Genome-Wide Profiling of Blood Pressure in Adults and Children

Hendrik R. Taal, Germaine C. Verwoert, Ayse Demirkan, A. Cecile J.W. Janssens, Kenneth Rice, Georg Ehret, Albert V. Smith, Ben F.J. Verhaaren, Jacqueline C.M. Witteman, Albert Hofman, Meike W. Vernooij, Andre G. Uitterlinden, Fernando Rivadeneira, M. Arfan Ikram, Daniel Levy, Albert J. van der Heijden, for the Cohort for Heart and Aging Research in Genome Epidemiology and Early Genetics and Lifecourse Epidemiology consortia, Vincent W.V. Jaddoe, Cornelia M. van Duijn

Abstract—Hypertension is an important determinant of cardiovascular morbidity and mortality and has a substantial heritability, which is likely of polygenic origin. The aim of this study was to assess to what extent multiple genetic variants contribute to blood pressure regulation in both adults and children and to assess overlap in variants between different age groups, using genome-wide profiling. Single nucleotide polymorphism sets were defined based on a meta-analysis of genome-wide association studies on systolic blood pressure and diastolic blood pressure performed by the Cohort for Heart and Aging Research in Genome Epidemiology (n=29 136), using different P value thresholds for selecting single nucleotide polymorphisms. Subsequently, genetic risk scores for systolic blood pressure and diastolic blood pressure were calculated in an independent adult population (n=2072) and a child population (n=1034). The explained variance of the genetic risk scores was evaluated using linear regression models, including sex, age, and body mass index. Genetic risk scores, including also many nongenome-wide significant single nucleotide polymorphisms, explained more of the variance than scores based only on very significant single nucleotide polymorphisms in adults and children. Genetic risk scores significantly explained ≤1.2% (P=9.6*10^{-8}) of the variance in adult systolic blood pressure and 0.8% (P=0.004) in children. For diastolic blood pressure, the variance explained was similar in adults and children (1.7% [P=8.9*10^{-10}] and 1.4% [P=3.3*10^{-7}], respectively). These findings suggest the presence of many genetic loci with small effects on blood pressure regulation both in adults and children, indicating also a (partly) common polygenic regulation of blood pressure throughout different periods of life. (Hypertension. 2012;59:241-247.)

Key Words: genome-wide association ■ genome-wide profiling ■ genetic risk scores ■ blood pressure ■ hypertension

Elevated blood pressure is an important risk factor for stroke and ischemic heart disease and is estimated to contribute to half of the global risk for cardiovascular disease.1-2 Antihypertensive treatment has been an effective combination these variants may nevertheless explain a substantial proportion of hypertension in the general population. Several common genetic variants associated with blood pressure have been identified through genome-wide association studies (GWASs) in adult populations, only explaining ≈1% of the variance of blood pressure.7-8

Despite the large size of the consortia used for gene discovery, many common variants with small effects on blood pressure remain unidentified. Although their individual associations do not reach genome-wide significance, in combination these variants may nevertheless explain a substantial
within-study associations were combined by prospective meta-analysis. The methods have been described in detail previously.7

Next, SNPs were selected using the results from the meta-analyses of GWASs on SBP in the discovery sample, on the basis of their nominal probability value (P_{discovery}) for association with SBP, using different P_{discovery} thresholds, ranging from 1.0×10⁻⁷ to 1.0. Subsequently, these sets of SNPs, with different P_{discovery} thresholds, were used to calculate the genetic scores in the target samples. The same was done using the results from the meta-analysis of GWAS on DBP, creating separate genetic risk scores for DBP.

For each individual in the 2 independent target samples, the genetic risk scores were calculated by multiplying the number of risk alleles per SNP (0, 1, or 2) with the effect size of that SNP in the discovery meta-analysis (weighted approach), summed over all of the SNPs in the set and divided by the number of SNPs in the considered set. The calculations of individual scores for each set of SNPs were performed using the PLINK (version 1.07) software, specifically by "profile scoring" option.

Subsequently, linear regression models were used to test the association between the individual genetic risk scores and SBP and DBP in the target samples. For subjects using antihypertensive medication, we added 10 mm Hg to the observed SBP values and 5 mm Hg to the observed DBP values. Similar to the discovery analysis, the models were adjusted for sex, age, and body mass index. For analyzing the explained variance for adult hypertension (SBP ≥140 mm Hg, DBP ≥90 mm Hg, or use of antihypertensive medication) and childhood hypertension (SBP or DBP >95th percentile for age, sex, and height11), we used logistic regression. We performed sensitivity analyses in the adults excluding subjects meeting the criteria for hypertension in the Rotterdam Study III. We have repeated the main analyses after removing SNPs in high linkage disequilibrium (LD), using the LD-pruning option in the Plink software.12 We pruned the data considering a window of 200 SNPs, removing one of a pair of SNPs if the LD is >0.25 and shifting the window 5 SNPs forward to repeat the procedure.

