

## Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium

### Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts

Bruce M. Psaty, MD, PhD; Christopher J. O'Donnell, MD, MPH; Vilmundur Gudnason, MD, PhD; Kathryn L. Lunetta, PhD; Aaron R. Folsom, MD; Jerome I. Rotter, MD; André G. Uitterlinden, PhD; Tamara B. Harris, MD; Jacqueline C.M. Witteman, PhD; Eric Boerwinkle, PhD;  
on Behalf of the CHARGE Consortium

**Background**—The primary aim of genome-wide association studies is to identify novel genetic loci associated with interindividual variation in the levels of risk factors, the degree of subclinical disease, or the risk of clinical disease. The requirement for large sample sizes and the importance of replication have served as powerful incentives for scientific collaboration.

**Methods**—The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium was formed to facilitate genome-wide association studies meta-analyses and replication opportunities among multiple large population-based cohort studies, which collect data in a standardized fashion and represent the preferred method for estimating disease incidence. The design of the Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium includes 5 prospective cohort studies from the United States and Europe: the Age, Gene/Environment Susceptibility—Reykjavik Study, the Atherosclerosis Risk in Communities Study, the Cardiovascular Health Study, the Framingham Heart Study, and the Rotterdam Study. With genome-wide data on a total of about 38 000 individuals, these cohort studies have a large number of health-related phenotypes measured in similar ways. For each harmonized trait, within-cohort genome-wide association study analyses are combined by meta-analysis. A prospective meta-analysis of data from all 5 cohorts, with a properly selected level of genome-wide statistical significance, is a powerful approach to finding genuine phenotypic associations with novel genetic loci.

**Conclusions**—The Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium and collaborating non-member studies or consortia provide an excellent framework for the identification of the genetic determinants of risk factors, subclinical-disease measures, and clinical events. (*Circ Cardiovasc Genet.* 2009;2:73-80.)

**Key Words:** epidemiology ■ meta-analysis ■ genetics ■ genomics

Early efforts to localize and identify genes that contribute to the risk of common chronic diseases often used either candidate gene studies or family-based linkage studies, which suffered from low statistical power, lack of replication, and low precision.<sup>1</sup> Although there were successes, progress was generally slow. Recently, genome-wide association studies (GWAS) have proven to be productive when they have adequate sample sizes and replication opportunities. Their

primary aim is to identify novel genetic loci associated with interindividual variation in the levels of risk factors, the measures of subclinical disease, or the risk of clinical events. The method does not require assumptions about a priori biological involvement, is precise in its ability to localize genetic effects to relatively small regions of the genome, and can be extended to evaluate potential gene-environment interactions.

Received October 17, 2008; accepted December 17, 2008.

Guest editor for this article was Elizabeth R. Hauser, PhD.

From the Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services (B.M.P.), University of Wash; Center for Health Studies, Group Health (B.M.P.), Seattle, Wash; the National Heart, Lung and Blood Institute and the Framingham Heart Study (C.J.O.D.), Framingham, Mass; Icelandic Heart Association and the Department of Cardiovascular Genetics (Y.G.), University of Iceland, Reykjavik, Iceland; Department of Biostatistics (K.L.), Boston University School of Public Health, Mass; Division of Epidemiology and Community Health (A.R.F.), University of Minnesota, Minneapolis; Medical Genetics Institute (J.I.R.), Cedars-Sinai Medical Center, Los Angeles, Calif; Departments of Internal Medicine (A.G.U.) and Epidemiology (A.G.U., J.C.M.W.), Erasmus Medical Center, Rotterdam, The Netherlands; Laboratory of Epidemiology, Demography, and Biometry (T.B.H.), Intramural Research Program, National Institute on Aging, Bethesda, Md; and Human Genetics Center and Division of Epidemiology (E.B.), University of Texas, Houston.

The online Data Supplement is available at <http://circgenetics.ahajournals.org/cgi/content/full/2/1/73/DC1>.

Correspondence to Bruce M. Psaty, MD, PhD, Cardiovascular Health Research Unit, 1730 Minor Ave, Suite 1360, Seattle, WA 98101. E-mail [psaty@u.washington.edu](mailto:psaty@u.washington.edu)

© 2009 American Heart Association, Inc.

*Circ Cardiovasc Genet* is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.108.829747

GWAS have successfully identified genetic loci associated with a variety of conditions such as type 2 diabetes<sup>2</sup> and coronary disease.<sup>3–5</sup> The large number of statistical tests required in GWAS poses a special challenge because few studies that have DNA and high-quality phenotype data are sufficiently large to provide adequate statistical power for detecting small to modest effect sizes.<sup>6</sup> Meta-analyses combining previously published findings have improved the ability to detect new loci.<sup>2</sup> Even before the era of GWAS,<sup>7</sup> the requirement for large sample sizes and the importance of replication have served as powerful incentives for collaboration.

Our understanding of the risk factors for common chronic diseases has benefited from large population-based cohort studies. Although these studies are costly and time consuming, they are generally free of the survival and recall biases typically encountered in case-control studies. The cohort design, with its prospective standardized data collection, is often the preferred method for estimating disease incidence and evaluating risk factors. The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium was formed to facilitate GWAS meta-analyses and replication opportunities among multiple large and well-phenotyped cohort studies. The design of the CHARGE Consortium includes 5 prospective cohort studies from the United States and Europe: the Age, Gene/Environment Susceptibility (AGES)—Reykjavik Study, the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), and the Rotterdam Study (RS).

