003–2005
Age, yrs	Breast Cancer Incidence Rates, per 100,000	Genotype Frequency in Population (%) (95% CI)	Dominant Risk Ratio (95% PI)[a]	Age-Specific Incidence Rates, (95% SI)[b] (per 100,000)	Lifetime Risk % (95% SI)[b]	Age-Specific Incidence Rates (95% SI)[b] (per 100,000)	Lifetime Risk % (95% SI)[b]	Lifetime Risk, % (95% CI)[c]
0–19	0.05	61.7 (61.1–62.3)	1.35 (1.29–1.41)	0.1 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.0)
20–24	1.6	61.7 (61.1–62.3)	1.35 (1.29–1.41)	1.8 (1.3–2.2)	0.0 (0.0–0.0)	1.3 (1.0–1.7)	0.0 (0.0–0.0)	0.0 (0.0–0.0)
25–29	8.5	61.7 (61.1–62.3)	1.35 (1.29–1.41)	9.4 (8.3–10.5)	0.1 (0.0–0.1)	6.9 (6.1–7.8)	0.0 (0.0–0.1)	0.1 (0.1–0.1)
30–34	30.1	61.7 (61.1–62.3)	1.35 (1.29–1.41)	33.4 (31.5–35.4)	0.2 (0.2–0.2)	24.7 (23.2–26.4)	0.2 (0.2–0.2)	0.2 (0.2–0.2)
35–39	71.1	61.7 (61.1–62.3)	1.35 (1.29–1.41)	79.0 (75.9–82.0)	0.6 (0.6–0.6)	58.5 (55.8–61.3)	0.5 (0.4–0.5)	0.6 (0.6–0.6)
40–44	157.2	61.7 (61.1–62.3)	1.35 (1.29–1.41)	174.6 (169.8–179.3)	1.5 (1.4–1.5)	129.3 (124.5–134.4)	1.1 (1.0–1.1)	1.4 (1.3–1.4)
45–49	246.4	61.7 (61.1–62.3)	1.35 (1.29–1.41)	273.6 (267.1–280.1)	2.8 (2.7–2.9)	202.7 (195.5–210.0)	2.1 (2.0–2.1)	2.5 (2.5–2.6)
50–54	298.5	61.7 (61.1–62.3)	1.35 (1.29–1.41)	331.5 (323.9–338.9)	4.4 (4.3–4.5)	245.5 (237.1–254.5)	3.3 (3.2–3.3)	4.0 (3.9–4.0)
55–59	363.7	61.7 (61.1–62.3)	1.35 (1.29–1.41)	403.8 (394.7–412.8)	6.3 (6.2–6.4)	299.0 (288.6–310.0)	4.7 (4.6–4.8)	5.6 (5.6–5.7)
60–64	439.1	61.7 (61.1–62.3)	1.35 (1.29–1.41)	487.5 (476.3–498.7)	8.6 (8.5–8.7)	361.1 (348.3–374.1)	6.4 (6.3–6.5)	7.6 (7.5–7.7)
65–69	497.3	61.7 (61.1–62.3)	1.35 (1.29–1.41)	552.1 (539.2–565.2)	11.1 (11.0–11.2)	409.0 (394.7–424.1)	8.2 (8.1–8.3)	9.7 (9.6–9.8)
70–74	507.6	61.7 (61.1–62.3)	1.35 (1.29–1.41)	563.6 (550.0–577.1)	13.4 (13.3–13.6)	417.5 (402.6–433.5)	10.0 (9.8–10.1)	11.7 (11.6–11.7)
75–79	523.8	61.7 (61.1–62.3)	1.35 (1.29–1.41)	581.5 (567.7–595.3)	15.7 (15.5–15.8)	430.8 (415.3–446.8)	11.6 (11.5–11.8)	13.4 (13.3–13.5)
80+	429.1	61.7 (61.1–62.3)	1.35 (1.29–1.41)	476.4 (466.2–487.0)	16.8 (16.7–16.9)	352.9 (340.9–365.6)	12.5 (12.3–12.6)	15.8 (15.7–15.9)

Estimates of lifetime risk of developing breast cancer from birth among carriers and noncarriers of *FGFR2* genetic variant were based on the age-specific breast cancer incidence rates of U.S. non-Hispanic white women in 2003–2005. Abbreviations: CI, confidence interval; PI, prediction interval; SI, simulation interval.

[a] Prediction interval was calculated via the Higgins et al. approximation method.[40]

[b] Simulation interval was based on 10,000 Monte Carlo simulations. We assumed that the breast cancer incidence cases and number of deaths followed the Poisson distribution (when the number of cases were large [$n > 1000$], normal distribution was used to approximate Poisson distribution), genotype frequency, and dominant risk ratio followed normal distribution with mean as point estimate and standard deviation as the difference between upper and lower 95% confidence interval (CI) or 95% prediction interval (PI) divided by 3.92.

[c] Lifetime risk of developing breast cancer from birth among U.S. non-Hispanic white women derived from DevCan software (version 6.3.1).

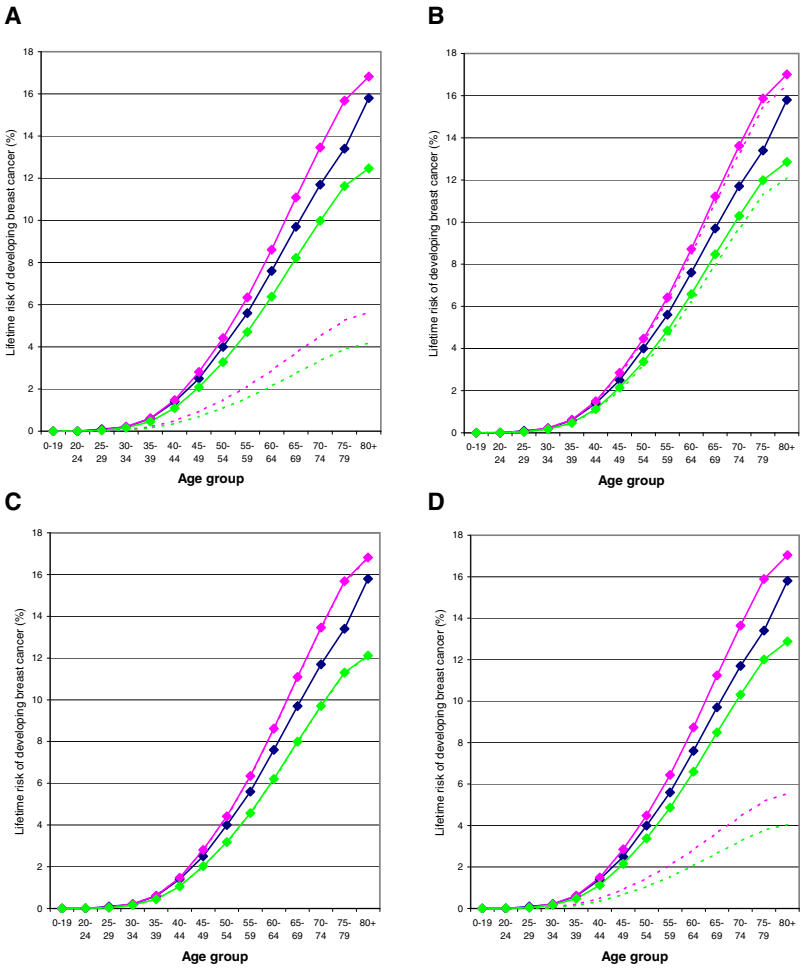

Figure 1. Effect of Epidemiologic Parameters on Lifetime Risk of Developing Breast Cancer from Birth among Carriers and Noncarriers of *FGFR2* Variant

Figures show the effect of breast cancer incidence rates, risk ratio, genotype frequency, and combinations of these parameters on lifetime risk of developing breast cancer from birth among carriers and noncarriers of *FGFR2* (rs2981582) genetic variant.