For both the Rotterdam Study III and the Generation R Study, regression analyses were performed in SPSS 17.0 for Windows (SPSS Inc, Chicago, IL). The difference of the explained variance in regression analyses were performed in SPSS 17.0 for Windows (SPSS Inc, Chicago, IL). The difference of the explained variance in the null (without genetic risk score) and alternative model (including a genetic risk score) was considered as the variance explained by the genetic score. A genetic risk score with a P<0.05 in the model was considered as significantly associated with the trait. We also assessed the difference between 2 subsequent models, both including a genetic risk score using the Akaiki Information Criterion (AIC).13

In the online Data Supplement (please see http://hyper.ahajournals.org), we describe the discovery and target samples, their respective data collection procedures, and quality control. The approaches for sensitivity analyses using unrelated traits as outcome and analysis using randomly selected SNP sets or an SNP set based on a biological pathway to calculate a genetic risk score are also presented in the online Data Supplement.

**Methods**

**Genome-Wide Profiling**

In genome-wide profiling, for a certain trait, genetic risk scores are constructed using data from a “discovery sample.” Sets of common variants to calculate genetic risk scores consist of all SNPs achieving a certain significance threshold (P_{discovery} threshold) in the discovery sample. In an independent “target sample,” a subject’s genetic risk score is computed across all of the SNPs with P value lower than the P_{discovery} threshold. The subject’s genotype (coded 0/1/2) is multiplied by the regression coefficient for that SNP as estimated in the discovery sample and divided by the total number of SNPs in that set. This risk score is calculated for all of the subjects in the independent target sample. Subsequently, an unbiased estimate of the variance explained by the genetic risk score is obtained by evaluating the increase in explained variance of the trait when adding the genetic risk score to a baseline model explaining that trait. The method has been described previously in detail.9

In our study, we used systolic blood pressure (SBP) and diastolic blood pressure (DBP) and hypertension as traits of interest. The discovery sample was the Cohort for Heart and Aging Research in Genome Epidemiology consortium, with a total sample size of 29,136 participants.7 We used 2 Dutch target samples, 1 adult sample (Rotterdam Study III) and 1 child sample (Generation R Study). In the discovery sample, within each cohort of the Cohort for Heart and Aging Research in Genome Epidemiology Study, regression models were fitted for SBP and DBP separately and allele dosage, using an additive genetic model. The models were adjusted for sex, age, age squared, and body mass index. Subsequently, the proportion of blood pressure. The extent to which unidentified common variants explain the missing heritability in blood pressure and other polygenic traits is an open and important question.

Recently, it has been shown that the presence of multiple variants affecting polygenic traits can be demonstrated by constructing genome-wide prediction models (genetic risk scores) of common variants.9–10 In a polygenic disease model, the more markers that are used in the model, the better the disease is predicted. Such a model also implies that everybody in the population carries a substantial number of risk variants with small effects on the disease, but patients carry more of these variants than nondiseased people. This has been demonstrated in a recent study of schizophrenia, showing that one can predict disease using both genome-wide significant and nonsignificant single nucleotide polymorphisms (SNPs), covering a large part of the genome.9

This approach can also be used to evaluate the evidence of overlap in genes affecting a continuous outcome as blood pressure in different age groups. In metabolic diseases, such as diabetes mellitus and blood pressure, there is increasing interest in the role of genetic determinants of blood pressure and other risk factors of cardiovascular disease to improve prevention of chronic disease and to identify targets for therapeutic interventions.

One may argue that many genes regulating blood pressure maintenance are similar across age groups. We used genome-wide profiling to evaluate to what extent multiple common genetic variants influence blood pressure in adults and, secondly, to test whether there is overlap in genes contributing to blood pressure levels in children and adults, which might indicate whether there is a common polygenic regulation of blood pressure throughout different periods of life.