With genome-wide data on about 38 000 individuals, these cohort studies have a large number of phenotypes measured in a similar way, and a prospective meta-analysis of within-study association data from the 5 studies, with a properly selected level of genome-wide significance, is a powerful approach to finding genuine phenotypic associations with novel genetic loci. The CHARGE Consortium provides a unique opportunity for collaborative investigation of the genetic determinants of risk factors, measures of subclinical disease, and clinical events.

## Design

### Cohorts

Participating studies<sup>8–14</sup> were prospective cohort studies that had multiple cardiovascular and aging phenotypes in common and that had genome-wide scans completed or in progress in 2007–2008 (Table 1). Briefly, the AGES-Reykjavik Study represents a sample drawn from the established population-based cohort, the Reykjavik Study.<sup>8</sup> The original Reykjavik Study comprised a random sample of 30 795 men and women living in Reykjavik in 1967 and born between 1907 and 1935. Over the years 1967 to 1996, 6 examinations were conducted in 6 subcohorts. Between 2002 and 2006, the AGES—Reykjavik study reexamined 5764 survivors of the original cohort. The ARIC study is a population-based prospective cohort study of cardiovascular disease and its risk factors sponsored by National Heart, Lung, and Blood Institute. ARIC included 15 792 individuals aged 45 to 64 years at baseline (1987–1989), chosen by

probability sampling from 4 US communities.<sup>9</sup> Cohort members completed 4 clinic examinations, conducted 3 years apart between 1987 and 1998. Follow-up for clinical events is annual. The CHS is a population-based National Heart, Lung, and Blood Institute-funded cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted at 4 field centers.<sup>10</sup> The original predominantly white cohort of 5201 persons was recruited in 1989–1990 from random samples of the Medicare lists. An additional 687 African-Americans were enrolled in 1992–1993. CHS participants completed standardized clinical examinations and questionnaires at study baseline and at 9 annual follow-up visits. Follow-up for clinical events occurs every 6 months. The FHS began in 1948 with the recruitment of an original cohort of 5209 men and women who were 28 to 62 years of age at entry.<sup>11</sup> Clinic examinations were performed approximately every 2 years. In 1971, a second generation of study participants, 5124 children and spouses of children of the original cohort were enrolled.<sup>12</sup> With 2 exceptions, clinic examinations took place approximately every 4 years. Enrollment of the third generation cohort of 4095 children of offspring cohort participants began in 2002.<sup>13</sup> The RS is a prospective population-based cohort study comprising 7983 subjects aged 55 years or older. A trained interviewer visited the individuals at home for a computerized questionnaire, and individuals were subsequently examined at a research center. Baseline data were collected between 1990 and 1993.<sup>14</sup> The original cohort underwent 3 additional examinations. In 2000–2001, an additional 3011 individuals aged 55 or older (mainly 55 to 64) were recruited and examined. Since 2006, an additional cohort of individuals aged 45 years or older (mainly 45 to 59 years) is being recruited, comprising 3236 subjects as of May 1, 2008. All of the CHARGE cohort studies were approved by their respective institutional review committee, and the subjects from all the cohorts provided written informed consent.

### Relationships Among the Studies

Each cohort study has its own administrative structure and set of investigators. Although investigators from several cohorts had occasionally collaborated on analyses, there was little precedent for consortia of cardiovascular epidemiology cohorts. In late 2007, it became clear that because all cohorts shared both a common prospective population-based design and a large number of phenotypes assessed by similar data-collection methods (Table 2), a cohort-level collaboration would facilitate a series of prospectively planned joint meta-analyses. The resulting CHARGE consortium represents a voluntary federation of 5 large complex studies. Between October 2007 and February 2008, the principles and procedures for the CHARGE consortium were developed and approved by the parent studies (public website: <http://web.chargeconsortium.com>).

### CHARGE Goals and Organization

The primary aim is the conduct of high-quality analyses that produce, in an efficient and timely manner, reliable and valid findings across multiple cardiovascular and aging-related phenotypes. The organizational structure is

**Table 1. Descriptions of Participating CHARGE Cohorts**

	AGES	ARIC	CHS	FHS	RS
Baseline exam	1967–1991 1991–1996	1987–1989	1989–1990 1992–1993	1948 1971 2002	1990–1993 2000–2001 2006–2008
N	19 381	15 792	5888	14 428	14 626*
Mean age, y	53	54	72	40	66
Women, %	52	55	57	53	59
Race/ethnicity, %					
Caucasian	100	73	84	100	91
African American	0	27	16	0	0
Other	0	0	0	0	9
Sites, N	1	4	4	1	1
GWAS plans					
Date of DNA	2002–2006	1987–1994	1989–1993	1971–2002	1990–2006
Other exclusion criteria	None	None	CVD at baseline	None	None
Eligible, N	3664	15 637	3980	9274	11 689
GWAS data available					
Subjects, N	3219	11 433	3865	8482	10 958
Mean age, y	77	54	72	55	65
Women, %	58	55	61	53	58
Race/ethnicity, %					
Caucasian	100	77	85	100	94
African American	0	23	15	0	0
Other	0	0	0	0	6
Event follow-up					
Through	6/2008	12/2005	6/2006	12/2007	1/2007
Successfully genotyped SNPs, N (autosomes)	325,094	869,224	306,655	502,197	530,683
Total SNPs, N	2 533 153	2 837 224	2 543 877	2 540 223	2 543 877

\*No. recruited through May 1, 2008.