(A) Effect of 3-fold lower breast cancer incidence rates; blue solid line with squares indicates lifetime risk of developing breast cancer among non-Hispanic white women; red solid line with squares indicates lifetime risk of developing breast cancer among non-Hispanic white women who carried *FGFR2* genetic variant; red dashed line indicates lifetime risk of developing breast cancer among carriers of *FGFR2* genetic variant assuming a 3-fold lower breast cancer incidence rates; green solid line with squares indicates lifetime risk of developing breast cancer among noncarriers of non-Hispanic white women; green dashed line indicates lifetime risk of developing breast cancer among noncarriers assuming a 3-fold lower breast cancer incidence rates.

(B) Effect of lower and upper 95% prediction interval (PI) of risk ratio; blue solid line with squares indicates lifetime risk of developing breast cancer among non-Hispanic white women; red solid line with squares indicates effect of using upper 95% PI risk ratio on lifetime risk among non-Hispanic white women who carried *FGFR2* genetic variant; red dashed line indicates effect of using lower 95% PI risk ratio on lifetime risk among non-Hispanic white women who carried *FGFR2* genetic variant; green solid line with square indicates effect of using upper 95% PI risk ratio on lifetime risk among noncarriers of non-Hispanic white women; green dashed line indicates effect of using lower 95% PI risk ratio on lifetime risk among noncarriers of non-Hispanic white women.

(C) Effect of lower and upper 95% confidence interval (CI) of genotype frequency; blue solid line with square indicates lifetime risk of developing breast cancer among non-Hispanic white women; red solid line with squares indicates effect of using lower 95% CI genotype frequency on lifetime risk among non-Hispanic white women who carried *FGFR2* genetic variant; red dashed line indicates effect of using upper 95% CI genotype frequency on lifetime risk among non-Hispanic white women who carried *FGFR2* genetic variant; green solid line with squares indicates effect of using lower 95% CI genotype frequency on lifetime risk among noncarriers of non-Hispanic white women; green dashed line indicates effect of using upper 95% CI genotype frequency on lifetime risk among noncarriers of non-Hispanic white women.

(D) Effect of combination of these parameters; blue solid line with square indicates lifetime risk of developing breast cancer among non-Hispanic white women; red solid line with square indicates combination effect of lower 95% CI genotype frequency and upper 95% PI risk ratio on lifetime risk among non-Hispanic white women who carried *FGFR2* genetic variant; red dashed line indicates combination effect of upper 95% CI genotype frequency and lower 95% PI risk ratio on lifetime risk among carriers of *FGFR2* genetic variant assuming a 3-fold lower breast cancer incidence rates; green solid line with square indicates combination effects of lower 95% CI genotype frequency and lower 95% PI risk ratio on lifetime risk among noncarriers of non-Hispanic white women; green dashed line indicates combination effects of upper 95% CI genotype frequency and upper 95% PI risk ratio on lifetime risk among noncarriers assuming a 3-fold lower breast cancer incidence rates.

variants to estimate disease risks for common complex diseases, because the cumulative uncertainty and variation of several epidemiologic parameters that influence the absolute disease risk can have a major impact on risk estimation.

A potentially detrimental scenario would provide consumers with risk estimates that use inappropriate population incidence rates and relative risks associated with genetic variants. Risk assessment among carriers and noncarriers could vary substantially depending on the uncertainties and variations in the combinations of population parameters. In particular, variations in population incidence rates can play an important role in lifetime risk estimates if there are marked differences across populations and ethnic groups. In our breast cancer example, 3-fold differences are documented even within white populations.[31,32] We did not consider nonwhite populations, but the breast cancer incidence rates in Asian countries such as China and Japan are less than one-fifth or one-fourth that of the U.S.[31,32] Extrapolation to populations of other ancestries would thus warrant further adjustments. We used breast cancer as an example in our

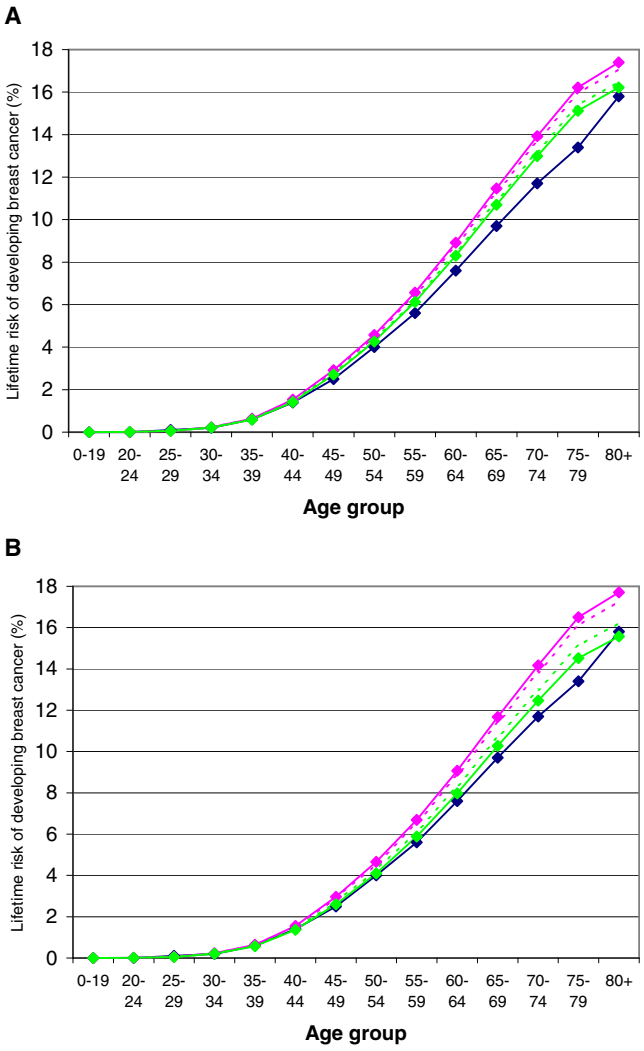

Figure 2. Effects of Varying Genotype Frequency and Risk Ratio on Lifetime Risk of Developing Breast Cancer from Birth among Carriers of *FGFR2* Variant

Figures show the effect of assuming 10% or 20% lower or higher values than the point estimates of genotype frequency and risk ratio of *FGFR2* (rs2981582) genetic variant on lifetime risk of developing breast cancer from birth among carriers of U.S. non-Hispanic white women in 2003–2005.

(A) Effect assuming 10% or 20% lower or higher genotype frequency; blue solid line with squares indicates lifetime risk of developing breast cancer among non-Hispanic white women; red solid line with squares indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 20% lower genotype frequency; red dashed line indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 10% lower genotype frequency; green solid line with squares indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 20% higher genotype frequency; and green dashed line indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 10% higher genotype frequency.