Demographic data of the Cohort for Heart and Aging Research in Genome Epidemiology consortium have been published previously.7 Table 1 shows the baseline characteristics for the Rotterdam Study III and the Generation R Study. The median age in Rotterdam Study III was 56.0 years (95% range: 47.7–62.3 years), in the Generation R Study the median age was 6.0 years (95% range: 5.8–6.7 years). Figure 1A and 1B shows the increase in explained variance of SBP by the genetic risk scores for a range of different P_{discovery} thresholds, in the adult and child target samples, respectively. When considering risk scores based on sets of SNPs with low P values in the discovery sample (up to P_{discovery} <1.0×10⁻⁴), the risk scores significantly explained ≥0.3% of variance in SBP in adults (P<0.01). These scores were nonsignificant in children, increasing the explained variance only up to 0.1% of variance in SBP. In DBP, the risk scores explained up to 0.7% (P<0.01) of the variance in adults.
Table 1. Subject Characteristics of the Generation R Study and the Rotterdam Study III

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Rotterdam Study III (n=2078)</th>
<th>Generation R Study (n=1034)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>56.0 (47.7–62.3)</td>
<td>6.0 (5.8–6.7)</td>
</tr>
<tr>
<td>Female, %</td>
<td>56.1</td>
<td>52.2</td>
</tr>
<tr>
<td>BMI, weight (kg)/length (cm)²</td>
<td>26.9 (21.6–36.7)</td>
<td>15.9 (13.9–18.4)</td>
</tr>
<tr>
<td>Mean systolic blood pressure, mm Hg</td>
<td>132.7 (19.2)</td>
<td>102.5 (8.1)</td>
</tr>
<tr>
<td>Mean diastolic blood pressure, mm Hg</td>
<td>82.6 (11.1)</td>
<td>60.5 (7.2)</td>
</tr>
<tr>
<td>Subjects with hypertension, %*</td>
<td>47.3</td>
<td></td>
</tr>
<tr>
<td>Subjects using antihypertensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medication, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children referred to nephrology for</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypertension, %†</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Prevalent type 2 diabetes mellitus, %</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>Serum total cholesterol, mmol/L</td>
<td>5.6 (1.1)</td>
<td></td>
</tr>
<tr>
<td>HDL cholesterol, mmol/L</td>
<td>1.5 (0.5)</td>
<td></td>
</tr>
</tbody>
</table>

Values are mean (SD) or median (95% range). HDL, high-density lipoprotein; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure.

*Adults are defined as SBP >140 mm Hg, DBP >90 mm Hg or use of antihypertensive medication and children as SBP or DBP >95th percentile for sex, age, and height.12

†Children were referred to pediatric nephrology department when the lowest blood pressure measurement was higher than the 99th percentile for sex, length, and age.12

The variance in systolic blood pressure for adults and children, respectively. We evaluated the difference in explained variance (percentage) of systolic blood pressure, when adding the genetic risk scores for different P value thresholds, to the baseline model for systolic blood pressure, including body mass index, sex, and age as covariates. The baseline model explained 12.8% and 5.2% of the variance in systolic blood pressure for adults and children, respectively. We evaluated the difference in explained variance between 2 subsequent models by calculating the Akaike Information Criterion (AIC) of each model. The difference in AIC follows a χ² distribution with 1 degree of freedom, from which the P value was derived. *P<0.05, **P<0.001. A, Rotterdam Study III. B, Generation R Study.
created random genetic risk scores of sufficient size (~565k SNPs, similar to the SNP set with $P_{\text{discovery}} < 0.3$) and evaluated the additional explained variance. The random genetic risk scores showed a significant increase in explained variance when added to the baseline model (see Table 2). The additional explained variance from the random genetic risk scores was 0.1% to 0.4% lower as compared with the original genetic risk scores (see Table 2). We also tested whether a set of SNPs from a biological pathway would lead to a higher increase in explained variance than SNP sets with a low $P_{\text{discovery}}$ threshold. We used SNPs in the fibroblast growth factor pathway as described by Tomaszewski et al. This genetic risk score did not explain additional variance in all of the phenotypes except for adult hypertension based on the systolic blood pressure score (increase in explained variance: 0.3; $P=0.014$; see Table 3).

In a strictly normotensive population, the additional explained variance of the genetic risk score including the most

---

**Figure 2.** Increase in explained variance in diastolic blood pressure by genetic risk scores. Bars represent the increase in explained variance (%) of diastolic blood pressure, when adding the genetic risk scores for different $P$ value thresholds, to the baseline model for diastolic blood pressure, including body mass index (BMI), sex and age as covariates. The baseline model explained 8.4% and 1.4% of the variance in diastolic blood pressure for adults and children, respectively. We evaluated the difference in explained variance between 2 subsequent models by calculating the Akaike Information Criterion (AIC) of each model. The difference in AIC follows a $\chi^2$ distribution with 1 degree of freedom, from which the $P$ value was derived. *$P<0.05$; **$P<0.001$. A, Rotterdam Study III. B, Generation R Study.