AGES indicates Age, Gene/Environment Susceptibility—Reykjavik Study; ARIC, Atherosclerosis Risk in Communities Study; CHS, The Cardiovascular Health Study; FHS, The Framingham Heart Study; RS, the Rotterdam Study.

Total SNPs=Sum of successfully genotyped+Successfully imputed using the criteria in Table 3.

simple and comprises a Research Steering Committee, an Analysis Committee, a Genotyping Committee, and approximately 20 phenotype-specific working groups. The Research Steering Committee, which has 2 representatives from each cohort, is responsible for establishing the other committees, for nominating working group members, and for developing general guidelines for collaboration, authorship, sharing of results, publication, and timely participation. The Analysis Committee develops guidelines that the working groups are encouraged to adopt or adapt, and the Genotyping Committee coordinates requests for follow-up genotyping.

The main scientific work takes place in the phenotype-specific working groups, which have responsibility for developing and executing the scientific plans. Working groups standardize phenotypes across the cohorts, decide whether and how to include other nonmember studies with similar phenotypes, and agree on analysis plans, often with input from the Analysis Committee. The phenotype working groups also develop plans for authorship and manuscripts, evaluate results, write manuscripts, and decide on the need for follow-up genotyping.

For each manuscript, the working-group investigators establish prespecified plans for analysis and timelines for participation. Before results are shared, each cohort must formally opt-in or opt-out of participation. For any phenotype, each cohort may work with other studies or consortia rather than CHARGE, and individual cohorts remain free to publish cohort-specific findings for any phenotype. The decision to opt-in represents a commitment to collaborate only with the CHARGE working group for that particular analysis until the manuscript is accepted for publication. Only investigators from cohorts that have opted-in have access to shared results. After the results have been shared, investigators cannot opt out to publish their findings on their own. Working group members agree not to share the GWAS findings with outside groups without the permission of the members who generated the data. Transparency, disclosure, and communications about all collaborations, additional follow-up experiments, or efforts to obtain additional funding have been essential to developing, ensuring, and maintaining trust within the consortium.

In practice, many of the CHARGE phenotype working groups have already engaged investigators from nonmember

**Table 2. Selected Phenotypes Available Across the Participating CHARGE Cohorts**

Selected Phenotype	AGES	ARIC	CHS	FHS	RS
Incident events					
MI	A	A	A	A	A
Stroke	A	A	A	A	A
TIA	A	A	A	A	A
Heart failure	A	A	A	A	A
PVD	ND	A	A	A	A
Mortality/longevity	A	A	A	A	A
Dementia	A	A	A	A	A
Subclinical measures					
Electrocardiography	MV	MV	MV	MV	MV
Holter monitor	ND	ND	MV	MV	ND
Echocardiography	P	P	MV	MV	SV
Cardiac MRI	P	ND	ND	SV	ND
Carotid IMT	SV	MV	MV	SV	MV
Cerebral MRI	SV	P	MV	SV	SV
Coronary calcium	SV	P	P	SV	SV
Ankle-brachial index	ND	MV	MV	MV	MV
Abd aortic ultrasound	ND	ND	SV	ND	SV
Bone density/geometry	SV	ND	P	MV	MV
Spine or hip CT	SV	ND	ND	SV	ND
Endothelial function	ND	ND	SV	SV	ND
Vessel wall stiffness	SV	SV	ND	SV	SV
Pulmonary function	P	MV	MV	MV	SV
Sleep studies	SV	P	P	P	SV
Retinal photography	SV	MV	SV	SV	MV
Traditional risk factors					
Diabetes	MV	MV	MV	MV	MV
Hypertension	MV	MV	MV	MV	MV
Atrial fibrillation	MV	MV	MV	MV	MV
Blood pressure	MV	MV	MV	MV	MV
Fasting lipids	MV	MV	MV	MV	MV
Fasting glucose	MV	MV	MV	MV	MV
Glucose tolerance test	MV	SV	SV	MV	ND
Behaviors (smoking)	MV	MV	MV	MV	MV
Medication use	MV	MV	MV	MV	MV
Height, weight	MV	MV	MV	MV	MV
Cognitive function	SV	MV	MV	MV	MV
Depression	SV	MV	MV	MV	MV
Quality of life	SV	ND	MV	MV	SV
Physical activity	MV	MV	MV	MV	SV
Renal function	MV	MV	MV	MV	MV
Biomarkers	MV	MV	MV	MV	MV

AGES indicates Age, Gene/Environment Susceptibility—Reykjavik Study; ARIC, The Atherosclerosis Risk in Communities Study; CHS, The Cardiovascular Health Study; FHS, The Framingham Heart Study; RS, the Rotterdam Study; MI, myocardial infarction; TIA, transient ischemic attack; PVD, peripheral vascular disease; IMT, intimal-medial thickness; MRI, magnetic resonance imaging; CT, computed tomography; A, assessed; MV, measures at multiple visits; SV, measures at single visit; P, measured once on part of the cohort; ND, not done.

studies as collaborators, including at least a dozen other studies from the United States and Europe. Collaborating nonmember studies either agree to the overall CHARGE principles, or the CHARGE working group develops and negotiates a new CHARGE-compatible agreement with the nonmember studies or consortia.