(B) Effect assuming 10% or 20% lower or higher genotype risk ratio; blue solid line with squares indicates lifetime risk among non-Hispanic white women; red solid line with squares indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 20% higher genotype risk ratio; red dashed line indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 10% higher genotype risk ratio; green solid line with squares indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 10% lower genotype risk ratio; and green dashed line indicates lifetime risk among carriers of *FGFR2* genetic variant assuming 20% lower genotype risk ratio.

analyses and assumed 3-fold differences in population incidence rates; it is not surprising to find that difference in population incidence rates was the largest contributor to variation in lifetime risk estimates. Many common diseases might have larger variations in population incidence rates, such as diabetes (International Diabetes Federation), and others might have less variation among different populations.[48] Other studies also suggested that the effects of genetic variants might differ by the pathological subtype of breast cancer.[40] This might be the case for other common diseases, adding more uncertainty in using appropriate population incidence rates and genotype risk ratios. Therefore, it is crucial that the appropriate population incidence rates are used in lifetime risk estimates. In addition, the rates of many common diseases might change greatly over time. For example, the incidence of diabetes increased 41% from 4.9 to 6.9 per 1000 population from 1997 to 2003 in the U.S.,[49] and could increase further in the future,[50] making risk estimates based on past data obsolete. SEER is a well-established population-based cancer registry that provides reliable estimates for many cancers. For many other diseases, population incidence rates may be unreliable or even unavailable.

Although the lifetime risk of developing breast cancer among carriers of multiple genetic variants increased compared with the risk in the general population, the assumed mode of joint effect among multiple genetic variants (additive versus multiplicative) also plays an important role in lifetime risk estimates. It remains unknown on what scale multiple genetic variants and environment risk factors interact to affect the risk for common diseases. Moreover, current studies do not have adequate statistical power to detect interactions, i.e., further significant deviations beyond these models.[51] Many studies argued for additive joint effects and others suggested that the multiple genetic variants interact multiplicatively.[52–55] Differences up to 3% can arise from consideration of multiplicative versus additive models alone among carriers of five genetic variants in our study.

However, compared to the variations in the population incidence rates and the assumed mode of joint effect of multiple genetic variants, the impact of uncertainties in the genotype risk ratio and frequency have little impact on lifetime risk estimates given that the risk associated with each SNP is moderate (RR < 1.5) and the genotype frequencies are common (>45%) in our study. For example, it appeared that a 10% variation around the point estimate of the common genotype frequency (SNP in *FGFR2* with a genotype frequency 61.7%) might produce ~0.5% difference, and a 10% variation in genotype risk ratio (SNP in *FGFR2* with a genotype risk ratio 1.35) is associated with ~1.0% difference in lifetime risk of developing breast cancer from birth to before age

80 years. There is no objective criteria concerning what changes might be considered substantial in lifetime risk estimates. From a practical point of view, <1.0% difference in lifetime risk might not be considered substantial. For strong-effect SNP with less common frequency in population, differences in risk ratio and genotype frequency could have significant impact on lifetime risk estimates (results not shown).[56] Presently, most GWAS have studied white populations, but the risk genotype frequencies and their associated risk might differ more for different populations.[57–61]

We used five selected SNPs for breast cancer that were confirmed by GWAS as an example in our analysis. There is preliminary evidence from some diseases such as type 1 and type 2 diabetes, hyperlipidemia, and Crohn's disease and theoretical reasons to believe that many common diseases are caused by a large number of common genetic variants that have weak effect each but high risk acting jointly.[62] It is widely accepted that one variant or even five of them typically would have limited predictive ability.[63,64] In the near future, as more of the genetic architecture of complex diseases is discovered, it should become more feasible to use many genetic variants that explain a larger share of the disease variance. One may speculate whether this would make a difference in our inferences. To examine the effects of many common genetic variants on lifetime risk estimates, we simulated a population with 100 SNPs assuming the genotype risk ratio = 1.1 and genotype frequency = 10% for each SNP. Assuming normal distribution of the number of SNPs, 95% of people would have between 4 and 16 SNPs in the simulated population.[64] By using the same incidence rates of non-Hispanic white women in 2003–2005 U.S., we calculated lifetime risks among the carriers of 10 or 20 SNPs for additive and multiplicative joint effect models, respectively. The lifetime risk of developing breast cancer from birth to before age 80 years increased from 25.5% to 34.8% among the carriers of 10 to 20 SNPs, assuming additive joint effects. It increased from 32.7% to 74.5% assuming multiplicative joint effects. It is possible to use the multiple SNPs to identify the high-risk individuals in population if used appropriately.[55,65] Nevertheless, the uncertainty and variation in crucial parameters as those that we have evaluated is unlikely to diminish and therefore substantial uncertainty around the point estimates may remain typical. The inappropriate application of lifetime risk estimate via GWAS-identified genetic variants could lead to misleading results and provide little help or even harm to consumers seeking risk assessments for common diseases.

In this report, we focused our analysis on the impact of joint effects of multiple genetic variants on lifetime risk estimates. In reality, it is important to consider the impact of gene-environment interactions. Many environment risk factors are common and have substantial larger effect size compared to the genetic variants. For example, the relative risk for breast cancer was 1.9 for women who first gave birth after age 30 years to those who gave their first birth before age 30 years, with an estimated 21% of non-Hispanic white women giving birth after age 30 years in U.S.[66,67] Depending on the mode of gene-environment interaction, the difference in lifetime risk could be substantial between those who are exposed and not exposed to environment risk factors.

We used an established and validated method to estimate the lifetime risk by using information derived from population-based disease registries and case-control studies that takes into account the competing risk.[24–29] Several companies offer genome-wide scans and provide lifetime risk estimate for the selected common diseases/conditions.[13,18] The predicted individual lifetime risk was obtained by multiplying the overall risk relative to population with the ethnicity- and region-specific average lifetime risk (Personalized Medicine Coalition). None of the companies has systematically examined the impact of multiple sources of uncertainty and variation of the epidemiologic parameters used in risk estimates, nor have any taken into account the effect of competing risk in their risk estimates. Without carefully considering the potential impact of the multiple sources of uncertainty and variation, personal genome tests could lead to inconsistent or misleading results in informing consumers on their disease risks and potential measures for disease prevention. The use of inappropriate methods would likely add additional uncertainty as well as inaccuracy in risk estimates. For example, mammography is an accepted preventive/screening measure for breast cancer, even if there is debate about its merits and indicated age of screening.[68,69] It has been argued that genetic testing might modify the indicated age of screening and might suggest that some women are at sufficiently low risk for screening not to be indicated.[65] However, this assumption relies on the ability to accurately estimate the lifetime risk, and this is not straightforward. Our analysis indicates that combined with variations in breast cancer incidence rates, uncertainties in the genotype frequencies, and the associated risk, the estimated lifetime risk of developing breast cancer can range more than 3-fold for a single genetic variant. The range of uncertainty in lifetime risk estimates is likely to be wider when considering gene-gene and gene-environment interactions.

There are some limitations to our approach. First, we assumed constant risk ratio across age groups. Nevertheless, there is no evidence for age modulation of the genetic effects. Second, we assumed that the mortality rates among carriers of genetic variants were the same as that of the general population. This assumption might be reasonable for a single GWAS genetic variant, but may be more questionable among carriers of multiple genetic variants. Simulations assuming that both total mortality and mortality rates other than cancer were 20% lower versus 20% higher than that of the population average among carriers of five genetic variants yielded relatively small differences (up to 0.7%) in the estimated lifetime risk of breast cancer (not shown in detail).