---

**Figure 3.** Increase in explained variance in hypertension by genetic risk scores. Bars represent the increase in explained variance (%) of hypertension in adults (systolic blood pressure [SBP] $>140$ mm Hg; diastolic blood pressure [DBP] $>90$ mm Hg or the use of antihypertensive medication), when adding systolic and diastolic genetic risk scores for different $P$ value thresholds, to the baseline model for systolic blood pressure, including body mass index (BMI), sex, and age as covariates. The baseline model explained 15.9% of the variance in hypertension in adults. We evaluated the difference in explained variance between 2 subsequent models by calculating the Akaike Information Criterion (AIC) of each model. The difference in AIC follows a $\chi^2$ distribution with 1 degree of freedom, from which the $P$ value was derived. *$P<0.05$; **$P<0.001$. A, Systolic blood pressure scores. B, Diastolic blood pressure scores.
increased the variance explained by that score. The results of these analyses are shown in the supplement (Figures S5A and S5B, S6A and S6B, and S7A and S7B).

### Discussion

Our findings indicate that, in addition to the blood pressure variants now identified, large numbers of common genetic variants affecting blood pressure remain to be identified, and these variants explain a significant part of the variance in blood pressure in adults and children. These nonsignificant, unidentified SNPs together explain a larger part of the variance than the genome-wide significant SNPs only. We also showed that adult-based genetic risk scores explained variance in blood pressure in children. This indicates not only that there is a polygenic effect on blood pressure in children, but, more importantly, it indicates that there is overlap in variants involved with blood pressure maintenance in adults and children and that these variants act throughout life.

In this study, we did not remove any SNPs that are in high LD with each other. Pruned analyses, as presented in the online supplemental Material section, do not change our conclusion that adding more nongenome-wide significant SNPs to genetic risk scores increases the variance explained by these scores. However, as expected, removing SNPs by LD pruning results in a reduction of the variance explained. Because SNPs in LD are removed randomly, in many cases informative SNPs are taken out of the analyses. Therefore, a pruned analysis is expected to underestimate the true effect on the explained variance by the genetic risk scores.

The additive explained variance of genetic risk scores on blood pressure is maximizing around the $P_{\text{discovery}}$ threshold of 0.3 and does not increase with more liberal thresholds. This genetic risk score includes $>550$k SNPs. This result suggests that SNPs with a $P_{\text{discovery}}$ value $<0.3$ in the discovery sample add to the explained variance of blood pressure and that many common variants associated with blood pressure regulation have not been identified yet. SNPs with a $P_{\text{discovery}}$ $>0.3$ are unlikely to be associated with blood pressure. Genetic risk scores of a similar size, consisting of randomly selected SNPs, still resulted in a significant increase in explained variance when added to the baseline model without a genetic risk score. The increase in explained variance based on the random genetic risk scores was lower than the increase based on the original genetic risk score models, although this difference was small and statistically significant in only half of the presented phenotypes. This result suggests that a sufficiently large number of SNPs tags many genes throughout the genome, which influence blood pressure regulation. These findings are in line with our hypothesis that blood pressure is a polygenic trait and that there are many more genes involved with blood pressure than are found so far in GWASs. Currently, the advantage of using genetic risk scores based on SNPs selected on their probability value in a GWAS discovery sample, as compared with genetic risk scores based on a random set of SNPs, seems to be limited. Larger GWAS discovery samples with identification of new common and rare SNPs might lead to higher explained variance.

### Table 2. Additional Explained Variance by a Random Genetic Risk Score vs Original Genetic Risk Score

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Additional Explained Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random Genetic Risk Score, %</td>
</tr>
<tr>
<td>Adult systolic blood pressure (Rotterdam Study III)</td>
<td>1.0†</td>
</tr>
<tr>
<td>Adult diastolic blood pressure (Rotterdam Study III)</td>
<td>1.3†</td>
</tr>
<tr>
<td>Child systolic blood pressure (Generation R)</td>
<td>0.7*</td>
</tr>
<tr>
<td>Child diastolic blood pressure (Generation R)</td>
<td>1.4†</td>
</tr>
<tr>
<td>Hypertension—systolic blood pressure scores (Rotterdam Study III)</td>
<td>1.1†</td>
</tr>
<tr>
<td>Hypertension—diastolic blood pressure scores (Rotterdam Study III)</td>
<td>1.7†</td>
</tr>
</tbody>
</table>

Random genetic risk score is the genetic risk score calculated on single nucleotide polymorphism (SNP) set containing ~565k SNPs, which were randomly selected out of the discovery meta-analysis, irrespective of their $P$ value for association. Original genetic risk score is the genetic risk score calculated on an SNP set with a $P_{\text{discovery}}$ threshold of 0.3. This SNP set contains ~565k SNPs and showed the largest increase in explained variance. The $P$ value for difference between the models was obtained by calculating the difference in Akaike Information Criterion between the random model and the original model. This difference follows a $\chi^2$ distribution with 1 degree of freedom.

* $P<0.05$.
† $P<0.001$.