Using traditional authorship criteria,<sup>15</sup> the CHARGE Research Steering Committee encourages the designation of multiple coequal first and last authors so that the authorship matches the scientific contributions of conducting and coordinating analyses from 5 complex studies. Special efforts are made to provide opportunities for young investigators. The original CHARGE consortium agreement calls for posting shared results on a public website once a manuscript is published in a journal. Recent changes in the NIH GWAS policy may affect this plan.<sup>16</sup> The consortium remains committed to the NIH GWAS policy on intellectual property.<sup>17</sup>

### Genotyping Methods

The CHARGE consortium was developed after each cohort study had contracted for their genotyping platforms and decided on the selection of the individuals to be included in the GWAS. Indeed, the 5 cohorts used 4 different platforms (Table 3), which have fewer than about 60 000 single-nucleotide polymorphisms (SNPs) in common. To maximize the availability of comparable genetic data and coverage of the genome, each cohort used recently developed methods<sup>18,19</sup> to impute for Europeans and European Americans their genotypes at each of the 2.5 million autosomal CEPH HapMap SNPs. Before imputation, individuals were excluded for low call rates or sex mismatches (Table 3). Next, criteria such as high levels of missingness, highly significant departures from Hardy-Weinberg equilibrium, or low minor allele frequencies were used to determine which SNPs to include in the imputation step. All the remaining individuals and SNPs entered the imputation process, which provided estimates for all the HapMap SNPs, including any that may have failed the data-cleaning criteria.

The ratio of the observed dosage variance to the expected binomial variance, the dosage-variance ratio, has proved to be useful metric of imputation quality. To assess accuracy of imputed genotypes, cohorts compared the imputation output with SNPs that had been previously genotyped on other platforms and that had not been used in the imputation process. In an internal analysis that compared the imputed SNPs to the actually genotyped SNPs in the RS, the mean concordance (number of concordant individuals/total number of individuals) between the imputed, and the genotyped SNPs was 0.989 for imputed SNPs with a dosage-variance ratio  $\geq 0.9$ . For ratios between 0.5 and 0.9, the concordance was 0.937; and for ratios  $\leq 0.5$ , it was 0.889. Validation efforts produced similar results in other cohorts.

### Analysis Methods

The CHARGE Analysis Committee developed a set of general plans as guidelines for all working groups. The issues include quality control of genotype data, decisions about what results to share across cohorts, formats for sharing data, strand alignments, coding of alleles, choice of covariates for

**Table 3. Genotyping Methods Used by the CHARGE Cohorts**

Study	AGES	ARIC	CHS	FHS	RS
Array type	Illumina 370CNV	Affymetrix 6.0	Illumina 370CNV	Affymetrix 500K and MIPS 50K combined	Version 3 Illumina Infinium II HumanHap550 SNP chip array
Genotyping center	NIA Genetic Laboratory	Broad Institute	Cedars-Sinai Medical Center	Affymetrix Core Laboratory	Erasmus Medical Center
Genotype calling	Illumina BeadStudio	Birdseed	Illumina BeadStudio	BRLMM	Beadstudio v0.3.10.14
Exclusion on SNPs used for imputation	Call rate <97%, HWE $P < 1 \times 10^{-6}$ , MAF <1%, Mishap $P < 1 \times 10^{-9}$ , A/T and G/C SNPs, Mismatches between Illumina, dbSNP and/or HapMap position	Call rate <90% MAF <1% pHWE < $10^{-6}$	Call rate $\leq 97\%$ , HWE $P < 10^{-5}$ , >2 replicate errors or Mendelian inconsistencies (for reference CEPH trios) heterozygote frequency=0, not in HapMap	Call rate <97%, HWE $P < 1 \times 10^{-6}$ , Mishap $P < 1 \times 10^{-9}$ , Mendelian errors > 100	Call rate <90%, No MAF/HWE filter
Exclusion on a per sample basis	Sex mismatch, Sample failure, Genotype mismatch with reference panel	Call rate <95%, sex mismatch, 1st degree relatives, genotype mismatch with reference panel, outliers based on IBS clustering or Eigenstrat	Call rate, <95%, sex mismatch, sample failure	Call rate <97%, subject heterozygosity >5 SD away from the mean, large Mendelian error rate	Call rate <97.5% sex mismatch, excess autosomal heterozygosity >0.336, outliers identified by the IBS clustering analysis
Imputation	MACH (version 1.00.16)	Mach1 v1.00.16	BIMBAM10 v0.99	MACH (version 1.00.15)	Mach 1.0
Imputation Backbone (NCBI build)	Hapmap release 22 CEU (build 36)	July 2006, phased haplotypes, HapMap release 21 (build 35)	HapMap CEU using release 21A, build 36	HapMap (release 22, build 36, CEU)	HapMap release 22 CEU (build 36)
Data handling and statistical tests	PLINK and R	PLINK, Mach2QTL	R	R packages kinship, gee, coxph	PLINK, Probable, Mach2QTL,

AGES, Age, Gene/Environment Susceptibility—Reykjavik Study; ARIC, The Atherosclerosis Risk in Communities Study; CHS, The Cardiovascular Health Study; FHS, The Framingham Heart Study; RS, the Rotterdam Study; MAF, minor allele frequency; HWE, Hardy-Weinberg equilibrium.

adjustments, detection of and correction for population structure, within-study phenotype analysis plans, between-study meta-analysis methods, and the importance of written analysis plans prior to sharing the results. The goal was to provide a flexible plan that could be adapted or adopted by working groups. For each stage in the analysis, there are several valid options available, and some are summarized briefly.