Table 3. Lifetime Risk of Developing Breast Cancer from Birth among Noncarriers and Carriers of Three and Five Genetic Variants

	Noncarriers				**Carriers of Three SNPs**				**Carriers of Five SNPs**			
	Using U.S. Breast cancer Incidence Rates of Non-Hispanic White Women in 2003–2005		**Assuming 1/3 of U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005**		**Using U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005**		**Assuming 1/3 of U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005**		**Using U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005**		**Assuming 1/3 of U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005**	
Age (years)	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**	**Age-Specific Incidence Rates (95% SI) (per 100,000)[a]**	**Lifetime Risk, % (95% SI)[a]**
Additive Model[b]												
0–19	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.1 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.1)	0.0 (0.0–0.0)	0.1 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.1)	0.0 (0.0–0.0)
20–24	1.0 (0.7–1.2)	0.0 (0.0–0.0)	0.3 (0.2–0.5)	0.0 (0.0–0.0)	2.0 (1.4–2.6)	0.0 (0.0–0.0)	0.7 (0.4–1.0)	0.0 (0.0–0.0)	2.0 (1.5–2.3)	0.0 (0.0–0.0)	0.7 (0.4–1.0)	0.0 (0.0–0.0)
25–29	5.0 (4.4–5.8)	0.0 (0.1–0.0)	1.7 (1.3–2.1)	0.0 (0.0–0.0)	10.4 (8.5–12.6)	0.1 (0.1–0.1)	3.4 (2.6–4.3)	0.0 (0.0–0.0)	10.9 (8.6–13.4)	0.1 (0.1–0.1)	3.6 (2.7–4.7)	0.0 (0.0–0.0)
30–34	17.9 (16.3–19.6)	0.1 (0.1–0.1)	5.9 (5.2–6.7)	0.0 (0.0–0.0)	36.9 (31.1–43.2)	0.2 (0.2–0.3)	12.3 (10.1–14.7)	0.1 (0.1–0.1)	39.4 (32.3–47.0)	0.2 (0.2–0.3)	12.9 (10.4–15.7)	0.1 (0.1–0.1)
35–39	42.4 (39.0–46.0)	0.3 (0.3–0.3)	14.1 (12.8–15.5)	0.1 (0.1–0.1)	87.2 (73.9–101.6)	0.7 (0.6–0.8)	29.0 (24.0–34.0)	0.2 (0.2–0.3)	91.3 (75.1–108.8)	0.7 (0.6–0.8)	30.5 (24.8–36.4)	0.2 (0.2–0.3)
40–44	93.6 (86.7–101.1)	0.8 (0.7–0.8)	31.1 (28.7–33.9)	0.3 (0.2–0.3)	193.0 (164.1–223.2)	1.6 (1.5–1.8)	64.2 (54.5–74.9)	0.5 (0.5–0.6)	201.8 (166.6–239.4)	1.7 (1.5–1.9)	67.4 (55.1–80.2)	0.6 (0.5–0.6)
45–49	146.7 (136.3–159.1)	1.5 (1.4–1.6)	48.8 (45.2–53.2)	0.5 (0.4–0.5)	302.7 (265.0–351.0)	3.1 (2.8–3.4)	100.7 (85.5–117.7)	1.1 (0.9–1.1)	316.9 (261.1–375.9)	3.3 (2.9–3.6)	105.7 (87.4–124.9)	1.1 (0.9–1.2)
50–54	177.6 (165.2–192.2)	2.4 (2.3–2.5)	59.1 (54.8–64.2)	0.8 (0.7–0.8)	366.2 (311.7–424.7)	4.9 (4.5–5.2)	121.8 (103.4–142.0)	1.6 (1.5–1.8)	383.7 (317.0–455.3)	5.1 (4.6–5.6)	127.7 (105.5–151.8)	1.7 (1.6–1.9)
55–59	216.5 (201.2–234.5)	3.4 (3.3–3.5)	72.1 (66.8–78.3)	1.1 (1.0–1.1)	445.7 (379.8–519.3)	7.0 (6.5–7.5)	148.5 (126.1–172.7)	2.3 (2.2–2.5)	466.9 (397.5–553.2)	7.3 (6.7–8.0)	155.7 (128.8–185.1)	2.5 (2.3–2.7)
60–64	261.4 (242.8–282.6)	4.6 (4.5–4.8)	87.1 (80.6–94.4)	1.5 (1.4–1.6)	539.9 (459.1–625.4)	9.5 (8.9–10.2)	179.5 (152.3–208.8)	3.2 (3.0–3.4)	564.7 (465.6–670.4)	10.0 (9.2–10.7)	187.8 (154.6–223.7)	3.3 (3.1–3.6)
65–69	296.1 (274.8–319.8)	6.0 (5.8–6.2)	98.6 (91.2–107.0)	2.0 (1.9–2.1)	610.3 (518.7–708.8)	12.2 (11.5–13.0)	202.9 (172.1–237.1)	4.1 (3.8–4.4)	640.1 (525.7–759.2)	12.8 (11.9–13.8)	212.7 (175.8–252.3)	4.3 (4.0–4.6)
70–74	302.2 (280.5–326.9)	7.2 (7.0–7.5)	100.6 (92.9–109.4)	2.4 (2.3–2.5)	621.6 (532.4–721.2)	14.9 (14.0–15.7)	207.3 (175.4–241.5)	5.0 (4.7–5.3)	652.9 (539.1–774.1)	15.6 (14.5–16.6)	217.1 (179.1–257.8)	5.2 (4.9–5.6)
75–79	311.9 (289.8–337.9)	8.4 (8.2–8.7)	103.9 (96.1–112.9)	2.8 (2.7–2.9)	642.1 (547.5–747.0)	17.3 (16.4–18.2)	214.1 (182.0–248.8)	5.8 (5.5–6.1)	674.1 (554.6–798.2)	18.1 (17.0–19.3)	224.1 (184.8–267.3)	6.1 (5.7–6.5)