### Table 3. Additional Explained Variance by a Genetic Risk Score Based on Fibroblast Growth Factor Signaling Pathway

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Additional Explained Variance by Signaling Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult systolic blood pressure (Rotterdam Study III)</td>
<td>0.1</td>
</tr>
<tr>
<td>Adult diastolic blood pressure (Rotterdam Study III)</td>
<td>0.0</td>
</tr>
<tr>
<td>Child systolic blood pressure (Generation R)</td>
<td>0.1</td>
</tr>
<tr>
<td>Child diastolic blood pressure (Generation R)</td>
<td>0.1</td>
</tr>
<tr>
<td>Hypertension—systolic blood pressure scores (Rotterdam Study III)</td>
<td>0.3*</td>
</tr>
<tr>
<td>Hypertension—diastolic blood pressure scores (Rotterdam Study III)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

* $P<0.05$. 
Although there have been several mutations described causing dominant, monogenic forms of hypertension and more of such rare variants may still be undiscovered, our results support the hypothesis that hypertension is a polygenic disease, which is in part explained by a large number of genes regulating blood pressure. In our adult population, genetic risk scores, including large numbers of SNPs, explain the largest proportion of variance in blood pressure, indicating the involvement of multiple genes in blood pressure regulation.

Risk scores containing highly significant SNPs, identified in large-scale genome wide association meta-analyses in adults, were significantly associated with blood pressure in adults but not in children. There are several explanations possible for this finding, including a smaller number of subjects and lower power in the children cohort. However, this finding might also indicate that the genome-wide significant SNPs found so far are related to late-onset pathology. It has been long hypothesized that there is a common polygenic regulation of blood pressure in adults and children. This is the first study showing evidence of such a mechanism.

It has been shown in literature that blood pressure tracks from childhood to adulthood. This study indicates that genes are explaining a part of the blood pressure tracking over life. We show that the same set of genes, based on an adult discovery sample, explain part of the variation in blood pressure in both adults and children. We also showed that these SNP sets explain variation in hypertension in adults, indicating also a polygenic origin of hypertension. It has been shown that high blood pressure in childhood predisposes to hypertension in adulthood. Adult-based genetic risk scores do not explain variance in childhood hypertension in children significantly. This fits with the common view that causes of juvenile hypertension are different from adult hypertension.

The percentage of explained variance by genetic risk scores is still low, although the heritability has been shown to be substantial, yet compared with the variance explained by the genome-wide significant SNPs on blood pressure found so far, there is a 4- to 5-fold increase in explained variance of blood pressure in our target samples.

In the coming years, the variance explained by polygenic models may be improved further, using technology, such as whole genome sequencing, which can be used to identify low-frequency variants. We used common variants only (minor allele frequency: >0.01) to create genetic risk scores. Low-frequency variants may add to the variance explained by these genetic risk scores. Also, we assumed an additive model, similar to the discovery analysis. We have to recognize that the biology of the genes involved in blood pressure regulation and possible interactions between these genes are unknown. Another possibility would be to construct genetic risk scores based on SNPs included in candidate biological pathways. A genetic risk score including SNPs from the fibroblast growth factor signaling pathway seemed to explain a larger proportion of explained variance in hypertension as compared with a genetic risk score including a similar number of SNPs based on the previous top SNPs from the GWAS. This indicates that previous knowledge on biological models and underlying mechanisms might improve the explained variance by genetic risk scores. Alternative methods of constructing genetic risk scores may be better when further research reveals more of the underlying genetic biology of blood pressure regulation.

Specific common variants that are associated with blood pressure still need to be identified. Much research is still needed to identify more and specific genes associated with blood pressure regulation in adults. If these common variants overlap with blood pressure regulation in children, they could provide clues for early etiology of hypertension.

**Perspectives**

At this stage, individual prediction is not yet feasible. Without a doubt, the prediction of blood pressure will improve and might contribute to predicting high blood pressure in the future. Genetic profiling might be a way of identifying subgroups at genetically high risk for increased blood pressure at a population level, but whether it will be enough for personalized medicine and early treatment of people at risk for high blood pressure (and possibly also other risk factors for cardiovascular disease) remains to be determined. Our study shows that this may require the identification of many more common variants with small effects on blood pressure.

**Acknowledgments**

The Generation R study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty of Social Sciences of Erasmus University Rotterdam; the Municipal Health Service Rotterdam area, Rotterdam; the Rotterdam Homecare Foundation, Rotterdam; and the Stichting Trombosecentrum en Artsenlaboratorium Rijnmond, Rotterdam. We gratefully acknowledge the contributions of general practitioners, hospital, midwives, and pharmacies in Rotterdam.