Special features of the CHARGE consortium are the large overall sample size, the population-based recruitment of cohort members, the standardized methods of data collection, and the prospective follow-up for clinical events. In case-control studies, it is not usually possible to obtain DNA from fatal cases. With the cohort design, DNA is generally available for all events, including the fatal ones; and failure-time models are recommended for associations with incident disease.

For most traits, the additive or the 1 degree-of-freedom regression model is used to assess the association between the phenotype and the number of copies of a specified allele. For many patterns of “true” associations, tests derived from this model have good power compared with other approaches.<sup>20</sup> The single regression coefficient is readily interpreted and easily used in meta-analysis. When imputed genotypes are used, the observed allele count is simply replaced by the imputation’s “estimated dosage.” Standard errors for the regression estimates are usually calculated with model-robust (‘sandwich’) methods. Routine adjustment is anticipated for age and sex though specific studies may also adjust for site (CHS, ARIC), for family relationships (FHS), or for cohort (FHS, RS). When necessary, principal components analysis is used to correct for within-study population structure.<sup>21</sup> Additionally, the method of genomic control is used to correct both within-study and meta-analyzed GWAS results for possible stratification.<sup>22</sup>

For the additive model, the regression coefficients estimate the difference in phenotype associated with each extra copy of the minor allele. Due to low power and potentially misleading results, meta-analyses are not reported for those SNPs for which the minor allele frequencies or the effective sample size, across CHARGE, is too small.<sup>23</sup> The acceptable lower threshold of minor allele frequencies depends on the total sample size for continuous traits or on the total number of events across all cohorts for dichotomous traits.

The analysis of 2.5 million SNPs across the genome poses an obvious multiple-testing problem. Before sharing results, working groups select a probability value threshold to identify a set of genotype-phenotype associations, almost all of which can be expected to replicate in similar populations. With 2.5 million tests, the use of a Bonferroni correction to control the Family-Wise Error rate at 0.05 yields a threshold probability value of  $2E-8$ . Another way to interpret this threshold is to estimate the expected number of false-positive (EFP) tests: if there are no true associations, each test contributes on average  $2E-8$  false-positives and, across the genome, yields an expected total of 0.05 false-positive results. Similarly, a threshold of  $1/2.5$  million, which equals  $4E-7$ , gives an expectation of one false-positive result for all tests. Unlike the Family-Wise Error rate interpretation, the control of EFP is not “conservative” for correlated tests.<sup>24</sup>

The CHARGE Analysis Committee recommends prespecifying a fixed probability value threshold as well as a number of tests, but the decision about the exact threshold to use is left up the working groups. The Analysis Committee has also provided power calculations for both continuous (Supplemental Figure 1) and binary phenotypes (Supplemental Figure 2).

When promising results from GWAS meta-analyses arise from SNPs that were imputed in some or all of the cohorts, genotyping the imputed markers in a sample of the existing cohort members serves to validate the imputation process. For the purpose of replication, genotyping high-signal SNPs in independent samples provides additional evidence about the presence or absence of an association. The number of SNPs and the number of independent individuals to be genotyped depend on the available resources and populations. Key follow-up efforts—resequencing high signal areas, fine mapping and functional studies—are likely to require new resources.

### Example of Coronary Heart Disease

The cohort-study methods papers provide detail about many of the phenotypes listed in Table 2. For coronary heart disease, investigators knowledgeable about the phenotype in each study decided to focus on fatal and nonfatal myocardial infarction (MI) as the primary outcome because the MI criteria differed in only trivial ways among the studies. There were some minor differences in the definition of the composite outcome of MI, fatal coronary heart disease, and sudden death, which became the secondary outcome. Only subjects at risk for an incident event were included in the analysis. MI survivors whose DNA was drawn after the event were not eligible. The primary analysis was restricted to Europeans or European Americans. Patients entered the analysis at the time of the DNA blood draw, and were followed until an event, death, loss to follow up, or the last visit. The main recommendations of the Analysis Committee were adopted, and a threshold of  $5 \times 10^{-8}$  was selected for genome-wide statistical significance. Analyses in progress include about 1700 MIs and 2300 coronary heart disease events among about 29 000 eligible patients. Each cohort conducted its own analysis, and results were uploaded to a secure share site for the fixed-effects meta-analysis. Even with this number of events (Supplemental Figure 2), power is good for only for relatively high minor allele frequencies ( $>0.25$ ) and large relative risks ( $>1.3$ ).

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

### Discussion

In thousands of published papers, the 5 CHARGE cohort studies and many of the collaborating studies have already characterized the risk factors for and the incidence and prognosis of a variety of aging-related and cardiovascular conditions. The analysis of the incident MI, for instance, is free from the survival bias typically associated with cross-sectional or case-control studies. The methodologic advantages of the prospective population-based cohort design, the

similarity of phenotypes across 5 studies, the availability of genome-wide genotyping data in each cohort, and the need for large sample sizes to provide reliable estimates of genotype-phenotype associations have served as the primary incentives for the formation of the CHARGE consortium, which includes GWAS data on about 38 000 individuals. The consortium effort relies on collaborative methods that are similar to those used by the individual contributing cohorts.

Phenotype experts who know the studies and the data well are responsible for phenotype-standardization across cohorts. The coordinated prospectively planned meta-analyses of CHARGE provide results that are virtually identical to a cohort-adjusted pooled analysis of individual level data. This approach—the within-study analysis followed by a between-study meta-analysis—avoids the human subjects issues associated with individual-level data sharing.