80+	255.4 (237.3–275.8)	9.0 (8.8–9.3)	85.1 (78.7–92.2)	3.0 (2.9–3.1)	526.0 (446.7–611.5)	18.5 (17.6–19.5)	174.9 (148.7–203.6)	6.2 (5.9–6.6)	551.6 (455.3–653.5)	19.4 (18.2–20.6)	183.5 (151.9–217.7)	6.5 (6.1–6.9)
Multiplicative Model[b]												
0–19	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.0 (0.0–0.0)	0.1 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.1)	0.0 (0.0–0.0)	0.1 (0.0–0.1)	0.0 (0.0–0.0)	0.0 (0.0–0.1)	0.0 (0.0–0.0)
20–24	0.8 (0.6–0.1)	0.0 (0.0–0.0)	0.3 (0.2–0.4)	0.0 (0.0–0.0)	2.1 (1.5–2.9)	0.0 (0.0–0.0)	0.7 (0.4–1.1)	0.0 (0.0–0.0)	2.4 (1.5–3.5)	0.0 (0.0–0.0)	0.8 (0.4–1.3)	0.0 (0.0–0.0)
25–29	4.5 (3.9–5.3)	0.0 (0.0–0.0)	1.5 (1.2–1.9)	0.0 (0.0–0.0)	11.2 (8.6–14.3)	0.1 (0.0–0.1)	3.7 (2.7–5.0)	0.0 (0.0–0.0)	12.6 (8.8–17.2)	0.1 (0.0–0.1)	4.2 (2.3–5.9)	0.0 (0.0–0.0)
30–34	16.0 (14.3–18.1)	0.1 (0.0–0.1)	5.3 (4.6–6.2)	0.0 (0.0–0.0)	40.0 (31.2–49.4)	0.3 (0.2–0.3)	13.2 (10.0–16.8)	0.1 (0.1–0.1)	44.8 (31.7–60.2)	0.3 (0.2–0.4)	14.9 (10.4–19.9)	0.1 (0.1–0.1)
35–39	38.0 (34.1–42.4)	0.3 (0.2–0.3)	12.6 (11.2–14.3)	0.1 (0.1–0.1)	94.3 (74.2–116.7)	0.7 (0.6–0.9)	31.4 (24.7–39.0)	0.2 (0.2–0.3)	106.2 (74.5–140.3)	0.8 (0.7–1.0)	35.2 (24.7–47.2)	0.3 (0.2–0.3)
40–44	83.9 (75.6–93.3)	0.7 (0.6–0.7)	27.9 (25.0–31.2)	0.2 (0.2–0.2)	209.0 (165.5–257.8)	1.8 (1.5–2.0)	69.4 (54.6–85.9)	0.6 (0.5–0.7)	235.2 (164.3–307.7)	2.0 (1.6–2.4)	78.0 (54.7–103.2)	0.7 (0.5–0.8)
45–49	131.5 (118.7–147.1)	1.4 (1.3–1.5)	43.8 (39.4–49.1)	0.5 (0.4–0.5)	326.8 (259.9–402.0)	3.4 (2.9–3.8)	108.9 (85.4–134.6)	1.1 (1.0–1.3)	367.2 (260.0–485.0)	3.8 (3.1–4.5)	122.8 (86.2–160.5)	1.3 (1.0–1.5)
50–54	159.1 (144.0–177.6)	2.1 (2.0–2.3)	52.9 (47.8–59.2)	0.7 (0.6–0.7)	396.1 (311.6–488.4)	5.3 (4.7–5.9)	132.0 (104.2–163.2)	1.8 (1.6–2.0)	446.3 (315.0–589.5)	5.9 (5.0–6.9)	148.0 (104.2–195.8)	2.0 (1.7–2.3)
55–59	194.2 (175.3–216.6)	3.1 (2.9–3.2)	64.6 (58.3-72.3)	1.0 (0.9–1.0)	483.8 (380.2–593.2)	7.6 (6.8–8.4)	160.9 (126.6–198.7)	2.5 (2.3–2.8)	542.6 (382.6–714.6)	8.5 (7.3–9.8)	180.6 (127.0–239.6)	2.9 (2.4–3.3)
60–64	234.5 (211.4–261.3)	4.2 (4.0–4.4)	78.1 (70.3–87.1)	1.4 (1.3–1.5)	582.7 (460.2–720.3)	10.3 (9.3–11.3)	194.2 (153.2–238.5)	3.5 (3.1–3.8)	656.0 (456.5–860.1)	11.5 (10.1–13.1)	218.2 (153.9–288.2)	3.9 (3.4–4.4)
65–69	265.5 (239.5–295.9)	5.4 (5.1–5.6)	88.4 (79.5–98.8)	1.8 (1.7–1.9)	662.0 (518.3–815.1)	13.3 (12.1–14.4)	220.0 (173.2–270.3)	4.5 (4.1–4.8)	744.2 (518.8–981.7)	14.9 (13.4–16.7)	247.3 (172.3–328.6)	5.0 (4.4–5.6)
70–74	271.0 (244.1–301.6)	6.5 (6.3–6.8)	90.2 (81.0–101.0)	2.2 (2.1–2.3)	674.2 (530.6–831.2)	16.1 (14.8–17.4)	224.0 (178.0–276.4)	5.4 (5.0–5.9)	755.2 (539.1–996.1)	18.0 (16.0–20.1)	252.7 (175.9–334.1)	6.1 (5.4–6.8)
75–79	279.7 (252.3–312.4)	7.6 (7.3–7.9)	93.1 (83.8–104.2)	2.5 (2.4–2.6)	697.3 (549.8–860.7)	18.7 (17.3–20.1)	231.7 (182.7–286.5)	6.3 (5.8–6.8)	781.1 (545.1–1036.6)	21.0 (18.8–23.3)	260.4 (183.7–344.5)	7.1 (6.3–7.9)
80+	229.0 (206.8–254.9)	8.1 (7.8–8.4)	76.3 (68.6–85.1)	2.7 (2.6–2.8)	570.4 (450.2–704.6)	20.1 (18.6–21.5)	189.7 (149.3–234.0)	6.7 (6.3–7.3)	639.5 (446.6–847.5)	22.5 (20.3–24.8)	213.5 (149.8–282.3)	7.6 (6.9–8.4)

Estimates of lifetime risk of developing breast cancer from birth among noncarriers and carriers of genetic variants were based on the age-specific breast cancer incidence rates of U.S. non-Hispanic white women in 2003–2005 and by assuming 1/3 of U.S. breast cancer incidence rates. Abbreviations: CI, confidence interval; PI, prediction interval; SI, simulation interval.

[a] Simulation interval was based on 10,000 Monte Carlo simulations. We assumed that the breast cancer incidence cases and number of deaths followed the Poisson distribution (when the number of incidence cases or number of deaths were large ($n > 1000$), normal distribution was used to approximate Poisson distribution), genotype frequency and dominant risk ratio followed the normal distribution with mean as point estimate and standard deviation as the difference between upper and lower 95% confidence interval (CI) or 95% prediction interval (PI) divided by 3.92.

[b] The risk ratios of joint effect of three or five SNPs on additive or multiplicative scale were calculated as defined in Material and Methods section. We took the lower 95% prediction interval (PI) of risk ratio of each SNP, calculated the risk ratios of joint effect on additive and multiplicative scale respectively and defined them as the minimum risk ratio in the Monte Carlo simulation. In a similar fashion, we took the upper 95% PI of risk ratio of each SNP, calculated the risk ratios on additive and multiplicative scale and defined them as the maximum risk ratio in the simulation (defined as the triangle distribution with the mean value as the point estimate of risk ratio on additive or multiplicative scale).

The recent explosion of DTC personal genome tests offered by the companies in different countries raises the concerns among the scientific communities and oversight groups regarding the possible health benefits and the undesired consequences for individuals.[13,18,19,56,69] Our study focused on examining the uncertainty in risk assessment as part of clinical validity of genetic testing. Our results indicated that it is important to recognize the impact of cumulative uncertainty and variation in the epidemiologic parameters involved in computation of disease risk. Providing consumers with these risk estimates without proper interpretation of the uncertainty and variation around these estimates could be seriously misleading.

Appendix

Estimations of Age-Specific Incidence Rate among Carriers of Multiple Genetic Variants

For simplicity, we consider N independent biallelic disease-susceptibility loci. Let G_1, G_2… G_n, be genotype frequency in population, and R_1, R_2,…,R_n, be the risk ratio for disease for subjects with risk genotype compared with subjects without risk genotype. The odds ratio from case-control association studies is a proxy of the population-level risk ratio. Let i_1, i_2,…,i_n, be binary numbers (0/1) depending on the (presence/absence) of the genetic variants. We have limited knowledge about how the multiple genetic variants might act in consort to affect the disease risk. There is a long debate about how to define and measure interaction in epidemiologic studies.[42,70] For simplicity, we consider the joint effects of a panel of genetic variants on either an additive or multiplicative scale without any extra interaction effect.[42,43] To illustrate the additive and multiplicative effect model, let us consider two independent biallelic disease susceptibility loci. Let G_1 and G_2 be genotype frequencies in population, and let Rg_{11}, Rg_{10}, and Rg_{01} be risk ratios of having both genetic variants, genetic variant 1 only (G_1) and genetic variant 2 only (G_2), respectively. The joint effect on an additive scale is defined as: $Rg_{11} = Rg_{10} + Rg_{01} - 1$. The joint effect on a multiplicative scale is given as: $Rg_{11} = Rg_{10}$*Rg_{01}. Assuming additive effects, the age-specific incidence rate among subjects with different combination of multiple genetic variants is obtained by:

$$P_k(D \mid G_1 = i_1, G_2 = i_2, \ldots, G_n = i_n) = \frac{P(D \mid G_1 = i_1, G_2 = i_2, \ldots, G_n = i_n)P(D)}{\sum_{i,i_2,\ldots i_k} P(D \mid G_1 = i_1, G_2 = i_2, \ldots, G_n = i_n)P(G_1 = i_1, G_2 = i_2, \ldots, G_n = i_n)} = \frac{\left[1 + \sum_{j=1}^{n} i_j(R_j - 1)\right]P(D)}{1 + \sum_{j=1}^{n} G_j(R_j - 1)} \quad (1)$$