**Sources of Funding**

This Rotterdam Study III support for genotyping was provided by The Netherlands Organization for Scientific Research (Netherlands Organisation for Scientific Research Groot, 175.010.2005.011, 911.03.012) and Research Institute for Diseases in the Elderly (RIDE). This study was supported by The Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research project No. 050-060-810. This study was further financially supported by The Netherlands Heart Foundation grant 2009B102 (to B.F.J.V. and M.A.I.). The first phase of the Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam; Erasmus University Rotterdam; and The Netherlands Organization for Health Research and Development (ZonMW). Additional support was provided by a grant from the Dutch Kidney Foundation (C08.2251; RT). The funding sources of other cohorts in the Cohort for Heart and Aging Research in Genome Epidemiology consortium are described in the online Data Supplement (please see http://hyper.ahajournals.org).

**Disclosures**

None.
References


ONLINE SUPPLEMENT

Genome-wide profiling of blood pressure in adults and children

Hendrik R. Taal, MD, Germaine C. Verwoert, MSc, Ayse Demirkan, MSc, A. Cecile J.W. Janssens, PhD, Kenneth Rice, PhD, Georg Ehret, MD, Albert V. Smith, PhD, Ben F.J. Verhaaren, MD, MSc, Jacqueline C.M. Witteman, MD, PhD, Albert Hofman, MD, PhD, Meike W. Vernooij, MD, PhD, Andre G. Uitterlinden, MD, PhD, Fernando Rivadeneira, MD, PhD, M. Arfan Ikram, MD, PhD, Daniel Levy, MD, PhD, Albert J. van der Heijden, MD, PhD, for the CHARGE and EAGLE consortia, Vincent W.V. Jaddoe, MD, PhD§, Cornelia M. van Duijn, PhD§

§ Authors contributed equally to this work
Expanded material and methods

Discovery Sample

The discovery sample consisted of a meta-analysis of genome wide association studies on blood pressure in 6 cohort studies conducted in the CHARGE-consortium (the Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study, the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), Rotterdam Study-I (RS-I), Rotterdam Study-II (RS-II)) \(^1\). Blood pressure was measured using a random zero sphygmanometer in ARIC, CHS, RS-1 and RS-II and using a mercury sphygmanometer in AGES and FHS. Blood pressure was measured in sitting position, except for AGES where it was measured in supine position. Quality control of the genotype data as well as the statistical methods and the results were published previously \(^2\). Briefly, each cohort applied quality control before imputing their data, excluding SNPs with a low minor allele frequency (<0.01), SNPs who violated the Hardy Weinberg equilibrium (P<1x10\(^{-6}\)) and SNPs with low call rates (<95-98%).

In brief, a meta-analysis of genome wide association studies on blood pressure in adults was performed. The total sample size of the meta-analysis was 29,136. For participants who were taking anti-hypertensive medication 10 mmHg was added to the observed SBP values and 5 mmHg to the observed DBP values \(^3\). Genotyping was performed on Illumina and Affymetrix platforms and a clean dataset was used for imputation to a HapMap set of ~ 2.5 million SNPs using MACH v0.99 or BIMBAM10.

Target Samples

Rotterdam Study III (RS-III)

The RS is a prospective population-based cohort study comprising 7,983 subjects aged 55 years or older. Participants completed an interview at home and at the research center, where participants were subsequently examined. Baseline data were collected between 1990 and 1993. In 1999, inhabitants who turned 55 years of age or moved into the study district since the start of the study were invited to participate in an extension of the Rotterdam Study (RS-II) of whom 3011 participated (67% response rate). In 2006 a further extension of the cohort was initiated in which 3,932 subjects were included (Rotterdam Study-III), aged 45 years and older, out of 6,057 subjects invited, living in the Ommoord district \(^4\).

At the research center, two seated blood pressure measurements at the right brachial artery were obtained with a random zero sphygmomanometer. The mean of two consecutive measurements was used in association analyses. Similar to the discovery set, anti-hypertensive medication was corrected for by adding 10 mmHg to observed SBP values and 5 mmHg to DBP values \(^3\).

Genotyping was done using the Illumina Human 610 Quad Arrays and for imputation of genotypes to the HapMap set of approximately 2.5 million SNPs, MACH was used. Prior to imputation, SNPs with a low minor allele frequency (<0.01), SNPs which violated the Hardy Weinberg equilibrium (P<1x10\(^{-6}\)) and SNPs with low call rates (<95%) were excluded. Blood pressure measurements and GWAS-data were available for 2,072 Caucasian subjects. No overlap consists between participants from the Rotterdam Study used in the discovery study and persons from the Rotterdam Study III, examined in the current study. The Medical Ethics Committee of the Erasmus Medical Center Rotterdam has approved the study.
**Generation R Study**

The Generation R Study is a population-based prospective cohort study from fetal life until young adulthood in Rotterdam, The Netherlands. Written informed consent was obtained from the mothers at enrolment and one of the parents at their children’s follow-up visit at the age of six years. The Medical Ethics Committee of the Erasmus Medical Center Rotterdam has approved the study.