Editors, reviewers, and readers expect replication as the standard in science.<sup>6</sup> The finding of a genetic association in one population with evidence for replication in multiple independent populations provides moderate assurance against false-positive reports and helps to establish the validity of the original finding. In a single experiment, the discovery-replication structure is traditionally embodied in a 2-stage design. The CHARGE consortium includes up to 5 independent replicate samples as well as additional collaborating studies for some phenotype working groups, so that it would have been possible to set up analysis plans within CHARGE to mimic the traditional 2-stage design for replication. For instance, the 2 largest cohorts could have served as the discovery set and the others as the replication set. However, attaining the extremely small probability values expected in GWAS requires large sample sizes. For any phenotype, a prospective meta-analysis of all participating cohorts, with a properly selected level of genome-wide statistical significance to minimize the chance of false-positives, is the most powerful approach to finding new genuine associations for genetic loci.<sup>25</sup> When findings narrowly miss the prespecified significance threshold, genotyping individuals in other independent populations provides additional evidence about the association. For findings that substantially exceed pre-established significance thresholds, the results of a CHARGE meta-analysis effectively provide evidence of a multistudy replication.

The effort to assemble and manage the CHARGE consortium has provided some interesting and unanticipated challenges. Participating cohorts often had relationships with outside study groups that predated the formation of CHARGE. Timelines for genotyping and imputation have shifted. Purchases of new computer systems for the volume of work were sometimes necessary. Each cohort came to the consortium with their own traditions for methods of analysis, organization, and authorship policies that, while appropriate for their own work, were not always optimal for collaboration with multiple external groups. Within each cohort, the investigators had often formed working groups that divided up the large number of available phenotypes in ways that made sense locally but did not necessarily match the configuration that had been adopted by other cohorts. The Research Steering Committee has attempted to create a set of

CHARGE working groups that accommodate the needs and the conventions of the various cohorts. Transparency, disclosure, and professional collaborative behavior by all participating investigators have been essential to the process.

Resource limitations are another challenge. Grant applications that funded the original single-study genome-wide genotyping effort typically imagined a much simpler design. The CHS whole-genome study had as its primary aim, for instance, the analysis of data on 3 endpoints, coronary disease, stroke and heart failure. With a score of active phenotype working groups, the CHARGE collaboration broadened the scope of the short-term work well beyond initial expectations for all the participating cohorts.

One of the premier challenges has been communications among scores of investigators at a dozen sites. CHS and ARIC are themselves multi-site studies. To be successful, the CHARGE collaboration has required effective communications: (1) within each cohort; (2) between cohorts; (3) within the CHARGE working groups; and (4) among the major CHARGE committees. In addition to the traditional methods of conference calls and email, the CHARGE “wiki,” set up by Dr J. Bis (Seattle, Wash), has provided a crucial and highly functional user-driven website for calendars, minutes, guidelines, working group analysis plans, manuscript proposals, and other documents. In the end, there is no substitute for face-to-face meetings, especially at the beginning of the collaboration, and this complex meta-organization has benefited from several CHARGE-wide meetings.

The major emerging opportunity is the collaboration with other studies and consortia. Many working groups have already incorporated nonmember studies into their efforts. Several working groups have coordinated submissions of initial manuscripts with the parallel submission of manuscripts from other studies or consortia. Several working groups have embarked on plans for joint meta-analyses between CHARGE and other consortia. CHARGE has tried to acknowledge and reward the efforts of champions, who assume leadership responsibility for moving these large complex projects forward and who are often hard-working young investigators, the key to the future success of population science.

The CHARGE Consortium represents an innovative model of collaborative research conducted by research teams that know well the strengths, the limitations, and the data from 5 prospective population-based cohort studies. By leveraging the dense genotyping, deep phenotyping and the diverse expertise, prospective meta-analyses are underway to identify and replicate the major common genetic determinants of risk factors, measures of subclinical disease, and clinical events for cardiovascular disease and aging.

### Acknowledgments

The authors thank Drs Josh Bis, Nicole Glazer, and Ken Rice for comments on earlier drafts. A full list of investigators from the CHARGE cohorts appears at: <http://web.chargeconsortium.com>.

### Sources of Funding

Age, Gene/Environment Susceptibility—Reykjavik Study has been funded by NIH contract N01-AG-12100, the NIA Intramural Re-

search Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament).

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, and R01HL087641; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Cardiovascular Health Study: The research reported in this article was supported by contract numbers N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295 and R01 HL087652 from the National Heart, Lung, and Blood Institute, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at <http://www.chs-nhlbi.org/pi.htm>. Framingham Heart Study: From the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278), and by grants from the National Institute of Neurological Disorders and Stroke (NS17950; PAW) and the National Institute of Aging, (AG08122, AG16495; PAW). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. Rotterdam Study: The GWA database of the Rotterdam Study was funded through the Netherlands Organization of Scientific Research NWO (nr. 175.010.2005.011). The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University, Rotterdam; the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

## Disclosures

None.