Assuming multiplicative effects, the age-specific incidence rate among subjects with different combination of multiple genetic variants is obtained by:

$$P_k(D \mid G_1 = i_1, G_2 = i_2, \ldots, G_n = i_n) = \frac{R_1^{i_1} R_2^{i_2} \ldots R_n^{i_n} P(D)}{[R_1 G_1 + (1 - G_1)][R_2 G_2 + (1 - G_2)] \ldots [R_n G_n + (1 - G_n)]} = \prod_{j=1}^{n} \frac{R_j^{ij} P(D)}{[R_j G_j + (1 - G_j)]} \quad (2)$$

Calculate Age-Specific Incidence Rate by Genotype

For a single biallelic locus with alleles A and a, there are three possible genotypes: AA (homozygous dominant), Aa (heterozygous), and aa (homozygous recessive). Let G_0, G_1, and G_2 denote the prevalence of homozygous recessive, heterozygous, and homozygous dominant genotypes in a population, respectively. Let ψ_2 and ψ_1 be risk among subjects with genotypes AA and Aa compared with risk among those with genotype aa, respectively. The age-specific incidence rate among subjects with AA or Aa genotypes can also be obtained by:

$$P_k(D \mid G_i) = \frac{\psi_{ki} P_k(D)}{[P_k(G_2)\psi_{k2} + P_k(G_1)\psi_{k1} + P(G_0)]} \quad (3)$$

where ψ_{ki} (I = 1, 2) is risk ratio for disease among subjects with homozygous and heterozygous genotypes in the k_{th} age interval, respectively; and $P_k(G_i)$ is the prevalence of the genotypes in a population.

Lifetime Risk Models

Lifetime and residual lifetime risk (age-conditional probability) estimates of developing breast cancer are derived with life-table methods adjusted for the competing risk of death. In brief, estimates of the lifetime risk of developing breast cancer are computed on the basis of population-based age-specific breast cancer incidence rates and all-causes (excluding breast cancer) mortality rates. A life table is constructed for a hypothetical cancer-free cohort of individuals who are exposed to the age-specific breast cancer rates as they age. The number of cancer-free individuals at the start of a subsequent age interval is computed by subtracting the number who develop cancer and the number who die of other causes from the number of

Table A1. Residual Lifetime Risk of Developing Breast Cancer before Age 80 Years by Baseline Age at Test among *FGFR2* Variant Carriers, Noncarriers, and in the General Population

	Using U.S. Breast Cancer Incidence Rates of Non-Hispanic White in 2003–2005		Assuming 1/3 of U.S. Breast Cancer Incidence Rates of Non-Hispanic White in 2003–2005		U.S. Non-Hispanic White Women in 2003–2005
Baseline Age (yrs)	***FGFR2* Variant Carriers, % (95% SI)[a]**	**Noncarriers, % (95% SI)[a]**	***FGFR2* Variant Carriers, % (95% SI)[a]**	**Noncarriers, % (95% SI)[a]**	**% (95% CI)[b]**
Birth	15.7 (15.5–15.8)	11.6 (11.5–11.8)	5.3 (5.2–5.3)	3.9 (3.8–4.0)	13.4 (13.3–13.5)
20	15.7 (15.6–15.8)	11.7 (11.5–11.8)	5.3 (5.2–5.3)	3.9 (3.8–4.0)	13.6 (13.5–13.6)
30	15.7 (15.5–15.8)	11.6 (11.5–11.8)	5.3 (5.2–5.3)	3.9 (3.8–4.0)	13.6 (13.5–13.5)
40	15.3 (15.2–15.4)	11.3 (11.2–11.5)	5.1 (5.0–5.2)	3.8 (3.7–3.8)	13.2 (13.1–13.3)
50	13.3 (13.1–13.4)	9.8 (9.7–10.0)	4.4 (4.3–4.5)	3.3 (3.2–3.3)	11.7 (11.6–11.7)
60	9.9 (9.8–10.0)	7.3 (7.2–7.5)	3.3 (3.2–3.3)	2.4 (2.3–2.5)	9.0 (8.9–9.0)
70	5.2 (5.1–5.3)	3.9 (3.8–4.0)	1.7 (1.7–1.8)	1.3 (1.2–1.3)	5.0 (5.0–5.1)

Abbreviations: SI, simulation interval; CI, confidence interval.
[a] Simulation interval was based on 10,000 Monte Carlo simulations. We assumed that the breast cancer incidence cases and number of deaths followed the Poisson distribution (when the number of incidence cases or number of deaths were large [n > 1000], normal distribution was used to approximate Poisson distribution), genotype frequency and dominant risk ratio followed normal distribution with mean as point estimate and standard deviation as the difference between upper and lower 95% confidence interval (CI) or 95% prediction interval (PI) divided by 3.92.
[b] Residual lifetime risk of developing breast cancer among U.S. non-Hispanic white women in 2003–2005 U.S. derived from DevCan software (version 6.3.1).

cancer-free individuals at the start of the interval. The lifetime risk of developing breast cancer (from birth) is estimated by dividing the sum of all expected breast cancer cases in the life table by the number of individuals in the initial birth cohort. The residual lifetime risk (age conditional probability) of developing breast cancer (from any specific age until certain age, e.g., age 80 years) is estimated by dividing the sum of all expected breast cancer cases from the age at test onward in the life table by the number of individuals at age interval of the test.[28,29] Detailed descriptions of the lifetime table methodology to estimate lifetime and residual lifetime risk have been published elsewhere.[28–30] In the estimates of lifetime risk among the carriers of the genetic variants, we have replaced the population-based age-specific incidence rates with our estimated age-specific incidence rates as described in the Material and Methods section. We also calculated the residual lifetime risk of developing breast cancer before age 80 when testing is not performed at birth but at different ages X, e.g., at X = 20, 30, 40, etc., years of age via the life table method.[31] Residual lifetime risk was estimated separately for U.S. non-Hispanic women breast cancer incidence rates from 2003–2005 and for the 3-fold downward adjusted breast cancer incidence rates.

Table A1 lists the residual lifetime risk of developing breast cancer before age 80 by baseline age at test among carriers of *FGFR2* genetic variant, noncarriers, and in general population stratified by the U.S. non-Hispanic women breast cancer incidence rates from 2003–2005 and the 3-fold downward adjusted breast cancer incidence rates. Table A2 lists the estimated residual lifetime risk of developing breast cancer by baseline age at test assuming additive or multiplicative joint effects of five genetic variants stratified by the different breast cancer incidence rates.

Acknowledgments

We thank Tiebin Liu for his help with the meta-analysis of the selected SNPs.

Disclaimer: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention or of the Tufts Clinical and Translational Science Institute.