Blood pressure measurements in children were conducted around the age of six years in a dedicated research center in the Erasmus Medical Center, Rotterdam, the Netherlands. Blood pressure was measured at the right brachial artery in supine position, four times with one minute intervals. Mean systolic and diastolic blood pressure were defined as the mean of all available measurements, excluding the first.

DNA was collected from cord blood samples. Cord blood for DNA isolation was available for 59% of all participating children of the Generation R cohort. Genotyping was carried out using the Illumina Human 610 Quad Arrays. Imputation was performed using MACH. Prior to imputation, SNPs with a low minor allele frequency (<0.01), SNPs which violated the Hardy Weinberg equilibrium (P<1x10^-6) and SNPs with low call rates (<95%) were excluded. Blood pressure measurements and GWAS-data were available for 1,034 Caucasian subjects.

**Exclusion of SNPs**

SNPs with an imputation quality of R²<0.95 were excluded from the analyses. This was done for both the discovery and target samples (140k SNPs excluded in the discovery sample, 136k SNPs in adult target sample and 142k SNPs in the child target sample). For the discovery sample meta-analyses, SNPs available in less than 4 studies were excluded (572k SNPs). After all quality checks and exclusions, about 1.8 million SNPs were left in the discovery sample and 2.4 million SNPs in the target samples. Genetic risk scores were calculated using SNPs available in both samples (about 1.8 million SNPs).

**Data analysis**

**Statistical methods**

For linear regression models, the P-value for association with an increase in explained variance, was obtained from the F statistic which is calculated using the following formula:

\[
F = \frac{\left( SS_{reg} \text{ (model with genetic risk score)} - SS_{reg} \text{ (baseline model)} \right) / k}{MS_{res} \text{ (model with genetic risk score)}}
\]

Under the null hypothesis that both models perform equally well, F follows an F-distribution with df₁=k and df₂ = n-p-1, with k being the number of variables added to the model, and p the total number of variables in the full model. For logistic regression models we calculated the change in -2 log likelihood of the two models which follows a chi square distribution with degrees of freedom equal to the number of added variables (in our study this is always one) To evaluate the difference between two subsequent models both including a genetic risk score we calculated the Akaike Information Criterion (AIC) for each of the models. We evaluated the difference in the AIC between two models, which follows a chi square distribution with one degree of freedom under the null hypothesis. The AIC for linear regression...
models is derived from the following formula: \( AIC = N \ln(RSS/N) + 2(k+1) \) and in logistic regression the \( AIC = -2 \log \text{likelihood} + 2k \), with \( k \) being the number of variables in the model, \( RSS \) the residual sum of squares, and \( N \) the number of subjects in your sample.

**Unrelated phenotype**
As a control, associations of the genetic risk scores were also tested with two traits unrelated to blood pressure; intracranial volume in adults, measured as described previously\(^7\), and head circumference in children. Adjustment was made for sex and age.

**Random genetic risk scores.**
We conducted sensitivity analyses to assess whether similar results could be obtained by calculating a genetic risk score based on a randomly selected SNP set. We constructed five SNP sets consisting of \(~565k\) randomly selected SNPs, irrespective of their \( P \)-value in the discovery analysis. We subsequently used these randomly created SNP sets to calculate genetic risk scores for each individual and assessed the increase in explained variance similar to the main analysis. We calculated the average increase in explained variance of risk scores and the average Akaike Information Criterion (AIC) of the models\(^8\-\(^9\). Subsequently, we tested the difference in AIC between the random model and the original genetic risk score. The difference in AIC follows a \( \chi^2 \) distribution with one degree of freedom\(^8\).

**Genetic risk score based on a biological pathway**
To explore whether a genetic risk score comprised of SNPs in a signaling pathway, we included SNPs in the FGF (fibroblast growth factor) pathway as described by Tomaszewski et al in a genetic risk score\(^10\). Of the 79 SNPs described in the supplement of this paper, 31 SNPs were directly available in the CHARGE discovery sample. For the remainder of the SNPs we searched for good proxies (HapMap \( r^2 >0.8\)) to include in the SNP set. We were able to include an additional 7 proxy SNPs, adding up to a total of 38 SNPs in this SNP set based on a biological pathway. Using this SNP set we calculated genetic risk scores and evaluated the increase in explained variance added to the baseline model.
Funding sources for the cohorts in the CHARGE consortium

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


Childhood hypertension

**Figure S1a**
Increase in explained variance in childhood hypertension by systolic genetic risk scores in the Generation R Study.