## References

- Ioannidis JPA, Nizani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet*. 2001;29:306–309.
- Zeggini E, Scott L, Saxena R, Boight BF; for the Diabetes and large-scale replication and Meta-analysis (DIAGRAM) Consortium. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008;40:638–645.
- The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature*. 2007;447:661–678.
- McPherson R, Pertsemlidis A, Kavasslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Hansen AT, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491.
- Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, Jonasdóttir A, Jonasdóttir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson D, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiassdóttir S, Jonsdóttir T, Paulson S, Einarsdóttir H, Gunnarsdóttir S, Gylafalson A, Vaccarino V, Hooper WG, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorleifsson G, Thorsteinsdóttir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491–1493.
- NCI-NHGRI Working Group on Replication in Association Studies. Replicating genotype-phenotype associations. *Nature*. 2007;447:655–660.
- Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, Vineis P, Balding DJ, Chokkalingam A, Dolan SM, Flanders W, Higgins JPT, McCarthy M, McDermott DH, Page GP, Rebbeck TR, Seminara D, Khoury MJ. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol*. 2008;37:120–132.
- Harris T, Launer L, Eiriksdóttir G, Kjartansson O, Jonsson PV, Sigurdsson G, Thorgeirsson G, Aspelund T, Garcia MF, Hoffman HJ, Gudnason V. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol*. 2007;165:1076–1087.
- The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol*. 1989;129:687–702.
- Fried LP, Borhani NO, Enright P, Furberg C, Gardin J, Kronmal R, Kuller LH, Manolio T, Mittelmark M, Newman A, O'Leary DH, Psaty B, Rautaharju P, Tracy RP, Weiler P. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.
- Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health*. 1951;41:279–286.
- Feinleib M, Kannel W, Garrison R, McNamara P, Castelli W. The Framingham Offspring Study: design and preliminary data. *Prev Med*. 1975;4:518–525.
- Splansky G, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin E, D'Agostino Sr RB, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165:1328–1335.
- Hofman A, Breteler MMB, van Duijn CM, Krestin GP, Pols HA, Stricker BHC, Tiemeier H, Uitterlinden AG, Vingerling JR, Witteman JCM. The Rotterdam Study: objectives and design update. *Eur J Epidemiol*. 2007;22:819–829.
- International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication, updated October 2007. <http://www.icmje.org>; Accessed Sept 19, 2008.
- National Institutes of Health. Modification to genome-wide association studies (GWAS) data access, Aug 28, 2008. [http://grant.nih.gov/grants/gwas/data\\_sharing\\_policy\\_modifications\\_20080828.pdf](http://grant.nih.gov/grants/gwas/data_sharing_policy_modifications_20080828.pdf); accessed Aug 29, 2008.
- National Institutes of Health. Policy for sharing of data Obtained in NIH supported or conducted genome-wide association studies (GWAS). *Federal Register*. 2007;72:4920–4927.
- Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet*. 2007;3:e114.
- Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*. 2006;S79:2290; <http://www.sph.umich.edu/csg/abecasis/MACH/>; accessed September 19, 2008.
- Lette G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol*. 2007;31:358–362.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–909.
- Devlin B. Genomic control for association studies. *Biometrics*. 1999;55:997–1004.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9:356–369.
- Gordon A, Glazko G, Qiu X, Yakovlev A. Control of the mean number of false discoveries, Bonferroni and stability of multiple testing. *Ann Appl Statistics*. 2007;1:179–190.
- Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome wide association studies. *Nat Genet*. 2006;38:209–213.

**Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)  
Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies  
From 5 Cohorts**

Bruce M. Psaty, Christopher J. O'Donnell, Vilmundur Gudnason, Kathryn L. Lunetta, Aaron R. Folsom, Jerome I. Rotter, André G. Uitterlinden, Tamara B. Harris, Jacqueline C.M. Witteman  
and Eric Boerwinkle  
on Behalf of the CHARGE Consortium

*Circ Cardiovasc Genet.* 2009;2:73-80  
doi: 10.1161/CIRCGENETICS.108.829747

*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue,  
Dallas, TX 75231

Copyright © 2009 American Heart Association, Inc. All rights reserved.  
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the  
World Wide Web at:

<http://circgenetics.ahajournals.org/content/2/1/73>

Data Supplement (unedited) at:

<http://circgenetics.ahajournals.org/content/suppl/2009/02/18/2.1.73.DC1>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Genetics* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation: Cardiovascular Genetics* is online at:  
<http://circgenetics.ahajournals.org/subscriptions/>