Received: July 13, 2009
Revised: October 15, 2009
Accepted: October 18, 2009
Published online: November 19, 2009

Web Resources

The URLs for data presented herein are as follows:

A Catalog of Published Genome-Wide Association Studies, http://www.genome.gov/26525384
DevCan, http://srab.cancer.gov/devcan/
HapMap, http://www.hapmap.org/
International Diabetes Federation: http://www.eatlas.idf.org/
More Gene Direct: a report on developments in the availability, marketing and regulation of genetic tests supplied directly to the public 2007, http://www.hgc.gov.uk/Client/document.asp?DocId=139&CAtegoryId=10
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
Personalized Medicine Coalition: Personal genomics and industry standards: scientific validity, http://www.personalizedmedicinecoalition.org/objects/pdfs/PMC%20personalgenomicsSci%20Valid15dec08.pdf
Screening for Breast Cancer. Feb 2002 Agency for Healthcare Research and Quality, Rockville, MD, http://www.ahrq.gov/clinic/3rduspstf/breastcancer

Table A2. Residual Lifetime Risk of Developing Breast Cancer before Age 80 Years by Baseline Age at Test among Carriers of Five Genetic Variants

	Using U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005		Assuming 1/3 of U.S. Breast Cancer Incidence Rates of Non-Hispanic White Women in 2003–2005	
Baseline Age (yrs)	**Additive Joint Effect % (95% SI)[a]**	**Multiplicative Joint Effect % (95% SI)[a]**	**Additive Joint Effect % (95% SI)[a]**	**Multiplicative Joint Effect % (95% SI)[a]**
Birth	18.1 (17.0–19.3)	21.0 (18.8–23.3)	6.1 (5.7–6.5)	7.1 (6.3–7.9)
20	18.1 (17.0–19.3)	21.0 (18.8–23.3)	6.1 (5.7–6.5)	7.1 (6.3–7.9)
30	18.1 (17.0–19.2)	11.0 (18.8–23.3)	6.1 (5.7–6.5)	7.1 (6.3–7.9)
40	17.7 (16.7–18.9)	20.1 (18.3–22.9)	5.9 (5.5–6.3)	6.9 (6.1–7.7)
50	15.4 (14.3–16.5)	17.9 (15.6–20.2)	5.1 (4.7–5.5)	5.9 (5.2–6.7)
60	11.5 (10.5–12.6)	13.4 (11.4–15.6)	3.8 (3.5–4.2)	4.4 (3.8–5.1)
70	6.0 (5.3–6.9)	7.0 (5.6–8.7)	2.0 (1.8–2.3)	2.3 (1.8–2.9)

Abbreviations: SI, simulation interval.
[a] Simulation interval was based on 10,000 Monte Carlo simulations. We assumed that the breast cancer incidence cases and number of deaths followed the Poisson distribution (when the number of incidence cases or number of deaths were large [n > 1000], normal distribution was used to approximate Poisson distribution), genotype frequency and dominant risk ratio followed the normal distribution with mean as point estimate and standard deviation as the difference between upper and lower 95% confidence interval (CI) or 95% prediction interval (PI) divided by 3.92. The risk ratios of joint effect of five SNPs on additive or multiplicative scale were calculated as defined in Material and Methods section. We took the lower 95% PI of risk ratio for each SNP, calculated the risk ratios of joint effect on additive and multiplicative scale, respectively, and defined them as the minimum risk ratio in the Monte Carlo simulation. In a similar fashion, we took the upper 95% PI of risk ratio for each SNP, calculated the risk ratios on additive and multiplicative scale, and defined them as the maximum risk ratio in the simulation (defined as the triangle distribution with the mean value as the point estimate of risk ratio on additive or multiplicative scale).

References

1. Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science *322*, 881–888.
2. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nat. Rev. Genet. *9*, 356–369.
3. Pearson, T.A., and Manolio, T.A. (2008). How to interpret a genome-wide association study. JAMA *299*, 1335–1344.
4. Collins, F.S. (2006). 2005 William Allan Award address. No longer just looking under the lamppost. Am. J. Hum. Genet. *79*, 421–426.
5. Collins, F.S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. Science *300*, 286–290.
6. Guttmacher, A.E., and Collins, F.S. (2002). Genomic medicine—A primer. N. Engl. J. Med. *347*, 1512–1520.
7. Guttmacher, A.E., and Collins, F.S. (2005). Realizing the promise of genomics in biomedical research. JAMA *294*, 1399–1402.
8. Jones, S., and Collins, F. (2003). Genomics in medicine: Hype or real promise? Interview by Ed Rabinowitz. Healthplan *44*, 20–24.
9. Goddard, K.A., Moore, C., Ottman, D., Szegda, K.L., Bradley, L., and Khoury, M.J. (2007). Awareness and use of direct-to-consumer nutrigenomic tests, United States, 2006. Genet. Med. *9*, 510–517.
10. Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A., and Bradley, L. (2007). The continuum of translation research in genomic medicine: How can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? Genet. Med. *9*, 665–674.
11. Khoury, M.J., McCabe, L.L., and McCabe, E.R. (2003). Population screening in the age of genomic medicine. N. Engl. J. Med. *348*, 50–58.
12. Hunter, D.J., Khoury, M.J., and Drazen, J.M. (2008). Letting the genome out of the bottle—Will we get our wish? N. Engl. J. Med. *358*, 105–107.
13. Hogarth, S., Javitt, G., and Melzer, D. (2008). The current landscape for direct-to-consumer genetic testing: Legal, ethical, and policy issues. Annu. Rev. Genomics Hum. Genet. *9*, 161–182.
14. Kaye, J. (2008). The regulation of direct-to-consumer genetic tests. Hum. Mol. Genet. *17*, R180–R183.
15. Burke, W., and Psaty, B.M. (2007). Personalized medicine in the era of genomics. JAMA *298*, 1682–1684.
16. Feero, W.G., Guttmacher, A.E., and Collins, F.S. (2008). The genome gets personal—Almost. JAMA *299*, 1351–1352.
17. Scheuner, M.T., Sieverding, P., and Shekelle, P.G. (2008). Delivery of genomic medicine for common chronic adult diseases: A systematic review. JAMA *299*, 1320–1334.
18. Melzer, D., Hogarth, S., Liddell, K., Ling, T., Sanderson, S., and Zimmern, R.L. (2008). Genetic tests for common diseases: New insights, old concerns. BMJ *336*, 590–593.
19. Offit, K. (2008). Genomic profiles for disease risk—Predictive or premature? JAMA *299*, 1353–1355.
20. Hudson, K., Javitt, G., Burke, W., and Byers, P. (2007). ASHG statement on direct-to-consumer genetic testing in the United States. Am. J. Hum. Genet. *81*, 635–637.
21. Task Force on Genetic Testing (U.S.), Holtzman, N.A., and Watson, M.S. (1998). Promoting Safe and Effective Genetic Testing in the United States: Final Report of the Task Force on Genetic Testing (Baltimore, MD: Johns Hopkins University Press).
22. Manolio, T.A., Bailey-Wilson, J.E., and Collins, F.S. (2006). Genes, environment and the value of prospective cohort studies. Nat. Rev. Genet. *7*, 812–820.
23. Rothman, K.J., and Greenland, S. (1998). Modern Epidemiology (Philadelphia, PA: Lippincott-Raven).