**Figure S1b**
Increase in explained variance in childhood hypertension by diastolic genetic risk score in the Generation R Study.

Figure S1.
Bars represent the increase in explained variance (%) of childhood hypertension (SBP or DBP > p95 for age, gender and height) for the systolic and diastolic genetic risk scores for different p-value thresholds. The genetic risk scores were added to the baseline model: High blood pressure = BMI + SEX + AGE and the increase in explained variance was evaluated. The baseline model explained 3.2% of the variance in high blood pressure in children.

* p < 0.05
** p < 0.001

S1a. Systolic blood pressure scores
S1b. Diastolic blood pressure scores
Figure S2. Bars represent the increase in explained variance (%) of intracranial volume when adding systolic and diastolic genetic risk scores for different p-value thresholds, to the base line model for systolic blood pressure including sex and age as covariates. The baseline model explained 34.2 % of the variance in intracranial volume in adults.

* p < 0.05
** p < 0.001

S2a. Systolic blood pressure scores
S2b. Diastolic blood pressure scores
UNRELATED PHENOTYPE

Figure S3a
Generation R Study – SBP scores

Figure S3b
Generation R Study – DBP scores

Figure S3.
Bars represent the increase in explained variance (%) of head circumference when adding systolic and diastolic genetic risk scores for different p-value thresholds, to the base line model for systolic blood pressure including sex and age as covariates. The baseline model explained 6.6 % of the variance in head circumference in 6 year old children.

* p < 0.05
** p < 0.001
S3a. Systolic blood pressure scores
S3b. Diastolic blood pressure scores
Figure S4a
Increase in explained variance in systolic blood pressure in a strictly normotensive population in the Rotterdam Study III

Figure S4b
Increase in explained variance in diastolic blood pressure in a strictly normotensive population in the Rotterdam Study III

Bars represent the increase in explained variance (%) of systolic (S4a) and diastolic blood pressure (S4b) in adults for the genetic risk scores for different p-value thresholds, excluding subjects on anti-hypertensive medication and all other subjects with hypertension. The genetic risk scores were added to the baseline model: High blood pressure = BMI + SEX + AGE and the increase in explained variance was evaluated. The baseline model explained 7.3% and 6% of the variance in systolic and diastolic blood pressure in adults.

* p < 0.05
** p < 0.001

S4a. Systolic blood pressure scores
S4b. Diastolic blood pressure scores
Analyses using pruned data (Figure S5a-S7b).

**Figure S5a**
Increase in explained variance in systolic blood pressure by pruned genetic risk scores in the Rotterdam Study III

**Figure S5b**
Increase in explained variance in systolic blood pressure by pruned genetic risk scores in the Generation R Study

Figure S5
Bars represent the increase in explained variance (%) of systolic blood pressure for the genetic risk scores for different p-value thresholds. The genetic risk scores for systolic blood pressure were added to the baseline model: systolic blood pressure = BMI + SEX + AGE and the increase in explained variance was evaluated. The baseline model explained 12.8% and 5.2 % of the variance in systolic blood pressure for adults and children respectively.

* p < 0.05
** p < 0.001

5a. Rotterdam Study III
5b. Generation R Study
Analyses using pruned data

Figure S6a
Increase in explained variance in diastolic blood pressure by pruned genetic risk scores in the Rotterdam Study III

Figure S6b
Increase in explained variance in diastolic blood pressure by pruned genetic risk scores in the Generation R Study

Figure S6
Bars represent the increase in explained variance (%) of diastolic blood pressure for the genetic risk scores for different p-value thresholds. The genetic risk scores for diastolic blood pressure were added to the baseline model: diastolic blood pressure = BMI + SEX + AGE and the increase in explained variance was evaluated. The baseline model explained 8.4% and 1.4% of the variance in diastolic blood pressure for adults and children respectively.

* p < 0.05
** p < 0.001
6a. Rotterdam Study III
6b. Generation R Study
Analyses using pruned data

Figure S7a
Increase in explained variance in hypertension disease status by pruned systolic genetic risk scores in the Rotterdam Study III

Bars represent the increase in explained variance (%) of hypertension in adults (SBP>140, DBP>90 or the use of anti hypertensive medication) for the systolic and diastolic genetic risk scores for different p-value thresholds. The genetic risk scores were added to the baseline model: Hypertension = BMI + SEX + AGE and the increase in explained variance was evaluated. The baseline model explained 15.9% of the variance in hypertension in adults.

* p < 0.05
** p < 0.001

7a. Systolic blood pressure scores
7b. Diastolic blood pressure scores