## Supplemental Material

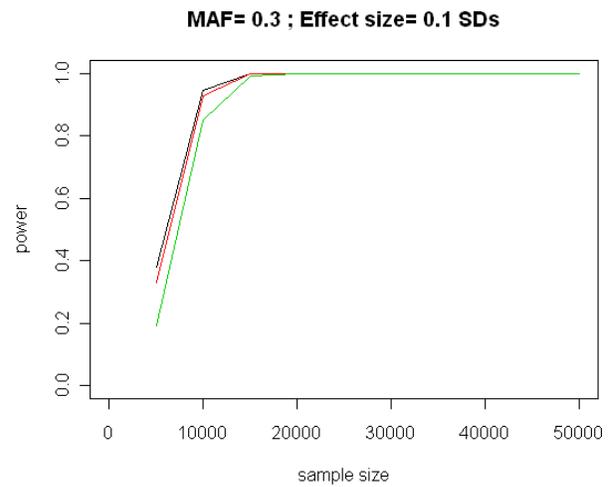
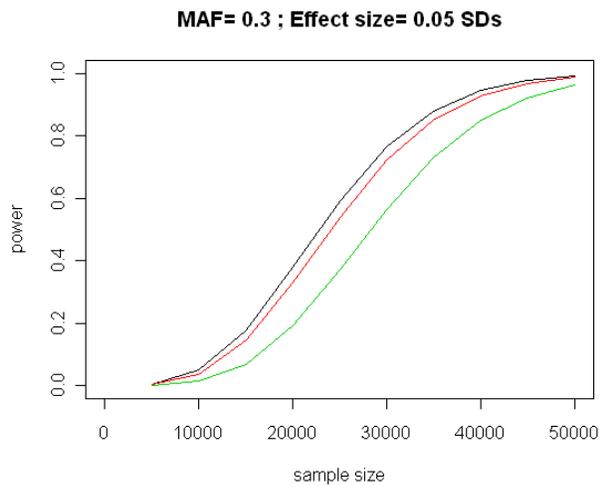
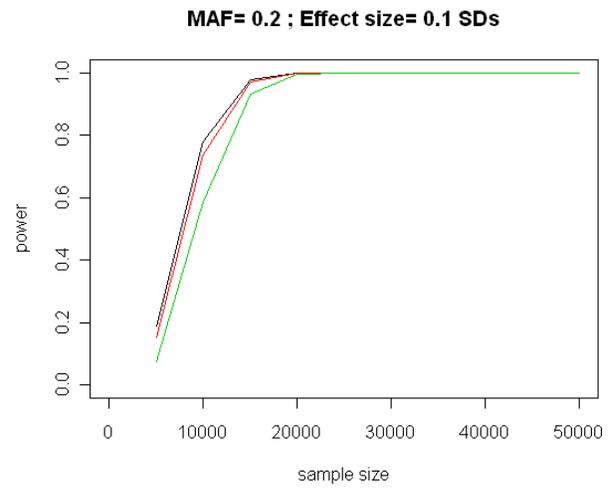
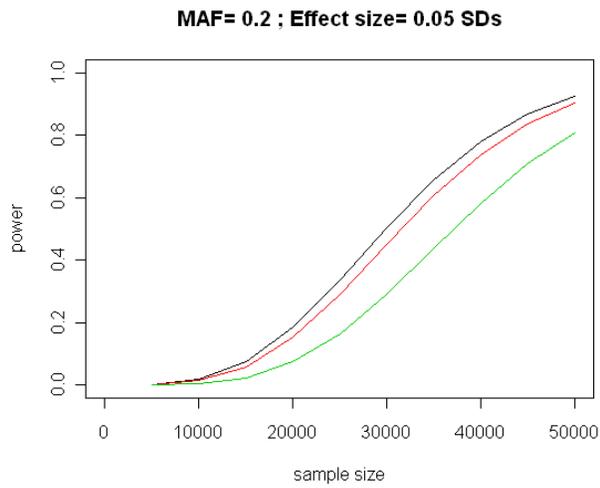
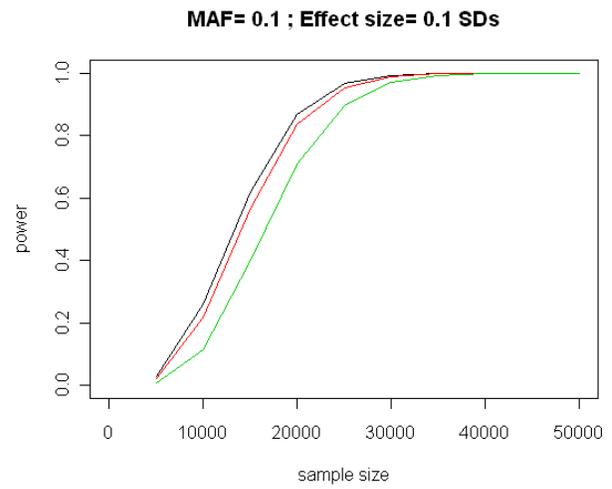
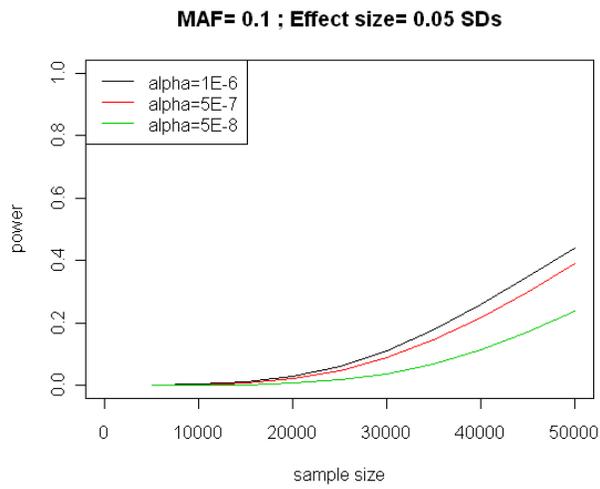
Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JCM, Boerwinkle E on behalf of the CHARGE Consortium. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ Cardiovasc Genet*.

## Supplemental figure legends

Supplemental Figure 1. Statistical power for detecting associations with a continuous phenotype. Abbreviations:  $n$ , number of subjects in all cohorts combined; Effect Size in units of standard deviation; MAF = minor allele frequency. Two-sides alpha levels of  $1E-6$  (black),  $5E-7$  (red), and  $5E-8$  (green).

Supplemental Figure 2. Statistical power for detecting associations with a binary phenotype. Abbreviations:  $n$ , number of clinical events from all cohorts; RelRisk = relative risk; MAF = minor allele frequency. Two-sides alpha levels of  $1E-6$  (black),  $5E-7$  (red), and  $5E-8$  (green).

(Supplemental Figure 1)



(Supplemental Figure 2)