24. Gail, M.H. (2008). Estimation and interpretation of models of absolute risk from epidemiologic data, including family-based studies. Lifetime Data Anal. *14*, 18–36.
25. Fu, R., Harris, E.L., Helfand, M., and Nelson, H.D. (2007). Estimating risk of breast cancer in carriers of BRCA1 and BRCA2 mutations: A meta-analytic approach. Stat. Med. *26*, 1775–1787.
26. Yang, Q., Khoury, M.J., Coughlin, S.S., Sun, F., and Flanders, W.D. (2000). On the use of population-based registries in the clinical validation of genetic tests for disease susceptibility. Genet. Med. *2*, 186–192.
27. Satagopan, J.M., Offit, K., Foulkes, W., Robson, M.E., Wacholder, S., Eng, C.M., Karp, S.E., and Begg, C.B. (2001). The lifetime risks of breast cancer in Ashkenazi Jewish carriers of BRCA1 and BRCA2 mutations. Cancer Epidemiol. Biomarkers Prev. *10*, 467–473.
28. Feuer, E.J., Wun, L.M., Boring, C.C., Flanders, W.D., Timmel, M.J., and Tong, T. (1993). The lifetime risk of developing breast cancer. J. Natl. Cancer Inst. *85*, 892–897.
29. Wun, L.M., Merrill, R.M., and Feuer, E.J. (1998). Estimating lifetime and age-conditional probabilities of developing cancer. Lifetime Data Anal. *4*, 169–186.
30. Fay, M.P., Pfeiffer, R., Cronin, K.A., Le, C., and Feuer, E.J. (2003). Age-conditional probabilities of developing cancer. Stat. Med. *22*, 1837–1848.
31. Jemal, A., Siegel, R., Ward, E., Hao, Y.P., Xu, J.Q., Murray, T., and Thun, M.J. (2008). Cancer statistics, 2008. CA Cancer J. Clin. *58*, 71–96.
32. Parkin, D.M., Bray, F., Ferlay, J., and Pisani, P. (2005). Global cancer statistics, 2002. CA Cancer J. Clin. *55*, 74–108.
33. Sasieni, P.D. (1997). From genotypes to genes: Doubling the sample size. Biometrics *53*, 1253–1261.
34. Zhang, J., and Yu, K.F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA *280*, 1690–1691.
35. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087–1093.
36. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. *39*, 870–874.
37. Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., et al. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. *39*, 865–869.
38. Higgins, J.P., Thompson, S.G., and Spiegelhalter, D.J. (2009). A re-evaluation of random-effects meta-analysis. J. R. Stat. Soc. Ser. A Stat. Soc. *172*, 137–159.
39. Moonesinghe, R., Khoury, M.J., Liu, T., and Ioannidis, J.P. (2008). Required sample size and nonreplicability thresholds for heterogeneous genetic associations. Proc. Natl. Acad. Sci. USA *105*, 617–622.
40. Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D.A., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Arias, J.I., et al. (2008). Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet *4*, e1000054.
41. Antoniou, A.C., Sinilnikova, O.M., McGuffog, L., Healey, S., Nevanlinna, H., Heikkinen, T., Simard, J., Spurdle, A.B., Beesley, J., Chen, X., et al. (2009). Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum. Mol. Genet. *18*, 4442–4456.
42. Rothman, K.J. (2002). Epidemiology: An Introduction (New York: Oxford University Press).
43. Yang, Q., Khoury, M.J., Friedman, J., Little, J., and Flanders, W.D. (2005). How many genes underlie the occurrence of common complex diseases in the population? Int. J. Epidemiol. *34*, 1129–1137.
44. Greenland, S. (2001). Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. Risk Anal. *21*, 579–583.
45. Steenland, K., and Greenland, S. (2004). Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am. J. Epidemiol. *160*, 384–392.
46. Jurek, A.M., Maldonado, G., Greenland, S., and Church, T.R. (2007). Uncertainty analysis: An example of its application to estimating a survey proportion. J. Epidemiol. Community Health *61*, 650–654.
47. Khoury, M.J., McBride, C.M., Schully, S.D., Ioannidis, J.P., Feero, W.G., Janssens, A.C., Gwinn, M., Simons-Morton, D.G., Bernhardt, J.M., Cargill, M., et al. (2009). The Scientific Foundation for personal genomics: recommendations from a National Institutes of Health-Centers for Disease Control and Prevention multidisciplinary workshop. Genet. Med. *11*, 559–567.
48. Mathers, C., Fat, D.M., Boerma, J.T., and World Health Organization.. (2008). The Global Burden of Disease: 2004 Update (Geneva, Switzerland: World Health Organization).
49. Geiss, L.S., Pan, L., Cadwell, B., Gregg, E.W., Benjamin, S.M., and Engelgau, M.M. (2006). Changes in incidence of diabetes in U.S. adults, 1997–2003. Am. J. Prev. Med. *30*, 371–377.
50. Narayan, K.M., Boyle, J.P., Geiss, L.S., Saaddine, J.B., and Thompson, T.J. (2006). Impact of recent increase in incidence on future diabetes burden: U.S., 2005–2050. Diabetes Care *29*, 2114–2116.
51. Burton, P.R., Hansell, A.L., Fortier, I., Manolio, T.A., Khoury, M.J., Little, J., and Elliott, P. (2009). Size matters: Just how big is BIG?: Quantifying realistic sample size requirements for human genome epidemiology. Int. J. Epidemiol. *38*, 263–273.
52. Bjornvold, M., Undlien, D.E., Joner, G., Dahl-Jorgensen, K., Njolstad, P.R., Akselsen, H.E., Gervin, K., Ronningen, K.S., and Stene, L.C. (2008). Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes. Diabetologia *51*, 589–596.
53. Hill, W.G., Goddard, M.E., and Visscher, P.M. (2008). Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet *4*, e1000008.
54. Jones, H.B., and Faham, M. (2005). Evidence and implications for multiplicative interactions among loci predisposing to human common disease. Hum. Hered. *59*, 176–184.
55. Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H.O., Hsu, F.C., Zhu, Y., Balter, K., et al. (2008). Cumulative association of five genetic variants with prostate cancer. N. Engl. J. Med. *358*, 910–919.

56. Ng, P.C., Murray, S.S., Levy, S., and Venter, J.C. (2009). An agenda for personalized medicine. Nature *461*, 724–726.
57. Botto, L.D., and Yang, Q. (2000). 5,10-Methylenetetrahydrofolate reductase gene variants and congenital anomalies: a HuGE review. Am. J. Epidemiol. *151*, 862–877.
58. Paracchini, V., Pedotti, P., and Taioli, E. (2005). Genetics of leptin and obesity: A HuGE review. Am. J. Epidemiol. *162*, 101–114.
59. Myles, S., Davison, D., Barrett, J., Stoneking, M., and Timpson, N. (2008). Worldwide population differentiation at disease-associated SNPs. BMC Med Genomics *1*, 22.
60. Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P.K., Holmkvist, J., Borch-Johnsen, K., Jorgensen, T., et al. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. Nat. Genet. *40*, 1098–1102.
61. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I.W., Abecasis, G.R., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat. Genet. *40*, 638–645.
62. Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science *273*, 1516–1517.
63. Gail, M.H. (2008). Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J. Natl. Cancer Inst. *100*, 1037–1041.
64. Janssens, A.C., Moonesinghe, R., Yang, Q., Steyerberg, E.W., van Duijn, C.M., and Khoury, M.J. (2007). The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. Genet. Med. *9*, 528–535.
65. Pharoah, P.D., Antoniou, A.C., Easton, D.F., and Ponder, B.A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. N. Engl. J. Med. *358*, 2796–2803.
66. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., and Mulvihill, J.J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J. Natl. Cancer Inst. *81*, 1879–1886.
67. Heck, K.E., Schoendorf, K.C., Ventura, S.J., and Kiely, J.L. (1997). Delayed childbearing by education level in the United States, 1969–1994. Matern. Child Health J. *1*, 81–88.
68. National Institutes of Health Consensus Development Panel. (1997). National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40–49, January 21–23, 1997. National Institutes of Health Consensus Development Panel. J. Natl. Cancer Inst. *89*, 1015–1026.
69. McGuire, A.L., and Burke, W. (2008). An unwelcome side effect of direct-to-consumer personal genome testing: Raiding the medical commons. JAMA *300*, 2669–2671.
70. Yang, Q., and Khoury, M.J. (1997). Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. Epidemiol. Rev. *19*, 33–43.